Check for updates

Inside the *Next Black Box*: Examining Students' Responses to Teacher Feedback in a Formative Assessment Context

Angela M. Lui*† and Heidi L. Andrade

Department of Educational and Counseling Psychology, University at Albany, Albany, NY, United States

Attempts to explain inconsistencies in findings on the effects of formative assessment and feedback have led us to study the next black box: how students interpret and subsequently use formative feedback from an external source. In this empirical study, we explore how students understand and process formative feedback and how they plan to use this information to inform next steps. We present findings from a study that examined students' affective and cognitive responses to feedback, operationalized as emotions, interpretations (i.e., judgments, meaning making, attributions), and decision-making. Relationships among these processes and students' initial motivational states were also explored. Survey data were collected from 93 students of a 7th grade English/Language Arts teacher who employed formative assessment practices. The results indicate that students tended to have positive emotions and judgments in response to their teacher's feedback and make controllable attributions. They generally made informative meaning of the feedback and constructive decisions about next steps. Correlational findings showed that (1) emotions, judgments, meaning making, and attributions are related; (2) judgments of and the meaning that students made about the feedback were most strongly related to decision-making about next steps; and (3) task value was the only motivation variable related to responses to feedback. We conclude with implications for research and practice based on the expected and unexpected findings from this study.

Keywords: formative assessment, formative feedback, responses to feedback, emotions, interpretations, decision making, motivational states, affective and cognitive processes

INTRODUCTION

In 1998, Black and Wiliam used the term *black box* to emphasize the fact that what happened in most classrooms was largely unknown: All we knew was that some inputs (e.g., teachers, resources, standards, and requirements) were fed into the box, and certain outputs (e.g., more knowledgeable and competent students, acceptable levels of achievement) could follow. They proposed that a promising new conception of assessment—formative assessment—could raise standards and increase student learning by yielding actionable information on learning as it occurs. The primary purpose of formative assessment is to collect evidence of students' current levels of understanding

to inform further instruction by teachers and/or next steps for students to take to enhance their learning (Wiliam, 2010).

Research on formative assessment has expanded a great deal since Black and Wiliam's seminal work. Much but not all of that research supports claims of its effectiveness. One potential explanation for this inconsistency is how students interpret and respond to the feedback: how do they make sense of the feedback? What judgments and attributions do they make? What conclusions do they draw? We have little knowledge of how students respond to feedback, including the interpretations, sense-making, and conclusions they draw from formative assessment feedback (Leighton, 2019); hence, we dub this as the *next black box*. This study sheds light into this next black box by using a novel model of students' responses to feedback as the framework to examine how secondary English/Language Arts (ELA) students accustomed to receiving formative feedback responded to their teachers' feedback, cognitively, and affectively.

Formative Assessment and Feedback

Formative assessment, or assessment *for* learning, is defined as the integration of processes and tools that generate meaningful feedback about learning that can support inferences about next steps in learning and instruction (Andrade, 2016). Feedback, therefore, is an integral part of formative assessment. It is most useful when there is an influence on future performance, or at least an attempt to use the information from feedback to improve learning (Ramaprasad, 1983; Sadler, 1989; Black and Wiliam, 1998; Wiliam, 2011). Feedback should close the gap between where students are and where they need to be (Kulhavy, 1977; Ramaprasad, 1983; Butler and Winne, 1995; Hattie and Timperley, 2007), ultimately improving academic achievement.

In a review of 12 meta-analyses on feedback in classrooms, Hattie (2009) concluded that, under the right conditions, feedback in a formative context can greatly contribute to students' achievement, with an average effect size of 0.73. Consistent with Hattie, other reviews and meta-analyses of research on feedback have reported that feedback generally tends to have a positive association with learning and achievement (Hattie and Timperley, 2007; Shute, 2008; Hattie, 2009; Lipnevich and Smith, 2009a; Wiliam, 2010, 2013; Ruiz-Primo and Li, 2013; Van der Kleij et al., 2015). Most recently, Wisniewski et al. (2020) conducted a meta-analysis of 435 studies ($k = 994$, $n > 61,000$) using a random-effects model and found a medium effect size of feedback on student learning ($d = 0.48$).

Meta-analyses tend to obscure important inconsistencies, however. Despite positive findings in the extant research literature, negligible and even negative relationships have also been noted (Bennett, 2011). For example, Kluger and DeNisi (1996) reported a mean effect size of 0.41 for formative feedback on performance, but 38% of these effects were negative. Shute (2008) noted in her review of the literature on formative feedback that "despite the plethora of research on the topic, the specific mechanisms relating feedback to learning are still mostly murky, with very few (if any) general conclusions" (p. 156). Although Wisniewski et al. (2020) found a medium effect size, their moderator analyses suggest that to get a more accurate understanding of the effects of feedback on

achievement, it is important to differentiate between cognitive, motor, motivational, and behavioral outcomes, as well as the types and amount of feedback information conveyed. These critiques have prompted researchers to look more closely at research findings to explain the inconsistencies. One crucial aspect of formative assessment that has been often discussed theoretically but too rarely studied empirically is students' responses to feedback or the internal mechanisms involved in the reception of feedback and their influences on learning and achievement (Butler and Winne, 1995; Kluger and DeNisi, 1996; Shute, 2008; Draper, 2009; Bennett, 2011; Dann, 2014).

THEORETICAL FRAMEWORK

Defining Responses to Feedback

Responses to feedback can be broadly defined as the internal cognitive and affective mechanisms and overt behavioral responses or actions taken in response to feedback (Lipnevich et al., 2016). Empirically, researchers have tended to focus on conceptions of, perceptions of, and emotions elicited by feedback as key internal processes, with evidence that these processes might be co-occurring (e.g., Lipnevich and Smith, 2009b; Gamlem and Smith, 2013). However, these terms are not always well-defined (Fisk, 2017; Van der Kleij and Lipnevich, 2021). Furthermore, theoretical discussions suggest that conceptions, perceptions, and emotions are necessary but insufficient conditions for responses to feedback.

Consequently, we developed and tested a model of the internal mechanisms of feedback processing to use as the framework for this study. This model was based on the literature on conceptions, perceptions, and emotions elicited by feedback (e.g., Lipnevich and Smith, 2009b; Gamlem and Smith, 2013; McMillan, 2016; Fisk, 2017; Van der Kleij and Lipnevich, 2021), as well as models related to feedback and responses to it (i.e., Bangert-Drowns et al., 1991; Butler and Winne, 1995; Tunstall and Gipps, 1996; Allal and Lopez, 2005; Hattie and Timperley, 2007; Draper, 2009; Andrade, 2013; Gamlem and Smith, 2013; Narciss, 2013; Lipnevich et al., 2016; Tulis et al., 2016; Winstone et al., 2017; Leighton, 2019; Panadero and Lipnevich, 2021). In combination, the models suggest that responses to feedback are comprised of *initial states*, including conceptions of or beliefs about assessment and feedback, motivational determinants (e.g., self-efficacy, mindset, goal orientation, task value), and prior knowledge; *internal responses to feedback*, including interpretations and perceptions of, and attributions made regarding the feedback, emotions elicited by the feedback, and decisions about next steps; and overt, observable *external responses to feedback* or behaviors in response to feedback.

The literature also suggests that *contextual factors* influence how students respond to feedback, particularly the culture of assessment and learning (e.g., Shepard, 2000; Hattie and Timperley, 2007; Andrade, 2010; Havnes et al., 2012; Robinson et al., 2013; Lipnevich et al., 2016; Tulis et al., 2016; Winstone et al., 2017), the task (Butler and Winne, 1995; Hattie and Timperley, 2007; Andrade, 2013; Lipnevich et al., 2016; Pellegrino et al., 2016; Leighton, 2019), and the purpose and source of

feedback (Panadero and Lipnevich, 2021). Lui and Andrade (2022) offer an in-depth review of the literature that informed the model's design. We briefly describe the model in the next section.
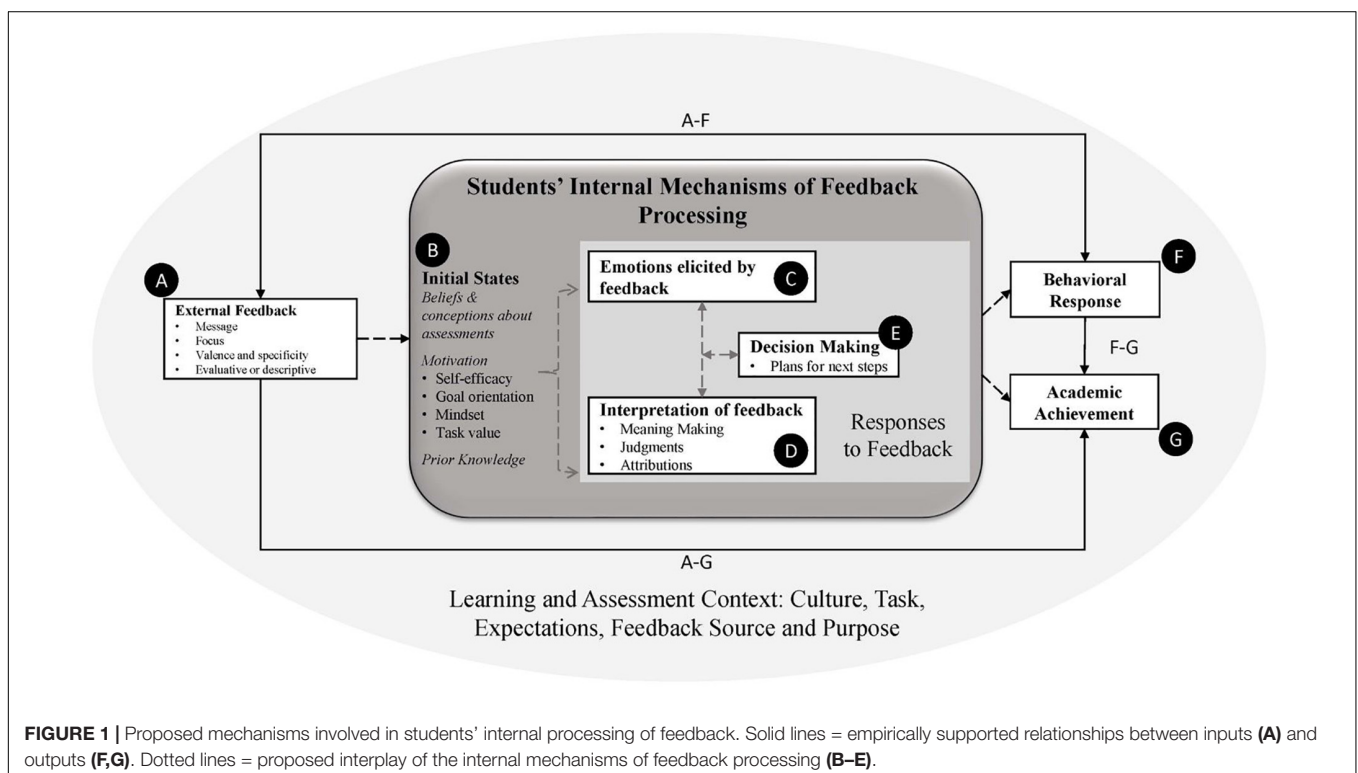
Model of the Internal Mechanisms of Feedback Processing

As shown in **Figure 1**, the external learning and assessment context, which includes the assessment and feedback culture, the task and the expectations regarding its quality, and feedback source and purpose, influences how assessment information and feedback are processed and used (e.g., Butler and Winne, 1995; Kluger and DeNisi, 1996; Shepard, 2000; Hattie and Timperley, 2007; Havnes et al., 2012; Andrade, 2013; Robinson et al., 2013; Harris et al., 2014; Lipnevich et al., 2016; Pellegrino et al., 2016; Tulis et al., 2016; Winstone et al., 2017). The internal mechanisms of feedback processing in **Figure 1** include inputs [external feedback (A)] and outputs [behavioral response (F) and academic achievement (G)]. As noted above, recent scholarship has focused largely on exploring the relationships between these inputs and outputs. Solid black lines represent the empirically tested relationships, which have been shown to be causally related (e.g., Coe et al., 2011; Phelan et al., 2012; Chen et al., 2017; Lee et al., 2020).

Hypothesized processes that lack sufficient empirical support include the internal mechanisms between the inputs and outputs (B-E). That is, there is well-established evidence of the existence of the initial states (B) in the dark gray box, including beliefs, motivation, and prior knowledge (e.g., Bangert-Drowns et al., 1991; Butler and Winne, 1995; Andrade,

2013; Narciss, 2013; Lipnevich et al., 2016; Tulis et al., 2016; Winstone et al., 2017; Panadero and Lipnevich, 2021; Van der Kleij and Lipnevich, 2021), as well as some evidence of associations between some but not all of those states (Wingate, 2010; Rakoczy et al., 2013; Robinson et al., 2013; Brown et al., 2016; Fatima et al., 2021). Evidence is lacking regarding the hypothesized ways in which, taken together, these initial states filter the information students receive and influence the ways they receive it—hence the dotted lines between B, C, D, and E in the figure. Similarly, research has demonstrated that feedback prompts emotions, interpretations (i.e., meaning making, judgments or perceptions, and attributions), and decisions about next steps (e.g., Draper, 2009; Lipnevich and Smith, 2009b; Gamlem and Smith, 2013; Leighton, 2019; Winstone et al., 2021), but the relationships between those processes and with students' initial states have not been empirically established, so they are also represented by dotted lines.

We hypothesize the following: positive emotions, controllable attributions, informative meaning making and positive judgments made about the feedback are positively correlated with each other, which will lead to adaptive decision making about next steps. These planned next steps might or might not translate into actual behavioral actions depending on how much time students are given to process and use the feedback, and how the processing might change during this time. These responses to feedback are driven by students' initial states, or the beliefs, motivation, and prior knowledge. These initial states are influenced by the context in which students learn. So, a student will likely possess the beliefs and motivation to respond positively to feedback if they are situated in an assessment and learning

FIGURE 1 | Proposed mechanisms involved in students' internal processing of feedback. Solid lines = empirically supported relationships between inputs **(A)** and outputs **(F,G)**. Dotted lines = proposed interplay of the internal mechanisms of feedback processing **(B–E)**.

context in which a culture of critique is established, errors and mistakes are treated as opportunities to learn, and students are given feedback with opportunities to revise.

Research Questions

This study explores our hypotheses about the detailed components of responses to feedback, as well as the relationships represented by the dotted lines in the model, in the context of a secondary ELA class in which formative assessment processes were implemented. This study was guided by the following research questions:

1. What are students' responses to their teacher's checklist-based written feedback on a draft essay? That is, what are the emotions elicited by the feedback, the meanings, judgments, and attributions made about the feedback, and decisions about next steps based on the feedback?
2. What are the relationships among students' responses to feedback, their motivational states, and the messages the feedback conveys?

MATERIALS AND METHODS

Sample and Context

Sample: Students' Demographic and Academic Information

A sample of 112 7th grade students was recruited from a large suburban middle school serving approximately 1,200 sixth- to eighth-grade students in the Northeast region of the United States. These students were in one of four classes taught by a white/European-American ELA teacher with over 17 years of teaching experience. A literacy education expert identified this teacher as an effective, experienced teacher who implements formative assessment practices into her regular classroom routines. When this study was conducted, she co-taught Social Arts—a combination of ELA and Social Studies—with a social studies teacher. Students who provided both assent and parental consent were included in the study; students whose data later reflected systematic missingness (i.e., items from entire scales were left unanswered) were excluded. Our final sample included 93 students (56% female; 12–13 years old). Demographic information is summarized in **Table 1**.

Six students (50% females) also participated in think aloud protocols. These students were selected to represent a balance of gender, race and ethnicity, and achievement. **Table 2** includes four sources of information about these students' achievement in ELA: Scores on the State examination and ELA course grade from the year before the study was conducted, grades for essays written in 7th grade immediately before the beginning of the study, and the grades ultimately assigned to the essay examined for this study after the formative assessment and data collection.

The Learning and Assessment Context

The first author visited the classroom regularly ($\sim$3$\times$/week) for five months to understand the learning context, including

TABLE 1 | Demographic and academic background information for quantitative sample ($n = 93$).

	Demographic information	n	% of total sample
Gender	Female	52	55.9%
	Male	41	44.09%
Age	12	37	39.78%
	13	56	60.22%
Ethnicity	Asian or Asian American	23	24.73%
	Black or African American	3	3.23%
	White or European American	67	72.04%
6th Grade ELA Perf	Level 1	14	15.05%
	Level 2	31	33.33%
	Level 3	18	19.35%
	Level 4	14	15.05%
English as a New Language	Yes	3	3.23%
Student with disability	Yes	9	9.68%
Economically disadvantaged	Yes	14	15.05%

TABLE 2 | Demographic and academic background information for qualitative sample ($n = 6$).

	Students					
Demographics	YN	DJ	KF	KG	SM	VS
Gender	F	M	M	F	M	F
Age	13	12	13	13	13	12
Race/Ethnicity*	A/AA	W/EA	A/AA	W/EA	A/AA	W/EA
Economically disadvantaged	N	N	Y	N	N	Y
Grade 6 Statewide ELA	Level 3	Level 2	Level 2	Level 1	Level 4	Level 3
Grade 6 ELA grade	95	93	93	91	95	99
Previous essay score	83	80	93	85	90	88
Present essay score	94	91	91	89	91	89

A/AA, Asian/Asian American; W/EA, White/European American.

the culture of learning and assessment and the teacher-student dynamics. These visits revealed that the teacher's 7th grade classes were student-centered and project-based, with formative assessment practices seamlessly embedded into routine classroom practices. The teacher often provided feedback to students based on patterns she saw in students' understanding and misconceptions. Teacher-created handouts and graphic organizers were used to support students' practice of self-regulated learning during class activities. For example, a reading calendar helped students plan and manage independent reading. Planning packets were used at the beginning of each new project to allow students to develop their ideas at their own pace.

Formative and summative assessment processes were incorporated into classroom practices. For example, the teacher communicated expectations or criteria by sharing and discussing assessment checklists at the beginning of every task, modeling answers, and role-playing conversations. She frequently provided

feedback, including delayed, written feedback with a rubric or checklist followed by opportunities to revise. She would also provide immediate, spontaneous, verbal feedback while walking around the classroom during individual and small group work. Students were explicitly given time to self-assess their work using checklists.

Grades were given only for the final versions of student work. The ELA teacher admitted that effort factored into the grades she assigned for the essays, especially effort in revising based on feedback. Because the assignment that was used for this study was based on social studies content, the teachers agreed to base final grades for ELA on effort (which was not included in a checklist or rubric), while social studies grades focused on completion of assignment and accuracy of content, about which students were given a checklist.

The Task: Writing Assignment on U.S. Presidents After the Revolutionary War

The first author and the teacher selected a writing assessment within the curriculum for this study. This assignment required students to write an essay about U.S. Presidents after the Revolutionary War. The teachers developed this based on the *Writing Like a Historian* curriculum.[1] Students were asked to identify and write about two Presidents who made the greatest impact and one who made the least impact, with justification and evidence.

The ELA teacher introduced the assignment to students by reviewing the description and checklist she created (see **Supplementary Appendix A**). Although the criteria on the checklist were not co-created with students, each item on the checklist was presented in detail and reviewed throughout this assignment. Students had many opportunities to ask questions about them. They were encouraged and expected to use the checklist for self-assessment and progress monitoring. Students were given three class periods and one study period, if needed, to work on the essay. Essays were handwritten on loose-leaf paper, double-spaced.

After students submitted the first draft of their essay, the ELA teacher spent four weeks writing feedback based on the checklist and focused on writing skills and the accuracy of the content. The teacher returned students' essays, marked up, with a filled-out checklist stapled to the top of the essay. Students were given one class period (80 min) plus extra time, if needed, to make revisions.

Study Procedures

We used a mixed methods design, with think alouds and surveys, conducted in a naturalistic setting to collect rich data about students' responses to feedback. **Figure 2** illustrates the data collection procedures. The timing was critical: Surveys of Motivational States were administered first to capture students' initial motivation beliefs one week before receiving feedback. One week later, the day students received their essays with feedback, data on their responses to feedback were collected for this study. Students were first given the opportunity to review and reflect on the feedback; think aloud participants

---

[1] https://sheg.stanford.edu/history-lessons

were drawn aside and asked to go through the feedback aloud. Immediately after, all students completed the Responses to Feedback (RtF) survey in Qualtrics on their laptops. Thus, the survey was administered when students first received feedback on the targeted writing assignment, capturing students' responses in their rawest form. Students had their essays with teacher feedback to reference as they completed the survey, which took no more than 15 min to complete.

Measures
Survey of Motivational States

Three established scales were used to measure students' motivational states: the Implicit Theories of Intelligence Questionnaire (Dweck, 2000) was used to measure mindset; the Personal Achievement Goal Orientation Scale (Midgley et al., 2000) was used to measure students' goal orientation; and the self-efficacy for writing scale from Andrade et al. (2009) was adapted to measure self-efficacy for writing. See **Supplementary Appendix B** for a copy of this survey.

Mindset Scale

With a total of six items, three of the items on the Implicit Theories of Intelligence Questionnaire (Dweck, 2006) refer to fixed views of intelligence, and three items refer to growth views of intelligence. The six items are scored from 1 (*strongly disagree*) to 6 (*strongly agree*). Fixed mindset items were reversed coded then averaged with the growth mindset items so that higher scores represent a growth mindset. The internal consistency reliability estimate of the mindset scale for this study was $\alpha = 0.84$.

Personal Achievement Goal Orientation Scale

The Personal Achievement Goal Orientation (PAGO) scale is a part of the Patterns of Adaptive Learning Scales (PALS; Midgley et al., 2000), developed to examine relationships between students' learning environment and motivation, affect, and behavior. The PAGO uses a five-point Likert-type scale, ranging from 1 (*not at all true*) to 5 (*very true*). Three subscales of goal orientation were measured, including (a) Mastery (5 items; $\alpha = 0.89$), which describes students who perform to develop their competence; (b) Performance-Approach (5 items; $\alpha = 0.90$), which describes students who perform to show competence, and (c) Performance-Avoidance (4 items; $\alpha = 0.89$), which describes students who avoid the demonstration of incompetence.

Self-Efficacy for Writing

The 11-item self-efficacy for writing scale was adapted based on students' writing skills and knowledge that the ELA teacher deemed essential and relevant to the target writing assignment. Students were directed to rate their confidence level on a scale of 0–100. The 0–100 format has shown stronger psychometric properties than a scale of 1–10, both in terms of factor structure and reliability (Pajares et al., 2001). The internal consistency reliability estimates of this scale for this study was $\alpha = 0.93$.

Task Value

Three items were used to measure students' task value, specifically their perceived task importance, relevance, and interest. Response options ranged from 1 (*very low*) to 5 (*very high*). The internal

FIGURE 2 | Data collection procedures.
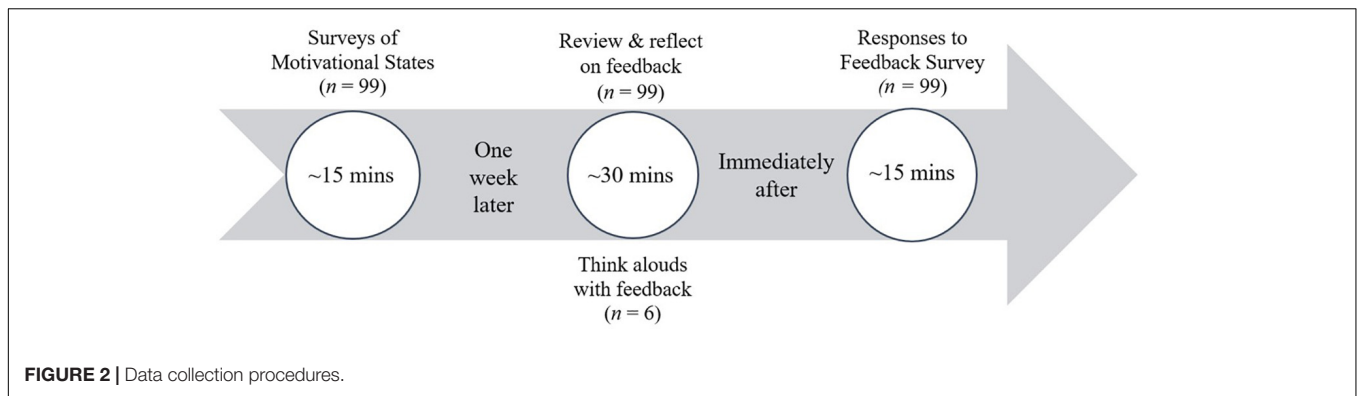
consistency reliability estimates of this scale for this study was $\alpha = 0.88$.

The Teacher Feedback: Handwritten With a Checklist

Students received feedback from their teacher in the form of an annotated checklist and comments and corrections written on the essay. Each essay had a checklist with teacher marks next to each criterion; notes are also written under each criterion that needed more work. The checklist was attached to the top of students' essays, on which handwritten feedback was also written with colored ink (not red), including corrections or flagging; underlines, circled words or sections with abbreviations denoted (e.g., MW = missing word; WC = word choice; PT = past tense); and questions and suggestions in the margins. No grades were included on any of the essays as part of the feedback.

There was a median of 32 discrete feedback instances on the checklist (range = 11–38) and 53 discrete feedback instances on the essays (range = 10 to 201). Using *a priori* codes adapted from the Meaning Making items, feedback messages were coded by one trained rater and the first author. **Table 3** summarizes the types of feedback messages, examples, and numbers and percentages of essays. Each type of feedback message was present on the checklist attached to the essays and the essays themselves. Rater training was conducted with six randomly selected essays. Two essays were scored together with the coding protocol; discrepancies indicated areas needing clarification. Minor revisions and additions were made to the protocol based on these discussions. The other four essays were rated independently then reviewed together. Once we achieved exact percent agreement above 85%, essays were randomly and equally split among the two raters, with one calibration essay after every sixth essay to monitor drift in interrater reliability. Disagreements were discussed and resolved.

Think Aloud Protocol and Coding Protocol

Think alouds were used to examine how students responded to feedback immediately upon receiving the feedback. The protocol was designed based on guidelines suggested by Ericsson and Simon (1993) and Green and Gilhooly (1996). During the think alouds, students were asked to state aloud what they were thinking and how they were feeling as they were first reading through their essay with their teacher's feedback. Upon extended silence, students were prompted with questions like, "What are

you feeling?" "What are you thinking?" and "What do you see?" Students were asked follow-up questions after they had reviewed their teacher's feedback aloud. These questions included, "What does the feedback mean to you?" "What does it make you think or feel?" and "What do you think you'll do about the feedback?" The think alouds were audio-recorded and transcribed.

The Responses to Feedback Survey

The Responses to Feedback (RtF) survey is based on the model of the internal mechanisms of feedback processing (**Figure 1**). The survey was piloted, and validity and reliability evidence (construct relevance, representativeness, response processes, and internal consistency) were examined as a part of a larger study (Lui, 2020). The final version of the survey (see **Supplementary Appendix C**) consists of five scales—Emotions, Judgments, Meaning Making, Attributions, and Decision-making—with six-point Likert-type items using one of two anchor types: 1 (*strongly disagree*) to 6 (*strongly agree*), or 1 (*definitely not*) to 6 (*most definitely*). The five scales are parsed into eight subscales. See **Table 4** for item functioning and Cronbach's alpha internal consistency estimates. Because the Decision-making scale comprises items that measure concrete and distinct decisions about next steps, they were treated as 12 individual items. The Uninformative Meaning Making subscale was omitted from this study because of unacceptable internal consistency.

Data Analysis
Research Question 1

Quantitative and qualitative methods were used to examine students' responses to feedback. Their responses were examined quantitatively using the self-report RtF survey. Descriptive statistics of students' responses to this survey were calculated using the Psych R Package (Revelle, 2019) to examine the nature of students' Emotions, Judgments, Meaning Making, and Attributions. Because the Decision-making items are treated as individual items for analyses, stacked barplots were created using the Likert R Package (Bryer and Speerschneider, 2016) to visualize frequency distributions of responses for each item.

Qualitatively, think alouds were conducted with six students during their first exposure to their teacher's feedback to capture their responses to feedback in their rawest form. To code the transcriptions, the first author and one trained rater first

TABLE 3 | Intended meaning of feedback: Examples and count of essays in which feedback was present.

Feedback codes	Example feedback on checklist[t]	Number of essays (%)	Example feedback on essay*	Number of essays (%)
Expectations	The checklist itself	93 (100%)	• Underlined *"The Missouri Compromis was a disagreement between the South and the North?"* with comment, "How does this show Monroe had the least impact?" • Circled *"are great presidents"* with a comment "Stick to the essay questions"	17 (18%)
Questions	"Washington?" connected to a single underline of "major" in the criterion, *"Explain a major event that shaped the nation during their term(s)"*	6 (6%)	• Circled *"disaster"* in *"That lead to the Embargo Act which was a disaster"* with a question, "How so?" • Underlined a paragraph with a question, "Why is this important?"	89 (96%)
Corrections	No feedback instances for this code	0	Commas added; words circled and noted as "Unnecessary"	90 (97%)
Comments— strengths	• Criteria that are checked • Comments on what was done well. (e.g., "Great draft! I've left some revision feedback ☺")	90 (97%)	Circled *"irrelevant"* with a comment, "Great word choice!"	1 (1%)
Improve—not how	Criteria that are marked with a circle and accompanied by a comment (Feedback with this code are more specific than those coded under "Improve—not how")	90 (97%)	"Repetitive"; "WC" for word choice; "SP" for spelling;	93 (100%)
Suggestions—next steps	Triple underline "Describe" in the criterion, *"Describe each president's time in office"* with an additional written comment underneath, "not much detail on Jefferson"	90 (97%)	Underlined a phrase with a comment, "Need to be more specific"	58 (62%)
Resources	–	0	–	0
Encourage help-seeking	–	0	–	0

*[t]Checklist criteria are italicized; *Students' writing is italicized.*

TABLE 4 | Scales and subscales of the responses to feedback (RtF) survey*.

Scales (# of items)	Anchor type	Subscales details (# of items, Cronbach's alpha)
Emotions (21 items)	*Very untrue of me* (1) to *Very true of me* (6)	Positive Emotions (10 items; α = 0.89) Negative Emotions (11 items; α = 0.92)
Judgment (18 items)	*Strongly disagree* (1) to *Strongly agree* (6)	Positive Judgments (9 items; α = 0.88) Negative Judgments (9 items; α = 0.89)
Meaning Making (12 items)	*Strongly disagree* (1) to *Strongly agree* (6)	Informative Meaning Making (9 items; α = 0.76) Uninformative Meaning Making (3 items; α = 0.59)
Attributions (11 items)	*Strongly disagree* (1) to *Strongly agree* (6)	Controllable (4 items; α = 0.83) Uncontrollable (7 items; α = 0.77)
Decision-making (12 items)	*Strongly disagree* (1) to *Strongly agree* (6)	N/A

*A copy of the survey is provided in **Supplementary Appendix C**.

read through the transcriptions at least once to get a general sense of the content. Then they separated the transcriptions into meaningful units and coded them using a coding protocol based on *a priori* domains from the model (i.e., emotions, judgments, meaning making, attributions, decision-making; see **Supplementary Appendix D** for coding protocol), with opportunities to create new codes. The first two transcriptions were coded together, with discussions to refine definitions and establish new codes, as appropriate. Once there was consistency in the understanding of the codes, raters coded the remaining transcriptions independently and met to discuss until there was consensus on all of the codes. The final codes are provided in **Table 5**.

Research Question 2

Correlations were used to examine relationships among responses to feedback (emotions, judgments, meaning making, attributions, decision-making), motivational states, and types of feedback messages. Pearson's correlations were used to examine relationships among emotions, judgments, meaning making, attributions, motivational states (self-efficacy, goal orientation, mindset, task value), and the presence or absence of types of feedback messages. Spearman correlations were used to examine the relationship between these variables and decision-making items, which are ordinal.

Before conducting correlational analyses, assumptions of linearity were checked using scatterplots, and the presence

TABLE 5 | Codes from the think alouds, by responses to feedback domains.

Domain definitions	Emerging codes		
Emotions: Feelings students expressed, which can be inferred by the raters, or explicitly noted by students.	● Appreciative/grateful ● Confused ● Deflated ● Disappointed	● Doubtful ● Happy ● Glad ● Relieved	● Satisfied ● Surprised ● Validated
Judgments: These are evaluations of the feedback. Could be in terms of usefulness, relevance, accuracy, clarity, respectfulness, tone.	● Accurate/Inaccurate ● Agree/Disagree ● Clear/Unclear ● Expected/Unexpected ● Fair/Unfair	● Positive/Negative ● Right/Wrong ● Straightforward ● Helpful/Unhelpful ● Informative/Uninformative	
Meaning Making: Processes to make sense of the feedback, including explaining the feedback (what is the feedback asking me about, telling me I did wrong/right, suggesting that I do or change?); explaining the original work to justify why the feedback is there; justifying the original work, and justifying the feedback.	● Apply the feedback (with and without uncertainty) ● Explain the feedback (with and without uncertainty) ● Justify original work ● Justify the feedback (at times by explaining original work) ● Reiterate original work ● Reiterate teacher feedback ● Uncertainty		
Attribution: Reasons or explanations students make for the feedback they received. Attributions that emerged from the data included attention, available resources, effort, growing up, random, technical knowledge.	● Attention (rushing, noticing) ● Effort ● Resources (or lack thereof) ● Knowledge and skills (or lack thereof)		
Decision Making: Statements about what students will/can do with the feedback, whether task specific or not. Codes that emerged from data included task specific behavioral next steps, essay-specific changes, transfer of skills, and uncertainty about next steps.	● Behavioral next steps (i.e., help seeking, researching, revise and edit, use available resources, think about it) ● Essay-specific changes to make (i.e., content changes, edits to grammar, spelling, conventions, word choice, include more information, make corrections teacher noted) ● Uncertain about next steps		
Other	● Judge own work/self-assessment		

of extreme outliers was checked using boxplots. No extreme violations of the assumptions were found according to the guidelines laid out by Cohen et al. (2003). Holm-Bonferroni adjustments, a less conservative and more powerful modification of the Bonferroni method (Holm, 1979), were made to account for the possibility of Type 1 error due to multiple comparisons. The GGcorrmat function in the ggstatsplot R package (Patil, 2021) was used to generate correlation matrices with Holm-Bonferroni adjustments, with the target alpha level at $\alpha = 0.05$.

RESULTS

Research Question 1: Students' Responses to Feedback

In general, think alouds revealed themes that are consistent with the five domains measured in the RtF survey (emotions, judgments, meaning making, attributions, and decision-making): emotions were elicited during students' processing of their teacher's feedback, they made judgments about and meaning of the feedback, attributed the feedback to controllable and uncontrollable factors, and decided on the next steps they would take in response to the feedback. Nuanced differences were found between the meaning making processes that emerged from students' think alouds and the processes represented by items on the RtF survey. One theme emerged from the think alouds that was not reflected in the RtF survey: reflection on and judgments of one's own work. Results are presented in more detail below

by the six domains (emotions, judgments, meaning making, attributions, decision-making, and reflection on and judgments of one's own work).

Emotions Elicited by the Feedback

According to the RtF survey results, the teacher's feedback elicited neither strongly positive nor negative feelings; but students' emotions were slightly more positive ($M = 3.32$, $SD = 0.99$) than negative ($M = 2.41$, $SD = 1.05$). At an item level, students *agreed* to *strongly agreed* that the feedback elicited feelings of "interest" (73% of students), "calm" (63% of students), "hopeful" (59% of students), "concern" (45% of students), "uncertainty" (41% of students), and "disappointment" (40% of students).

The think aloud data revealed a similar pattern: students were generally satisfied with the feedback, noting that it was well-deserved and informative. For example, YN said, "The word 'crucial' is misspelled. I had some doubt about that at first, so I'm glad I know that's not the correct spelling," and KG voiced, "I feel like what I did and what got circled was well-deserved. I am satisfied with it." Negative emotions elicited by the feedback were generally not about the feedback but the quality of their work: "Well, I didn't expect to make that many mistakes;" "I thought I did a little bit better. I agree with what they're saying now that I read over it, but I first was kind of like, oh, I thought I did a little bit better than this."

Judgments About the Feedback

Students reported strong positive judgments ($M = 4.42$, $SD = 0.91$) about the feedback compared to negative judgments

(M = 1.92, SD = 0.88). At an item level, the majority of students agreed that the feedback was "right" (95% of students), "helpful" (91% of students), "informative" (88% of students), "respectful" (87% of students), "clear" (85% of students), and "specific" (81% of students). A smaller proportion of students judged the feedback negatively: "surprising" (50% of students), "confusing" (32% of students), "unpleasant" (22% of students), and "vague" (19% of students).

This finding is similar to the judgments that emerged in the think alouds (e.g., agree/disagree, clear/confusing, useful, makes sense, fair, helpful, and manageable; see **Table 5**). Judgments were closely tied with meaning making processes. None of the three students who disagreed with the feedback explicitly stated that they thought the feedback was incorrect; it was through justifying their writing or reflecting aloud on it that their judgment of inaccuracy was revealed. For instance, KG responded to a question the teacher had about the accuracy of a piece of historical detail, "She circled, 'they came to the U.S. and seized' and she wrote 'accurate'? Pretty sure I read that somewhere in the textbook. Maybe...I don't know how that wouldn't be accurate. I'm not sure about that." SM said in response to a section that the teacher marked as fragmented sentences, "I wasn't sure like what was wrong with it because it seemed okay to me."

Making Meaning of Feedback

According to the responses on the RtF survey, students tended to make informative meaning out of the feedback they received (M = 4.01, SD = 0.79), particularly agreeing with the following items: "includes specific comments about what I should improve" (91% of students), "explains what I am expected to do for this assignment" (90% of students), "includes questions about what I wrote" (88% of students), and "gives suggestions on what I can do next" (81% of students). Items with the least agreement were "includes specific comments about what I did well" (32% of students) and "gives suggestions on resources I could use" (39% of students).

Think aloud data shed light on processes that students took in making meaning of the feedback, which involved explaining or justifying the feedback or original work. During the think alouds, students were asked to explain, in their own words, the information that the feedback was providing. While explaining, all six students justified the feedback, usually by explaining what they did or did not do to elicit it. For example, SM justified commas that the teacher added to his essay, "I can see she added some commas. That means it needs a pause, I think." KF thought aloud about a section of text in his essay that the teacher crossed out:

> She crossed out, "there was an expedition" and "which is called" and she just put "Jefferson" between "it" and "sent" so it sounds like, "The event happened in 1803 after they bought it. Jefferson sent Lewis and Clark...oh! Jefferson sent the Lewis and Clark expedition, who explored the territory," which is the same meaning. I don't know why I wrote so much.

Students also justified their work when making sense of the feedback, often when they disagreed with the feedback, at least initially. SM thought aloud as he made sense of one piece of

feedback that misaligned with his research and understanding of the Chesapeake and Ohio Canal, saying, "She said to put ['Canal' in Chesapeake and Ohio Canal] as plural, but I searched it up, and it's just one canal. But it has 'and' in its name and might be confusing." KF responded to a "how so?" question that the teacher had written in the margin about his claim that "Missouri Compromise was big but not as major or had bigger impact than the events of the other presidents did or were involved in," first by explaining the historical content that he wrote about, then adding, "And so like, [the Missouri Compromise] is not as important as like the Louisiana Purchase because without the Louisiana Purchase, Missouri wouldn't even happen. So, I don't think it's as important, but it's still important though, so that's why I included it in my thing, I guess."

Students also conveyed uncertainties during their review of the feedback. Starting with a slight chuckle and sigh, DJ admitted, "I honestly don't know why she put that arrow there," and then proceeded to the following feedback instance without trying to figure it out. VS made uncertain meaning out of markings that were confusing to her, saying, "[The teacher] circled the comma here. I put a comma there. I think I shouldn't have put it there, then?"

Attributions Made Based on the Feedback

Based on results from the RtF attributions scale, students tended to make more controllable attributions (M = 4.11, SD = 0.79) than uncontrollable attributions (M = 3.07, SD = 0.86). Controllable attributions with which students agreed include (1) "my understanding of the content" (72% of students), (2) "the strategies I used to write it" (72% of students), (3) "my understanding of what I had to do for the essay" (68% of students), and (4) "the effort I put into writing it" (68% of students). More than half of the students also agreed with two uncontrollable attribution items, which state that they received the feedback they did because of "what my teacher thinks about my understanding of the content" (66% of students), and "what my teacher thinks about my understanding of the assignment" (63% of students).

Think aloud data provided further insights into these attributions: More than half of each student's attributions were related to technical knowledge or specific writing techniques or methods they used or did not use. In response to a teacher's comment about connections between the conclusion and essay question, DJ said, "so I didn't really add a conclusion about the question that the essay was asking." Effort and attention were the second most common controllable attribution mentioned by five of six students. YN said, "So I should have edited that before I handed it in." KG expressed a similar thought as she retrospectively thought about her feedback, "And that was on me; I could have gone back and done that and checked it a little bit better."

Uncontrollable attributions emerged as well but less frequently. For example, in response to a content accuracy question the teacher asked, VS said, "We weren't allowed to open a textbook and look in the textbook for information because we were supposed to have it in our binder." Time was a reason for mistakes given by two students. KF noted, "I was one of the

lastest person to finish, so I was kind of getting rushed at the end. So, I was quickly writing things that would hopefully make sense." YN explained that she made mistakes "because I was in quite a rush."

Making Decisions About Next Steps

Almost all of the students agreed with constructive next steps (**Figure 3**), i.e., making the corrections the teacher noted (96% of students), reviewing requirements of the essay (94% of students), and rereading the feedback (92% of students); and disagreed with unconstructive or drastic next steps, including ignoring the feedback (96% of students), not changing the essay (93% of students), and starting over (90% of students). Most students also agreed that they would find more information or materials to include in their essays (84% of students).

Consistent with the survey findings, students who participated in the think alouds offered thoughts about what they would do next, most of which were task-specific and adaptive, especially when the feedback was in the form of corrections or error flagging. For example, in response to his teacher's "☺" on his use of the first-person pronoun, SM shared, "I might just want to take ['I can assure you'] out and say, 'This is the most exciting thing of all his presidency,' or take out the whole sentence."

Unless students disagreed with the feedback, their plans tended to be task-specific behavioral next steps, such as asking for help or referring to resources. For example, YN said in response to a comment she was unsure of, "I'm going to plan on asking [the teacher] what she meant, and then I'm going to revise that." In response to a question the teacher asked about the accuracy of historical content, VS said, "I would maybe look into the textbook again and rephrase it how the textbook said it." YN received similar feedback and responded, "I wrote this as a guess, not as, not exactly with confidence, so I have to research that and make sure that's accurate."

Students also acknowledged when they were unsure of their next steps. SM thought aloud, "So I'm trying to figure out a way to try it, like make this more clear that it's connecting, the connection between the two [ideas]." VS tried to make revisions on the spot during the think aloud, but eventually ended by saying, "I don't really know how to phrase it here, but I understand what she wrote."

Reflection on and Judgment of One's Own Work

Reflection is not a subscale in the RtF survey but emerged as an essential process during the think alouds. Students reflected on how their work would be once they implemented the feedback, or what they should have done before submitting their essay to have avoided the feedback. Usually, upon reading and making meaning of the feedback, students made statements that started with, "I should have done this. . ." or "There should be. . ." or immediately implemented a response to the feedback. For example, YN reflected on the teacher's correction on her essay, saying, "there should be a comma over here." KF responded to corrections that the teacher made to one of his sentences, first explaining then applying the feedback: "She completely crossed out 'people with money' delete-she said to delete 'also known as'

and just put 'wealthy' so it sounds like," "he supported common people, not just the wealthy."

There were also situations when students responded to feedback by reflecting on their original thinking and writing processes. For example, the teacher underlined the word "died" and asked, "because of force?" Understanding that this question was asking why the army died, SM reflected on his understanding of the content and why this detail was left out, stating, "I wasn't sure if it was that the thousands died because of the force or they were just dying because of the harsh conditions of having to walk all those miles." VS reflected more generally, "Oh yeah that makes sense. What was I thinking when I wrote this?!"

Two students judged their work during the process of responding to feedback. For example, DJ reflected on the overall picture he gathered from all the feedback he received: "Well, I think I didn't like do terribly bad. I like-I didn't do the greatest, but I didn't do the worst, so like I'm kind of in the middle because I got. I got four things circled and like six or seven things checked." In response to the teacher crossing out "event sorta says" on his essay, he judged his work as not good: "[. . .] which you don't really need that, cause it doesn't, well nobody wants to read that, okay? Because, I don't know. It—it—it's not good. 'Event sorta says'? Tha-tha-that's not good."
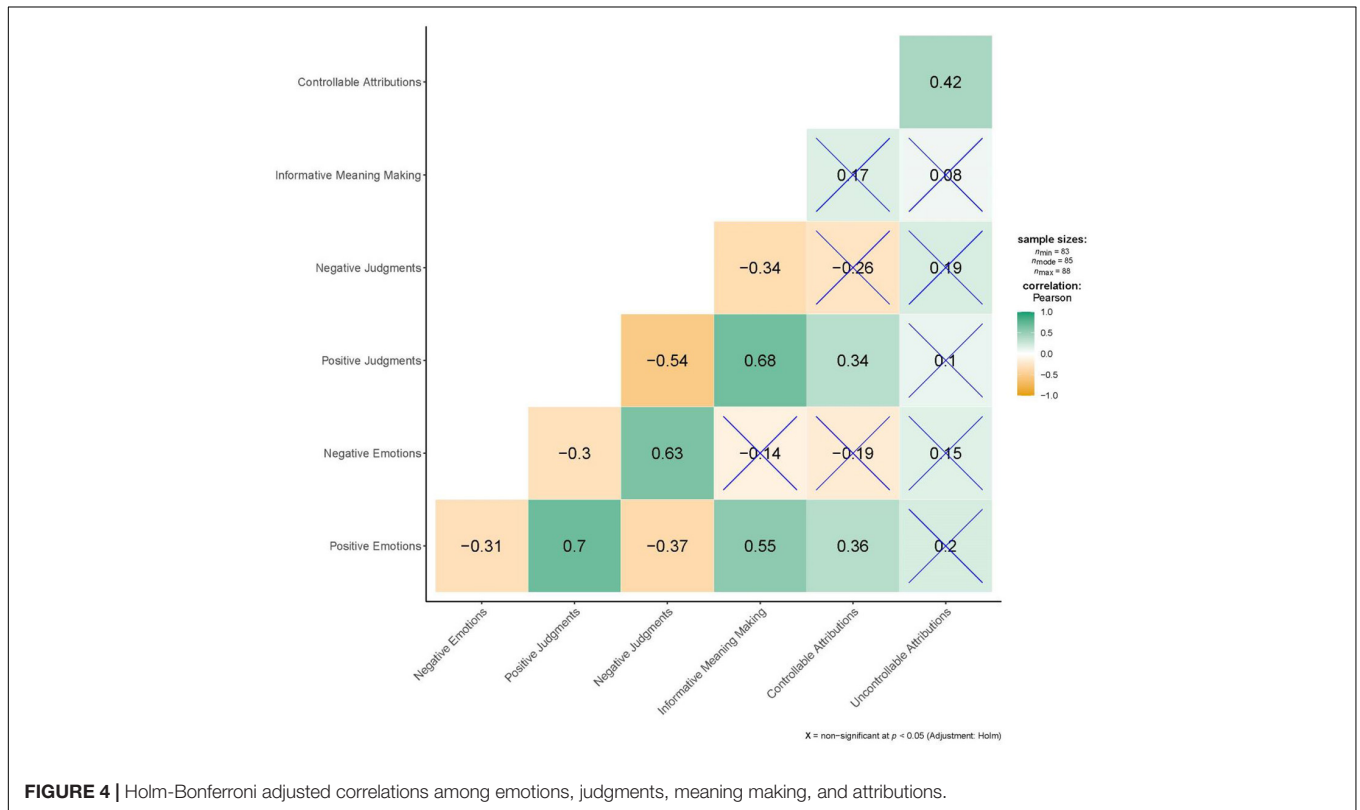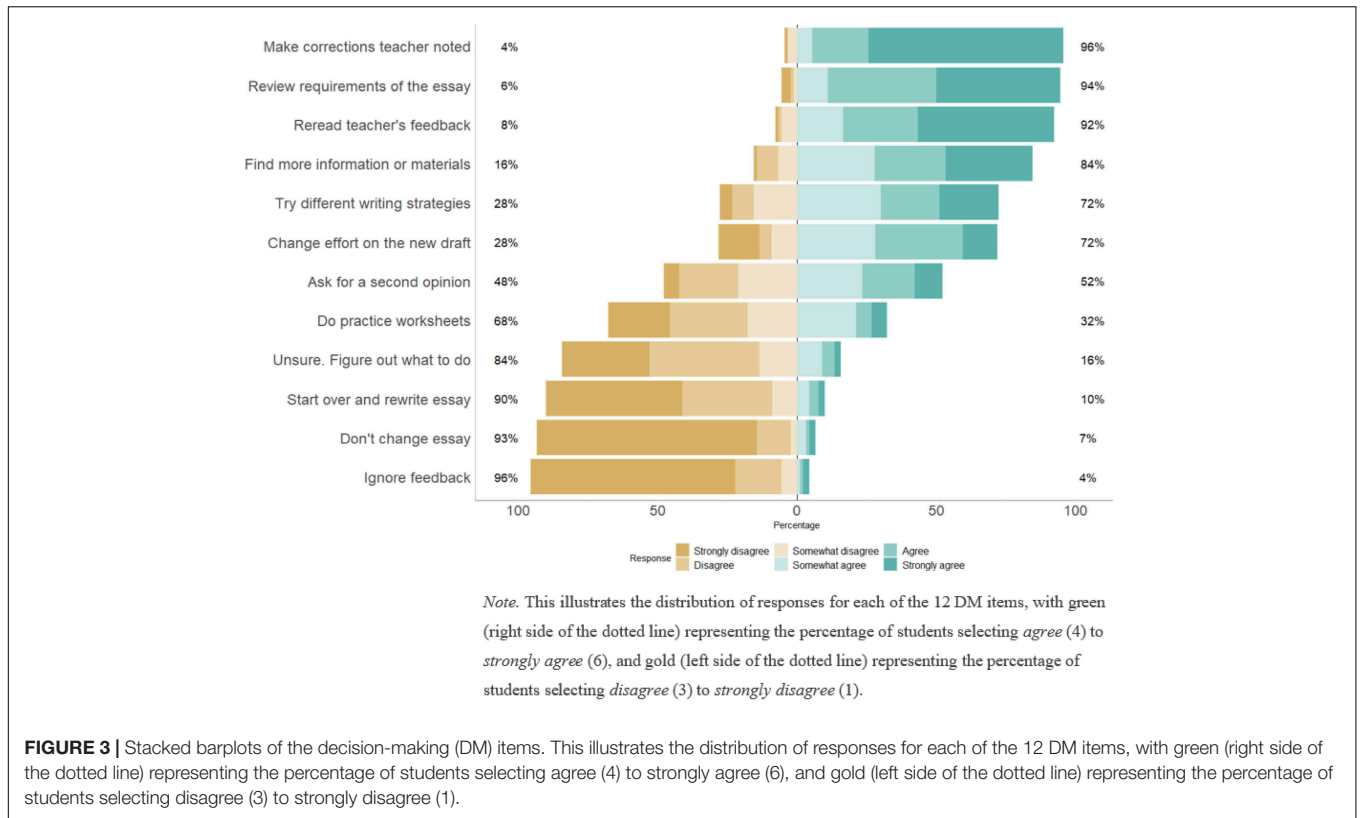
Research Question 2: Relationships Among Students' Responses to Feedback, Motivational States, and Feedback Messages

Emotions, Judgments, Meaning Making, and Attributions

Holm-Bonferroni adjusted Pearson's correlations among Emotions, Judgments, Meaning Making, and Attributions are shown in **Figure 4**. Generally, positive responses tended to be positively related to each other and negatively related to negative responses. In contrast, negative responses to feedback tended to be positively correlated with each other and negatively correlated with positive responses. For example, Positive Emotions were positively correlated with Positive Judgments ($r_{adj} = 0.70$), Informative Meaning Making ($r_{adj} = 0.55$), and Controllable Attributions ($r_{adj} = 0.36$), and negatively correlated with Negative Judgments ($r_{adj} = -0.37$). Positive Judgments were positively correlated with Informative Meaning Making ($r_{adj} = 0.68$), and Controllable Attributions ($r_{adj} = 0.34$), and negatively correlated with Negative Judgments ($r_{adj} = -0.54$). Consistent with this pattern, Negative Emotions were positively correlated with Negative Judgments ($r_{adj} = 0.63$), and Negative Judgments were negatively correlated with Informative Meaning Making ($r_{adj} = -0.34$). Surprisingly, Controllable Attributions had a significant, positive correlation with Uncontrollable Attributions ($r_{adj} = 0.42$).

Emotions, Judgments, Meaning Making, Attributions, and Decision-Making

Holm-Bonferroni adjusted Spearman's correlations between responses to feedback and decision-making suggest that positive emotions, positive and negative judgments, and informative
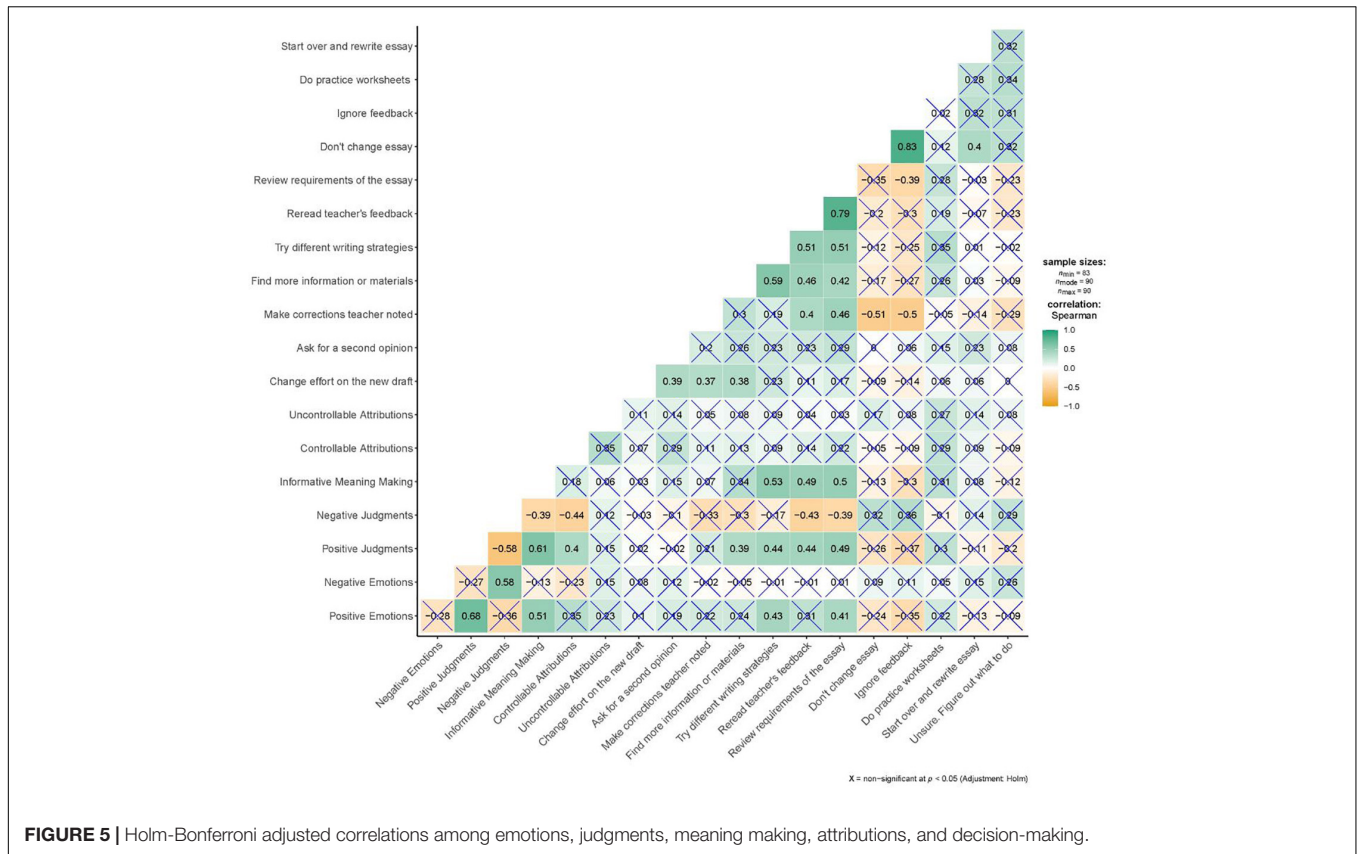
Note. This illustrates the distribution of responses for each of the 12 DM items, with green (right side of the dotted line) representing the percentage of students selecting *agree* (4) to *strongly agree* (6), and gold (left side of the dotted line) representing the percentage of students selecting *disagree* (3) to *strongly disagree* (1).

FIGURE 3 | Stacked barplots of the decision-making (DM) items. This illustrates the distribution of responses for each of the 12 DM items, with green (right side of the dotted line) representing the percentage of students selecting agree (4) to strongly agree (6), and gold (left side of the dotted line) representing the percentage of students selecting disagree (3) to strongly disagree (1).

FIGURE 4 | Holm-Bonferroni adjusted correlations among emotions, judgments, meaning making, and attributions.

FIGURE 5 | Holm-Bonferroni adjusted correlations among emotions, judgments, meaning making, attributions, and decision-making.

meaning making about the feedback all have significant relationships with certain decisions students report about next steps, particularly constructive next steps (**Figure 5**). For example, (1) finding more information to include in the next draft was positively correlated with Positive Judgments ($r_{adj} = 0.39$); and (2) trying different writing strategies was positively correlated with Positive Emotions ($r_{adj} = 0.43$), Positive Judgments ($r_{adj} = 0.44$), and Informative Meaning Making ($r_{adj} = 0.53$). Furthermore, (3) rereading the feedback was positively correlated with Positive Judgments ($r_{adj} = 0.44$) and Informative Meaning Making ($r_{adj} = 0.49$), and negatively correlated with Negative Judgments ($r_{adj} = -0.43$); and (4) reviewing the requirements was positively correlated with Positive Emotions ($r_{adj} = 0.41$), Positive Judgments ($r_{adj} = 0.49$), and Informative Meaning Making ($r_{adj} = 0.50$), and negatively correlated with Negative Judgments ($r_{adj} = -0.39$).

Responses to Feedback and the Motivational States

Summary statistics for students' motivational states are in **Table 6**, which revealed that, on average, students tended toward a growth mindset and reported a stronger orientation toward mastery than performance and avoidant orientations. They also claimed to be moderately self-efficacious about their writing and do not have strong values for or against the writing task at hand.

Pearson's correlations with Holm-Bonferroni adjustments between variables of motivational states and responses to feedback indicated that Task Value was the only motivational

state that resulted in statistically significant relationships with responses to feedback (**Figure 6**). That is, Task Value was positively correlated with Positive Emotions ($r_{adj} = 0.43$), Positive Judgments ($r_{adj} = 0.52$), and Informative Meaning Making ($r_{adj} = 0.39$), and negatively correlated with Negative Judgments ($r_{adj} = -0.38$). Neither self-efficacy, mindset, nor goal orientation correlated with responses to feedback.

The Motivational States and Decision-Making

Spearman's correlations between variables of motivational states and decision-making about next steps are in **Figure 7**. While the correlations trended in the expected direction (i.e., positive correlations between adaptive next steps and mindset, performance and mastery orientation, self-efficacy, and task value), none of these relationships were statistically significant.

TABLE 6 | Summary statistics of students' motivational states.

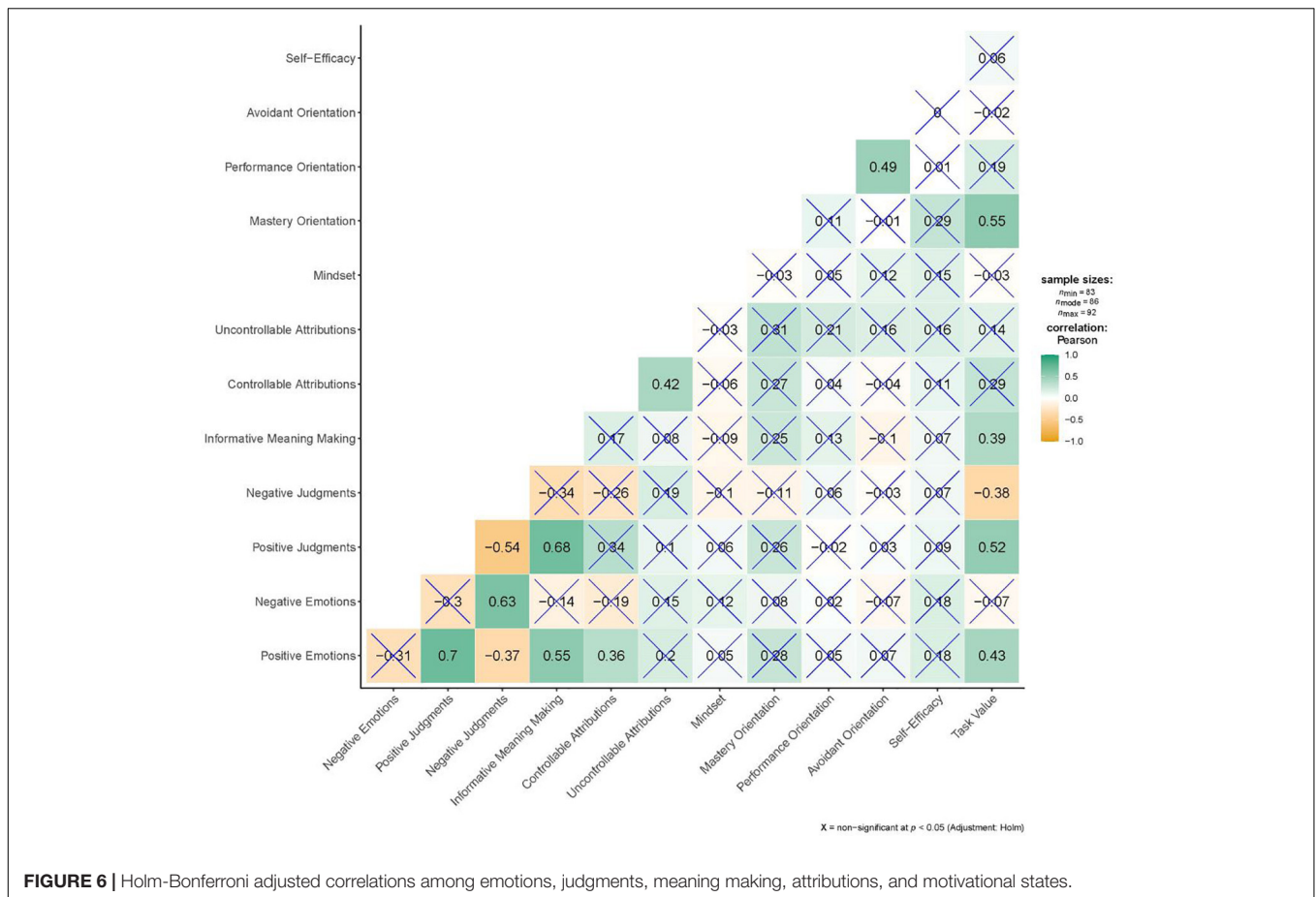|	N	Mean (SD)	Median	Min	Max
Mindset*	93	4.49 (0.90)	4.67	2.00	6.00
Mastery orientation	92	4.76 (0.86)	4.80	2.00	6.00
Performance orientation	89	2.76 (1.12)	2.80	1.00	6.00
Avoidant orientation	92	3.45 (1.15)	3.50	1.00	6.00
Self-efficacy	92	79.43 (15.54)	83.55	16.91	99.1
Task value	90	3.00 (1.09)	3.17	1.00	5.00

*Three items were reverse scored.

FIGURE 6 | Holm-Bonferroni adjusted correlations among emotions, judgments, meaning making, attributions, and motivational states.

Correlations Between Types of Feedback Messages and Responses to Feedback

When the total number of feedback instances was considered, there were significant negative correlations between the number of feedback instances and students' decision to reread teacher's feedback ($r_{unadj} = -0.25$, $p = 0.01$), and review requirements of the essay ($r_{unadj} = -0.27$, $p = 0.01$). With Holm-Bonferroni adjustments, these were no longer statistically significant.

Relationships between types of feedback messages, responses to feedback, and decision-making were also examined. As shown in **Figure 8**, there were no statistically significant relationships between feedback messages and responses to feedback.

There were also no statistically significant correlations between types of feedback messages and decision-making (**Figure 9**).
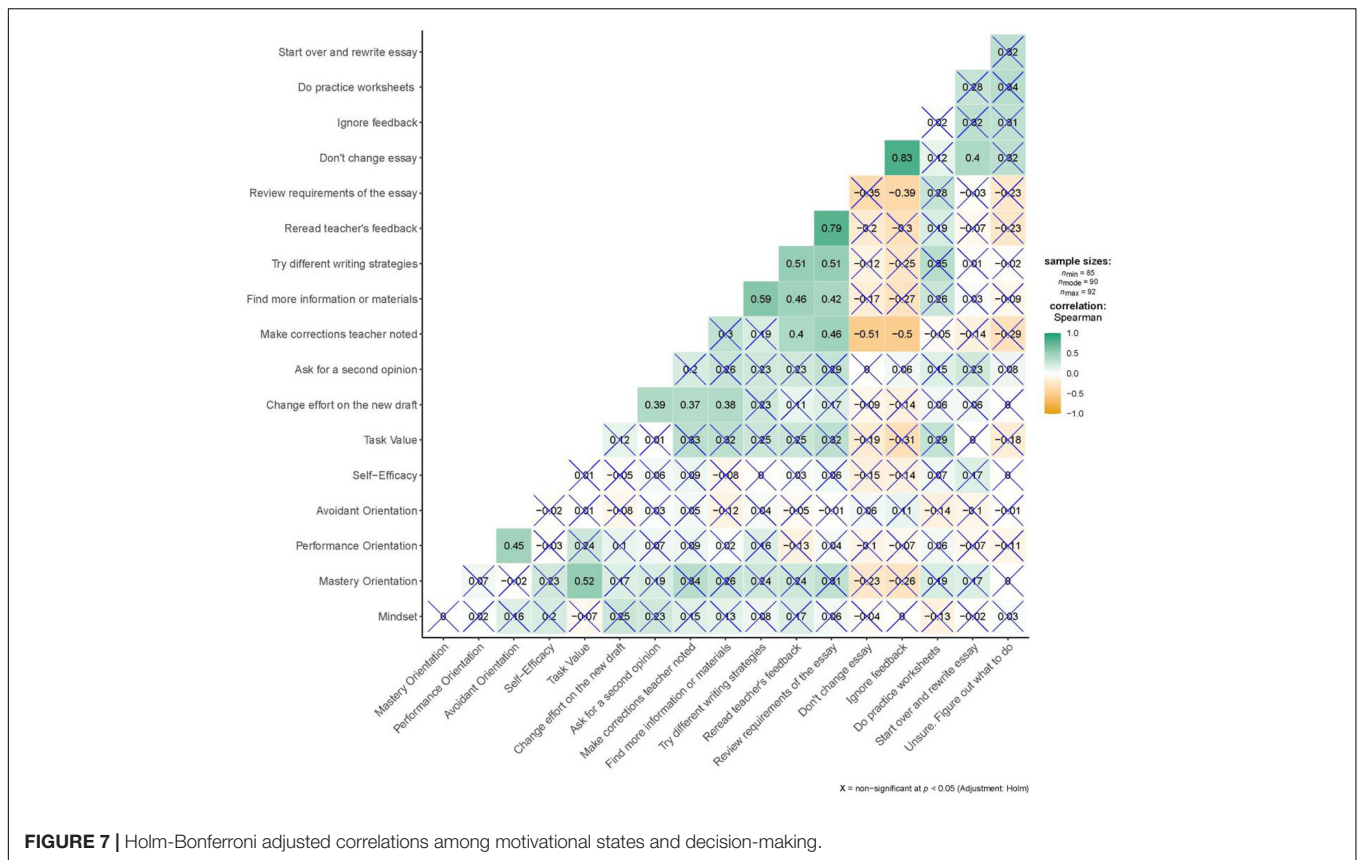
DISCUSSION

The purposes of this study were to explore our hypotheses about the detailed components of responses to feedback, as well as the relationships represented by the dotted lines in the model of students' internal mechanisms of feedback processing (**Figure 1**) in the context of a secondary ELA class in which formative assessment processes were implemented. The model illustrates hypothesized cognitive and affective processes that play a role in how students respond to feedback. These include Emotions, Judgments, Meaning Making, Attributions, and Decision-making, each of which were measured using the RtF survey and examined through think alouds. As part of a mixed methods design, 93 students participated in the survey component of the study, and six students participated in think alouds. Data were collected on students' responses to feedback upon their first exposure to their teacher's written and checklist-based feedback on a writing assessment about a month after they submitted the draft of their essays.

A 7th grade ELA teacher and her students were recruited for this study. The teacher had been identified as an effective teacher who implements formative assessment practices into her usual class routines. Five months of classroom visits by the first author confirmed that her instruction was student-centered and project-based, with formative and summative assessment practices seamlessly embedded into routine classroom practices.

Students' Responses to Feedback

Overall, the components of the model were supported, and one additional component (reflection on/judgments of one's own work) was revealed. Not surprisingly, given the context of

FIGURE 7 | Holm-Bonferroni adjusted correlations among motivational states and decision-making.

the study, in which receiving descriptive, formative feedback, typically accompanied by clear criteria was a routine matter, students' emotions tended to be slightly positive. Considering the number of corrective feedback students received for the assignment, it is also understandable that nearly half of the students felt concerned, unsure, and disappointed after reading the feedback. At the same time, the feedback was primarily informative, descriptive, and not evaluative—characteristics of effective feedback (Nicol and MacFarlane-Dick, 2006; Brookhart, 2007/2008; Wiggins, 2012)—which is likely why the other half of the students felt interested, calm, and hopeful.
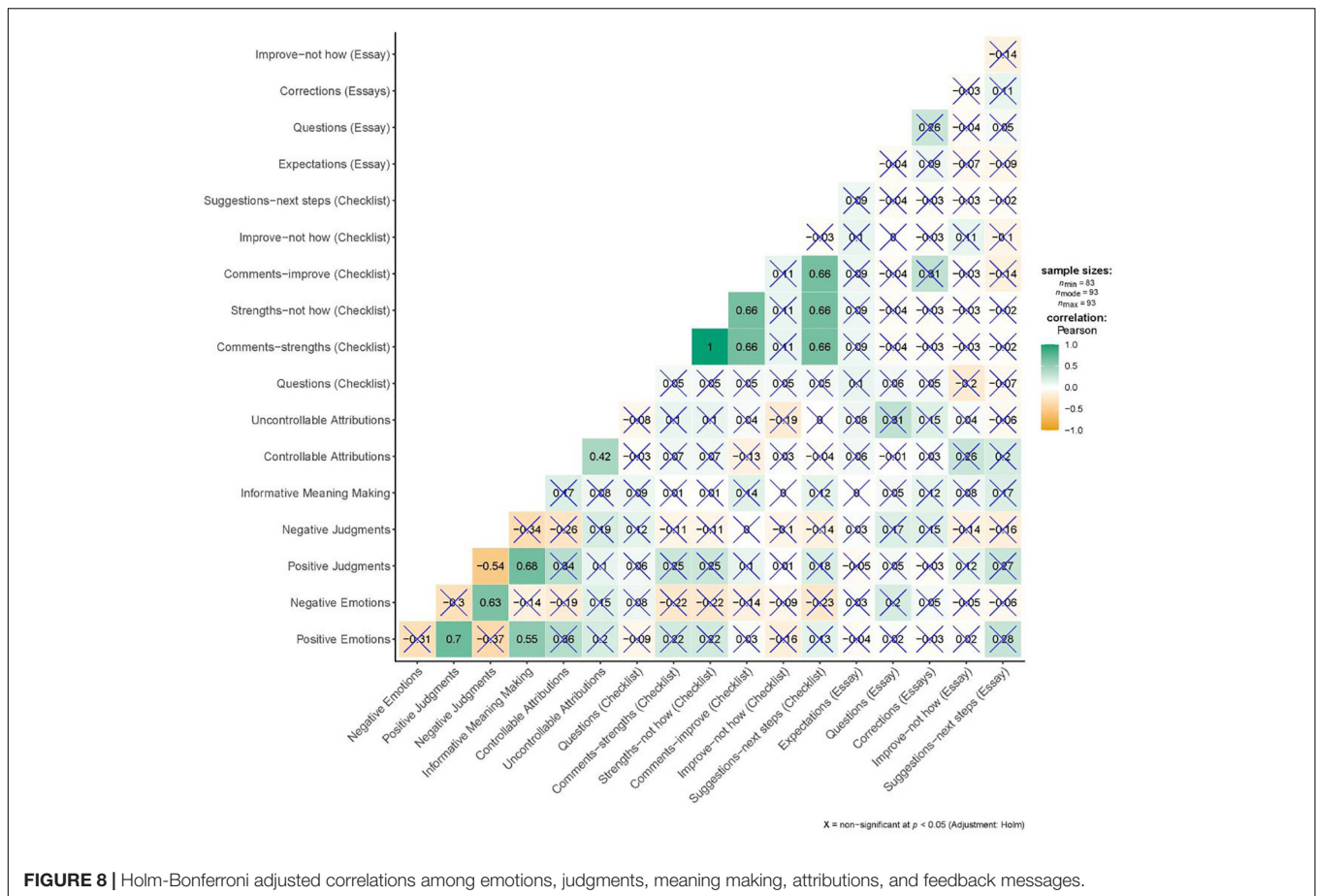
Results indicate that, in addition to judging the feedback as correct and helpful, students also tended to agree with positive judgments that are descriptive of effective formative feedback, such as informative, respectful, clear, and specific. Only a small percentage of students made negative judgments about the feedback (e.g., vague, unpleasant, confusing). These findings echo the literature on students' positive perceptions of effective formative feedback (Lipnevich and Smith, 2009b; Gamlem and Smith, 2013); thus, it is predictable that many students judged the feedback provided by the teacher as informative, respectful, clear, and specific.

Meaning Making was operationalized on the survey as either informative or uninformative, but only Informative Meaning Making was examined for this study because Uninformative Meaning Making subscale resulted in poor item and scale statistics. Our future research will include a revision to this

subscale. The results from this study indicate that students tended to make informative meaning of the feedback. For example, students tended to agree that the feedback provided them with information, in the form of comments and questions, about expectations of the assignment, areas of growth, and suggestions for improvement. Think aloud data also suggested that students interpret feedback by explaining their original work, the thought processes that went into it, and what the feedback means in the context of this work. This finding reveals that students tended to make meaning of the feedback, particularly in terms of what they referred to as their mistakes.

Regarding attributions, Cauley and McMillan (2010) explained that they can be cued by subtle messages that feedback sends to students. The routine exposure to formative feedback with opportunities to use it was likely to have cued students in this study to hold generally controllable attributions, agreeing most with attributions related to their understanding of the content and requirements of the assignment and strategies they used to complete the assignment.

A surprising attribution-related finding was that more than half of the students agreed that they received the feedback they did because of "what my teacher thinks about my understanding of the content" and "what my teacher thinks about my understanding of the assignment." These items were designed as uncontrollable attributions, so the results for these items contrast with the other Uncontrollable Attribution items, with which more than 50% of students disagreed. An obvious explanation

FIGURE 8 | Holm-Bonferroni adjusted correlations among emotions, judgments, meaning making, attributions, and feedback messages.

is that the wording of these two items is convoluted and could have been interpreted as something controllable: "Changing my understanding of the content and/or the assignment will change the feedback I receive" and "I can change my understanding." This also explains the peculiar positive correlation ($r = 0.42$) between controllable and uncontrollable attributions. The two items have been flagged for revision.

The finding that students generally planned to take adaptive or constructive next steps is consistent with theory (Butler and Winne, 1995; Draper, 2009; Lipnevich et al., 2016): because the feedback is generally clear and specific and students tend to agree with it, adaptive problem-solving strategies (e.g., rereading the feedback, finding more information) were activated. However, it is intriguing that although the teacher's feedback was generally thought to be informative, respectful, clear, and specific, 14 students (15%) were still *somewhat* to *very unsure* of what they would do next. An investigation of these 14 students showed that they also tended to judge the feedback more negatively than those who were more certain of what to do, particularly judging the feedback as confusing, unpleasant, and/or vague. They also tended to have slightly more negative and less positive emotions than the other students.

The think aloud data provide some insights into this result: students were, at times, uncertain about some of the feedback, and this uncertainty became a barrier to their decision-making.

For example, one student admitted, "I don't know how to do that, but I will figure it out," and "And then, she put an arrow there for something...?" It is important to reiterate that the think aloud and survey were conducted during students' first exposure to the feedback on an essay they had written a month earlier; therefore, there might be feedback that required more time to understand before decisions about next steps could be made. It could also be that students were questioning the accuracy of the feedback and needed time to process inconsistencies between their thinking and the feedback. For example, one student said, "I don't understand why. I don't know why I need to capitalize the *t*. I thought it was fine as a lowercase." These interpretations foreshadow the following discussion of relationships that were found among responses to feedback processes, in that Positive Judgments about the feedback were positively correlated with constructive next steps. Negative Judgments were negatively correlated with the decision to reread the feedback.

Relationships Among Responses to Feedback, Motivational States, and Feedback Messages

The model of the internal mechanisms of feedback processing (**Figure 1**) also depicts hypothesized relationships among types of feedback messages, initial and motivational states, responses
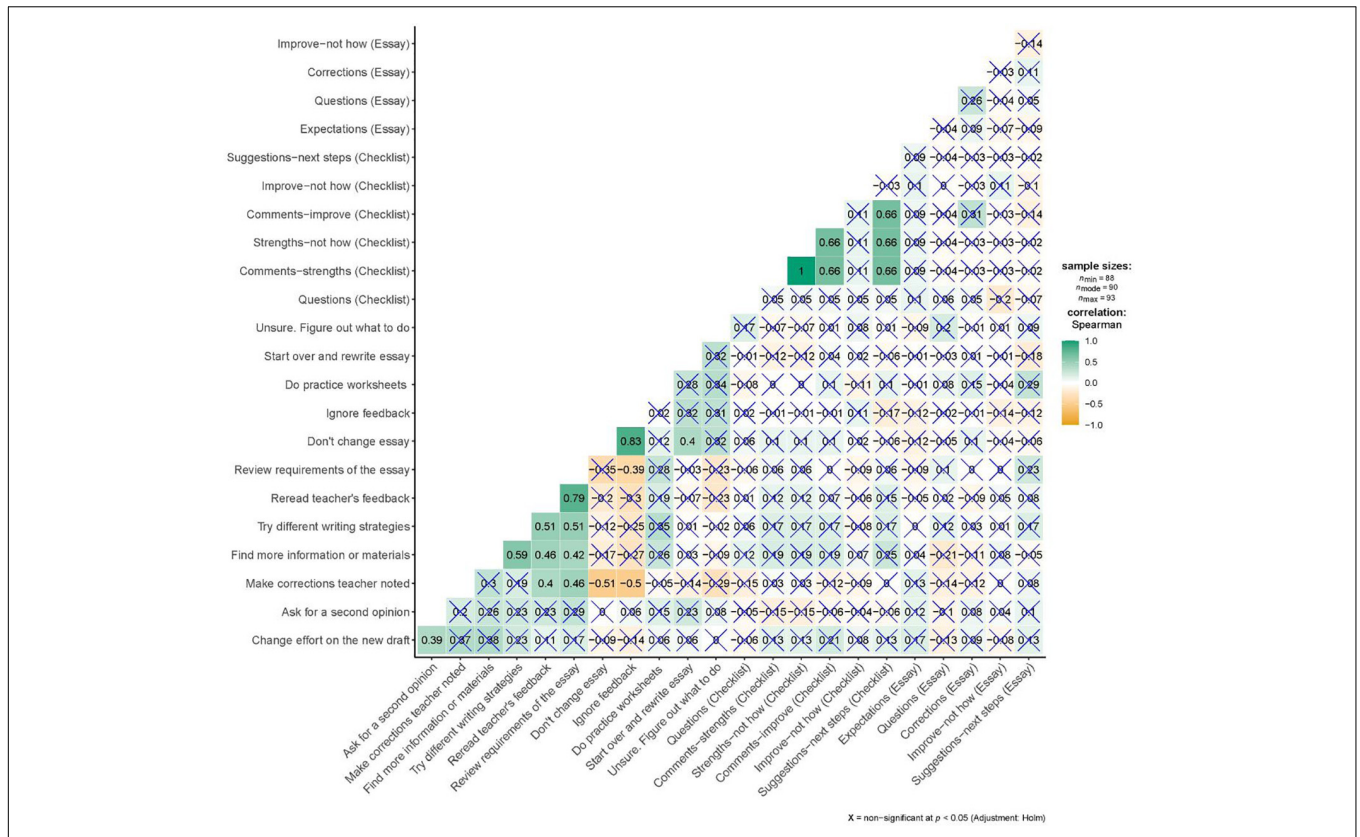
FIGURE 9 | Holm-Bonferroni adjusted correlations among decision-making and feedback messages.

to feedback, behavioral responses, and academic achievement. For this study, we examined the relationships among responses to feedback (i.e., Emotions, Judgments, Meaning Making, Attributions, and Decision-making), motivational states, and external feedback. Correlational findings revealed expected and unexpected results.

Relationships Among Responses to Feedback Processes

As expected, the emotions elicited by the feedback and students' judgments about the feedback were strongly related to each other. Positive Emotions and Judgments were also positively related to Informative Meaning Making and Controllable Attributions (**Figure 4**). These positive relationships among Emotions, Judgments, Meaning Making, and Attributions are consistent with findings from qualitative studies on students' emotions and perceptions about feedback (e.g., Sargeant et al., 2008; Lipnevich and Smith, 2009b; Gamlem and Smith, 2013). Furthermore, this study revealed that students' Judgments and Meaning Making about the feedback might be more explanatory of their decisions about next steps than emotions and attributions (**Figure 5**). Compared to Positive Emotions and Attributions, Positive Judgments and Meaning Making were strongly correlated with more adaptive Decision-making, including the plans to review the requirements, reread the feedback, try different strategies. Beyond these, Positive

Judgments were also associated with students' decision to find more information or materials.

Additional support for the conjecture that students' judgments of the accuracy, trustworthiness, usefulness, and clarity of their teacher's feedback strongly correlated with their decisions about next steps is found in the results regarding negative judgments. Negative judgments about the feedback were negatively correlated with students' plans to reread the feedback and review the requirements of the task. This finding supports Draper (2009) claim that students might choose to ignore feedback if they judged it inaccurate or irrelevant. This finding also echoes Lerner and Keltner (2000), who theorized that negative emotions and judgments could lead to maladaptive decisions: In this study, students with negative judgments tended to be less likely to report plans to reread their feedback and review the task requirements.

Relationships Between Motivational States and Responses to Feedback

Motivational states are known influences on how students respond to feedback (Bangert-Drowns et al., 1991; Butler and Winne, 1995; Bonner and Chen, 2019) and are postulated in our theoretical framework (**Figure 1B**). Therefore, it was unexpected to find no statistically significant relationships between responses to feedback and students' self-efficacy, mindset, or goal orientation after making Holm-Bonferroni

adjustments (**Figures 6**, **7**). Instead, statistically significant correlations between task value and the responses to feedback domains suggest that task value might be more critical than self-efficacy, mindset, and goal orientation in predicting these students' judgments, emotions, meaning making, and decision-making; at least for students who are regularly exposed to formative assessment and feedback practices.

The lack of significant relationships between students' responses to feedback and motivational states is perplexing, nonetheless, because this finding challenges the prevailing assumption that these motivational states influence the direction that students take, including how they "allocate their efforts, set goals, plan for study time, and seek or don't seek help if needed" (Bonner and Chen, 2019, p. 21). This is also contrary to findings from Wingate (2010), which suggested a positive relationship between self-efficacy for writing and emotions, and a study by Brown et al. (2016), which indicated that students' judgments and acceptance of the feedback were positively correlated with students' behaviors, self-efficacy, and performance. We propose several reasons for the null findings from this study, each of which warrants future investigation.

The first explanation for the lack of relationship between responses to feedback and motivational states points to the distinction between task-specific and domain-general constructs. In this study, students' goal orientation and mindset were measured as domain-general constructs, the way they are typically construed in the literature (Payne et al., 2007; Cellar et al., 2011; Gunderson et al., 2017; Burgoyne et al., 2020). In contrast, students' emotions, judgment, meaning making, attribution, and decision-making, as well as task value, were all measured as task-specific processes: students were asked to think about their essay on U.S. Presidents after the Revolutionary War as they completed the RtF survey and task value items. Furthermore, self-efficacy was domain-specific, but not task-specific: students were asked to think about their self-efficacy for writing, but not about their essay on U.S. Presidents after the Revolutionary War. It could be that treating mindset, goal orientation, and self-efficacy as task-specific motivational constructs might have produced stronger associations with responses to feedback.

The second explanation for the null relationships between responses to feedback and motivational states relates to the research design. With the assumption that the motivational states were stable and unlikely to change, data on self-efficacy, mindset, and goal orientation were collected one week before students' essays with teacher feedback were returned to minimize survey fatigue. However, the body of research on feedback and motivational states suggests that students' mindset, goal orientation, and self-efficacy are malleable and can be influenced by feedback (VandeWalle et al., 2001; Cimpian et al., 2007; Chan and Lam, 2010; Hier and Mahony, 2018). This raises the question of whether students' mindset, goal orientation, and self-efficacy might have changed in the week between when data on the motivational states were collected and when their essays with feedback were returned, and the RtF survey was administered. Perhaps these motivational states are not as stable as was assumed.

Finally, and perhaps optimistically, the null relationships between responses to feedback and motivational states could be explained by the restricted range of students' reported motivational states as well as their responses to feedback. That is, the assessment culture and formative assessment practices used by the teacher in this study likely cultivated a willingness and expectation to learn. As a result, students had developed a mastery orientation, growth mindset, and relatively high levels of self-efficacy for writing, as suggested in **Table 6**, albeit with several outliers. Furthermore, the established culture of critique offered a learning environment for students to appreciate and generally respond positively to feedback, especially because they were given the opportunity to use it.

Relationships Between Responses to Feedback and Feedback Messages

Another intriguing but non-significant finding alludes to the concept of too much feedback (Wiggins, 2012). When the overall number of feedback instances was considered, there were statistically significant negative correlations between the number of feedback instances and students' decision to reread the teacher's feedback and review the requirements of the essay. Once Type 1 error was adjusted for using the Holm-Bonferroni method, these correlations were no longer statistically significant. Nonetheless, the possibility of too much feedback being counterproductive is worth investigating with a larger sample.

The lack of relationships between responses to feedback and types of feedback messages (**Figures 8**, **9**) is also contrary to our hypotheses but perhaps expected because of the lack of variability in the teacher's feedback across students. Although the counts of feedback instances did vary among student essays, the types of feedback and feedback messages were generally the same across all student essays: there were correctives, as well as comments in the form of symbols, questions, and marks on the margins. Aside from checkmarks on the checklist, only one student received positive feedback or praise on the essay. Future studies are warranted with students of different teachers with different assessment and feedback practices.

Limitations and Future Directions

As this area of research is still in its nascent stages, this study had to be exploratory, with several limitations that warrant caution in the interpretation of results. First, the sample used for this study was small and purposively selected for the teacher's use of formative assessment practices. Replication studies with more classrooms, both with and without formative assessment practices, are warranted before generalizations to other settings can be made. It is, however, hypothesized that students' responses to feedback will differ depending on the context; classroom observations are recommended as a data source for replication studies. Furthermore, while this study examined students' immediate responses to feedback, including their emotions, interpretations, and plans for next steps, their subsequent thoughts, and feelings, as well as their behavioral next steps and academic achievement, were not examined. These limitations offer a range of future directions.

Finally, as the field of formative assessment has matured, it has shifted toward a discipline-specific perspective (Andrade et al., 2019). This perspective recognizes formative assessment as domain dependent, in which process characteristics such as articulating learning goals, feedback, questioning, peer and self-assessment, and agency "need to be instantiated in the substance—the content—of the discipline in question" (Cizek et al., 2019, p. 16). Furthermore, research related to responses to feedback has been primarily conducted in the higher education setting (e.g., Whitington et al., 2004; Weaver, 2006; Lipnevich and Smith, 2009a,b; Robinson et al., 2013; Pitt and Norton, 2017). Domain-specific studies of students' responses to feedback in K-12 contexts are needed.

SCHOLARLY AND PRACTICAL IMPLICATIONS

Despite its limitations, this study has contributed to the body of research that shifts the focus from feedback as something that is given to something that is received, shedding light into the next black box. Using the model of the internal mechanisms of feedback processing and the RtF survey, this study examined students' cognitive and affective processing of feedback that was provided in an environment where formative assessment was seamlessly embedded into classroom instruction. One of the most intriguing findings was the one that challenged the influence of motivational states. That is, self-efficacy, mindset, and goal orientation did not correlate with emotions, judgments, meaning making, attributions, or decision-making. Instead, students' task value was related to how they responded to feedback: the more task-motivated they were, the more positive emotions, judgments, and informative meaning making, and fewer negative judgments students had about the feedback. In addition, judgments of and the meaning that students made about the feedback were most strongly related to decision-making about next steps. Although only speculative at this point, these relationships could be due to the influence of a formative assessment environment.

Scholars of formative assessment assert that a classroom climate of trust, honesty, and mutual respect between teacher and students is essential in how students respond to feedback (Cowie, 2005; Hattie and Timperley, 2007; Gamlem and Smith, 2013; Tierney, 2013). Visits to this 7th grade ELA classroom revealed a positive assessment culture in which formative feedback was provided with opportunities to revise; checklists and rubrics were used to communicate and reiterate criteria and expectations to students; help-seeking was encouraged; and students openly sought help from the teacher, especially when feedback did not align with their thinking.

In addition to the culture of assessment and learning in which feedback was given and received, the nature of the feedback likely played a role in how students responded to it. Lipnevich and Smith (2009b) concluded in their study with college students that the type of feedback that students found most helpful was "just comments, tell me what I did wrong, where I could change it. Just comments and error marks" (p. 365). This is descriptive of the feedback that the teacher in this study provided to all her students. Lipnevich and Smith's study was conducted with students in higher education: Our study suggests that the kinds of feedback that middle school students prefer are similar to college students' preferences.

To respond positively to formative feedback, it is essential to cultivate a positive assessment culture in which formative assessment practices are effectively incorporated. Ekholm et al. (2015) noted that "feedback is one of the most effective interventions instructors can use to improve student writing; however, it is ineffectual if students do not welcome the feedback they receive" (p. 204). Similarly, Jonsson and Panadero (2018) asserted that "feedback needs to be perceived as useful by the students" (p. 546) as an important condition for the productive use of feedback. In this study, students who had more positive feedback judgments tended to report more constructive next steps. Also, students who reported to have more task value tended to have more positive judgments about the feedback. Then, it might be helpful to monitor students' responses to feedback, particularly judgments, and nurture students' task value.

It has been well established in the formative assessment field that feedback is not formative unless there are opportunities for students to use it (Lipnevich et al., 2014; Andrade, 2016; Jonsson and Panadero, 2018). Allowing for opportunities to revise conveys a purpose of the feedback to the students: they can actually use it. Several students outwardly expressed gratitude for this during the think aloud. One student shared,

I'm actually happy that they let us [write the essay], and then they did marks on it and helped us with it so that we can do it again and make it better and have all these help with it. I like how they did that. I feel like a lot of teachers wouldn't do this, and they would just, you know, do your essay, revise it yourself, and write it.

This study demonstrates the importance of opportunities for revision through students' perspectives.

Methodologically, qualitative studies on students' perceptions and emotions about feedback have depended mainly on semi-structured interviews and classroom observations. Although think aloud protocols have been used for decades (Ericsson and Simon, 1998; Schellings et al., 2013), this is among the first studies, to our knowledge, that has used think aloud protocols to capture what students were thinking and how they were feeling about the feedback that they receive upon their first exposure to the feedback. Much of the think aloud was spent on students making sense of what the feedback was trying to tell them. This study took a look at students' interpretation, or meaning making, of the feedback; as Leighton (2019) noted, much more could and should be done in this area, particularly in examining the alignment between students' interpretation of and teachers' intentions for the feedback.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University at Albany—SUNY. Written informed assent to participate in this study was provided by the participants, and consent was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

This manuscript was a part of AL's dissertation. HA and AL's dissertation chair, provided copious support and feedback on the ideas and writing. Both authors contributed equally to the manuscript revision and read and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2022.751549/full#supplementary-material

REFERENCES

Allal, L., and Lopez, L. M. (2005). "Formative assessment of learning: a review of publications in French," in *Formative Assessment: Improving Learning in Secondary Classrooms*, ed. OECD (Paris: OECD), 241–264.

Andrade, H. (2010). "Students as the definitive source of formative assessment: Academic selfassessment and the self-regulation of learning," in *Handbook of Formative Assessment*, eds H. Andrade, and G. Cizek (New York, NY: Routledge), 90–105.

Andrade, H. (2013). "Classroom assessment in the context of learning theory and research," in *SAGE Handbook of Research on Classroom Assessment*, ed. J. H. McMillan (New York, NY: SAGE), 17–34. doi: 10.4135/9781452218649.n2

Andrade, H. (2016). *Classroom Assessment and Learning: A Selective Review of Theory and Research [White paper]*. Washington, DC: National Academy of Sciences.

Andrade, H. L., Bennett, R. E., and Cizek, G. J. (eds) (2019). *Handbook of Formative Assessment in the Disciplines*. Milton Park: Routledge. doi: 10.4324/9781315166933

Andrade, H. L., Wang, X., Du, Y., and Akawi, R. L. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *J. Educ. Res.* 102, 287–302. doi: 10.3200/JOER.102.4.287-302

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., and Morgan, M. (1991). The instructional effect of feedback in test–like events. *Rev. Educ. Res.* 61, 213–238. doi: 10.3102/00346543061002213

Bennett, R. E. (2011). Formative assessment: a critical review. *Assess. Educ. Principles Policy Practice* 18, 5–25. doi: 10.1080/0969594X.2010.513678

Black, P., and Wiliam, D. (1998). Assessment and classroom learning. *Assess. Educ. Principles Policy Practice* 5, 7–74. doi: 10.1080/0969595980050102

Bonner, S. M., and Chen, P. P. (2019). *Systematic Classroom Assessment: An Approach for Learning and Self-regulation*. Milton Park: Routledge Press.

Brookhart, S. M. (2007/2008). Feedback that fits. *Educ. Leadership* 65, 54–59.

Brown, G. T., Peterson, E. R., and Yao, E. S. (2016). Student conceptions of feedback: impact on self-regulation, self-efficacy, and academic achievement. *Br. J. Educ. Psychol.* 86, 606–629. doi: 10.1111/bjep.12126

Bryer, M. J., and Speerschneider, K. (2016). *likert: Analysis and Visualization Likert Items. R Package Version 1.3.5*. Available online at: https://CRAN.R-project.org/package=likert.

Burgoyne, A. P., Hambrick, D. Z., and Macnamara, B. N. (2020). How firm are the foundations of mind-set theory? the claims appear stronger than the evidence. *Assoc. Psychol. Sci.* 31, 258–267. doi: 10.1177/0956797619897588

Butler, D. L., and Winne, P. H. (1995). Feedback and self-regulated learning: a theoretical synthesis. *Rev. Educ. Res.* 65, 245–281. doi: 10.3102/00346543065003245

Cauley, K. M., and McMillan, J. H. (2010). Formative assessment techniques to support student motivation and achievement. *Clear. House: J. Educ. Strategies Issues Ideas* 83, 1–6. doi: 10.1080/00098650903267784

Cellar, D. F., Stuhlmacher, A. F., Young, S. K., Fisher, D. M., Adair, C. K., Haynes, S., et al. (2011). Trait goal orientation, self-regulation, and performance: a meta-analysis. *J. Bus. Psychol.* 26, 467–483. doi: 10.1007/s10869-010-9201-6

Chan, J. C. Y., and Lam, S. F. (2010). Effects of different evaluative feedback on students' self-efficacy in learning. *Instruct. Sci.* 38, 37–58.

Chen, F., Lui, A., Andrade, H., Valle, C., and Mir, H. (2017). Criteria-referenced formative assessment in the arts. *Educ. Assess. Eval. Accountability* 29, 297–314. doi: 10.1080/00220671.2016.1255870

Cimpian, A., Arce, H. M. C., Markman, E. M., and Dweck, C. S. (2007). Subtle linguistic cues affect children's motivation. *Psychol. Sci.* 18, 314–316. doi: 10.1111/j.1467-9280.2007.01896.x

Cizek, G. J., Andrade, H. L., and Bennett, R. E. (2019). "Integrating measurement principles into formative assessment," in *Handbook of Formative Assessment in the Disciplines*, eds H. L. Andrade, R. E. Bennett, and G. J. Cizek (Milton Park: Routledge), 3–19.

Coe, M., Hanita, M., Nishioka, V., and Smiley, R. (2011). *An Investigation of the Impact of the 6+1 Trait Writing Model on Grade 5 Student Writing Achievement (NCEE 2012-4010)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2003). *Applied Multiple Regression/correlation Analysis for the Behavioral Sciences*, 3rd Edn. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cowie, B. (2005). Pupil commentary on assessment for learning. *Curriculum J.* 16, 137–151. doi: 10.1080/09585170500135921

Dann, R. (2014). Assessment as learning: blurring the boundaries of assessment and learning for theory, policy and practice. *Assess. Educ. Principles Policy Practice* 21, 149–166. doi: 10.1080/0969594X.2014.898128

Draper, S. W. (2009). What are learners actually regulating when given feedback? *Br. J. Educ. Technol.* 40, 306–315. 1 doi: 10.1111/j.1467-8535.2008.00930.x

Dweck, C. (2000). *Self-theories: Their Role in Motivation, Personality, and Development*. Hove: Psychology Press.

Dweck, C. (2006). *Mindset: The New Psychology of Success*. New York, NY: Random House.

Ekholm, E., Zumbrunn, S., and Conklin, S. (2015). The relation of college student self-efficacy toward writing and writing self-regulation aptitude: writing feedback perceptions as a mediating variable. *Teach. Higher Educ.: Critical Perspect.* 20, 197–207. doi: 10.1080/13562517.2014.974026

Ericsson, K. A., and Simon, H. A. (1993). *Protocol Analysis*. Cambridge, MA: MIT press.

Ericsson, K. A., and Simon, H. A. (1998). How to study thinking in everyday life: contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind Culture Act.* 5, 178–186. doi: 10.1207/s15327884mca0503_3

Fatima, S., Ali, M., and Saad, M. I. (2021). The effects of students' conceptions of feedback on academic self-efficacy and self-regulation: evidence from higher education in Pakistan. *J. Appl. Res. Higher Educ.* 14, 180–199. doi: 10.1108/JARHE-07-2020-0209

Fisk, S. M. (2017). *A Qualitative Inquiry of Students' and Teachers' Perceptions on Feedback in Three Queensland Secondary Schools. (Publication No. 112506)*. Doctoral dissertation, Brisbane City QLD: Queensland University of Technology.

Gamlem, S. M., and Smith, K. (2013). Students' perceptions of classroom feedback. *Assess. Educ. Principles Policy Practice* 20, 150–169. doi: 10.1080/0969594X.2012.749212

Green, C., and Gilhooly, K. (1996). "Protocol analysis: practical implementation," in *Handbook of Qualitative Research Methods for Psychology and Social Sciences*, ed. J. T. Richardson (Leicester: BPS Books), 55–74.

Gunderson, E. A., Hamdan, N., and Sorhagen, N. S. (2017). Who needs innate ability to succeed in math and literacy? academic-domain-specific theories of intelligence about peers versus adults. *Dev. Psychol.* 53, 1188–1205. doi: 10.1037/dev0000282

Harris, L. R., Brown, G. T., and Harnett, J. A. (2014). Understanding classroom feedback practices: a study of New Zealand student experiences, perceptions, and emotional responses. *Educ. Assess. Eval. Accountability* 26, 107–133. doi: 10.1007/s11092-013-9187-5

Hattie, J. A. C. (2009). *Visible Learning: A Synthesis of 800+ Meta-analyses on Achievement*. Milton Park: Routledge.

Hattie, J. A. C., and Timperley, H. (2007). The power of feedback. *Rev. Educ. Res.* 77, 81–112. doi: 10.3102/003465430298487

Havnes, A., Smith, K., Dysthe, O., and Ludvigsen, K. (2012). Formative assessment and feedback: making learning visible. *Stud. Educ. Eval.* 38, 21–27. doi: 10.1097/ACM.0000000000001922

Hier, B. O., and Mahony, K. E. (2018). The contribution of mastery experiences, performance feedback, and task effort to elementary-aged students' self-efficacy in writing. *School Psychol. Quarterly* 33, 408–418. doi: 10.1037/spq0000226

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Statistics* 6, 65–70. doi: 10.2307/4615733

Jonsson, A., and Panadero, E. (2018). "Facilitating students' active engagement with feedback," in *The Cambridge Handbook of Instructional Feedback*, eds A. A. Lipnevich and J. K. Smith (Cambridge: Cambridge University Press), 531–554.

Kluger, A. N., and DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol. Bull.* 119, 254–284. doi: 10.1037/0033-2909.119.2.254

Kulhavy, R. W. (1977). Feedback in written instruction. *Rev. Educ. Res.* 47, 211–232. doi: 10.3102/00346543047002211

Lee, H., Chung, H. Q., Zhang, Y., Abedi, J., and Warschauer, M. (2020). The effectiveness and features of formative assessment in US K-12 Education: a systematic review. *Appl. Measurement Educ.* 33, 124–140. doi: 10.1080/08957347.2020.1732383

Leighton, J. P. (2019). Students' interpretation of formative assessment feedback: three claims for why we know so little about something so important. *J. Educ. Measurement* 56, 793–814. doi: 10.1111/jedm.12237

Lerner, J. S., and Keltner, D. (2000). Beyond valence: toward a model of emotion-specific influences on judgement and choice. *Cogn. Emot.* 14, 473–493. doi: 10.1080/026999300402763

Lipnevich, A. A., and Smith, J. K. (2009a). Effects of differential feedback on students' examination performance. *J. Exp. Psychol. Appl.* 15, 319–333. doi: 10.1037/a0017841

Lipnevich, A. A., and Smith, J. K. (2009b). "I really need feedback to learn:" students' perspectives on the effectiveness of the differential feedback messages. *Educ. Assess. Eval. Accountability* 21, 347–367. doi: 10.1007/s11092-009-9082-2

Lipnevich, A. A., Berg, D. A. G., and Smith, J. K. (2016). "Toward a model of student response to feedback," in *The Handbook of Human and Social Conditions in Assessment*, eds G. T. L. Brown and L. R. Harris (Milton Park: Routledge), 169–185.

Lipnevich, A. A., McCallen, L. N., Miles, K. P., and Smith, J. K. (2014). Mind the gap! students' use of exemplars and detailed rubrics as formative assessment. *Instruct. Sci.* 42, 539–559. doi: 10.1007/s11251-013-9299-9

Lui, A. M. (2020). *Validity of the Responses to Feedback Survey: Operationalizing and Measuring Students' Cognitive and Affective Responses to Teacher Feedback*. Doctoral dissertation, Albany, NY: University at Albany, State University of New York.

Lui, A., and Andrade, H. (2022). The next black box of formative assessment: a model of the internal mechanisms of feedback processing. *Front. Educ.* 7:751548. doi: 10.3389/feduc.2022.751548

McMillan, J. H. (2016). "Section discussion: student perceptions of assessment," in *Handbook of Social and Human Conditions in Assessment*,

eds G. T. L. Brown and L. R. Harris (Milton Park: Routledge), 221–244.

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., et al. (2000). *Manual for the Patterns of Adaptive Learning Scales*. Ann Arbor, MI: University of Michigan.

Narciss, S. (2013). Designing and evaluating tutoring feedback strategies for digital learning. *Digit. Educ. Rev.* 23, 7–26.

Nicol, D. J., and MacFarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Stud. Higher Educ.* 31, 199–218. doi: 10.1080/03075070600572090

Pajares, F., Hartley, J., and Valiante, G. (2001). Response format in writing self-efficacy assessment: greater discrimination increases prediction. *Measurement Eval. Counseling Dev.* 33, 214–221. doi: 10.1080/07481756.2001.12069012

Panadero, E., and Lipnevich, A. (2021). A review of feedback typologies and models: towards an integrative model of feedback elements. *Educ. Res. Rev.* 35:100416. doi: 10.1016/j.edurev.2021.100416

Patil, I. (2021). *ggstatsplot: "ggplot2" Based Plots with Statistical Details. R Package Version 0.2.0*. Available online at: https://CRAN.R-project.org/package=ggstatsplot.

Payne, S. C., Youngcourt, S. S., and Beaubien, J. M. (2007). A meta-analytic examination of the goal orientation nomological net. *J. Appl. Psychol.* 92, 128–150. doi: 10.1037/0021-9010.92.1.128

Pellegrino, J. W., DiBello, L. V., and Goldman, S. R. (2016). A framework for conceptualizing and evaluating the validity of instructionally relevant assessments. *Educ. Psychol.* 51, 59–81. doi: 10.1080/00461520.2016.1145550

Phelan, J. C., Choi, K., Niemi, D., Vendlinski, T. P., Baker, E. L., and Herman, J. (2012). The effects of POWERSOURCE© assessments on middle-school students' math performance. *Assess. Educ. Principles Policy Practice* 19, 211–230. doi: 10.1080/0969594X.2010.532769

Pitt, E., and Norton, L. (2017). 'Now that's the feedback I want!' students' reactions to feedback on graded work and what they do with it. *Assess. Eval. Higher Educ.* 42, 499–516. doi: 10.1080/02602938.2016.1142500

Rakoczy, K., Harks, B., Klieme, E., Blum, W., and Hochweber, J. (2013). Written feedback in mathematics: mediated by students' perception, moderated by goal orientation. *Learn. Instruction* 27, 63–73. doi: 10.1016/j.learninstruc.2013.03.002

Ramaprasad, A. (1983). On the definition of feedback. *Behav. Sci.* 28, 4–13.

Revelle, W. (2019). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, IL: Northwestern University.

Robinson, S., Pope, D., and Holyoak, L. (2013). Can we meet their expectations? experiences and perceptions of feedback in first year undergraduate students. *Assess. Eval. Higher Educ.* 38, 260–272. doi: 10.1080/02602938.2011.629291

Ruiz-Primo, M. A., and Li, M. (2013). "Examining formative feedback in the classroom context: new research perspectives," in *SAGE Handbook of Research on Classroom Assessment*, ed. J. H. McMillan (New York, NY: SAGE), 215–232.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Sci.* 18, 119–144. doi: 10.1007/BF00117714

Sargeant, J., Mann, K., Sinclair, D., Vleuten Cees, Van Der, and Metsemakers, J. (2008). Understanding the influence of emotions and reflection upon multi-source feedback acceptance and use. *Adv. Health Sci. Educ.* 13, 275–288. doi: 10.1007/s10459-006-9039-x

Schellings, G. L. M., van Hout-Wolters, B. H. A. M., Veenman, M. V. J., and Meijer, J. (2013). Assessing metacognitive activities: the in-depth comparison of a task-specific questionnaire with think-aloud protocols. *Eur. J. Psychol. Educ.* 28, 963–990.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educ. Researcher* 29, 4–14. doi: 10.3102/0013189X029007004

Shute, V. (2008). Focus on formative feedback. *Rev. Educ. Research* 78, 153–189.

Tierney, R. (2013). "Fairness in classroom assessment," in *SAGE Handbook of Research on Classroom Assessment*, ed. J. H. McMillan (Thousand Oaks, CA: SAGE), 125–144.

Tulis, M., Steuer, G., and Dresel, M. (2016). Learning from errors: a model of individual processes. *Frontline Learn. Res.* 4:12–26. doi: 10.14786/flr.v4i2.168

Tunstall, P., and Gipps, C. (1996). Teacher feedback to young children in formative assessment: a typology. *Br. Educ. Res. J.* 22, 389–404. doi: 10.1080/0141192960220402

Van der Kleij, F. M., and Lipnevich, A. A. (2021). Student perceptions of assessment feedback: a critical scoping review and call for research.

Educ. Assess. Eval. Accountability 33, 345–373. doi: 10.1007/s11092-020-09331-x

Van der Kleij, F. M., Feskens, R. C. W., and Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: a meta-analysis. *Rev. Educ. Res.* 85, 475–511. doi: 10.3102/0034654314564881

VandeWalle, D., Cron, W. L., and Slocum, J. W. Jr. (2001). The role of goal orientation following performance feedback. *J. Appl. Psychol.* 86, 629–640. doi: 10.1037/0021-9010.86.4.629

Weaver, M. R. (2006). Do students value feedback? student perceptions of tutors' written responses. *Assess. Eval. Higher Educ.* 31, 379–394. doi: 10.1080/02602930500353061

Whitington, V., Glover, A., and Harley, F. (2004). Preservice early childhood students' perceptions of written feedback on their essays. *Early Child Dev. Care* 174, 321–337. doi: 10.1080/0300443032000153390

Wiggins, G. (2012). Seven keys to effective feedback. *Educ. Leadership* 70, 10–16. doi: 10.1371/journal.pone.0248628

Wiliam, D. (2010). "An integrative summary of the research literature and implications for a new theory of formative assessment," in *Handbook of Formative Assessment*, eds H. L. Andrade and G. J. Cizek (Milton Park: Routledge), 18–40.

Wiliam, D. (2011). What is assessment for learning? *Stud. Educ. Eval.* 37, 3–14. doi: 10.1016/j.stueduc.2011.03.001

Wiliam, D. (2013). "Feedback and instructional correctives," in *SAGE Handbook of Research on Classroom Assessment*, ed. J. H. McMillan (New York, NY: SAGE), 197–214.

Wingate, U. (2010). The impact of formative feedback on the development of academic writing. *Assess. Eval. Higher Educ.* 35, 519–533. doi: 10.1080/02602930903512909

Winstone, N. E., Nash, R. A., Parker, M., and Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: a systematic review and a taxonomy of recipience processes. *Educ. Psychol.* 52, 17–37. doi: 10.1080/00461520.2016.1207538

Winstone, N., Bourne, J., Medland, E., Niculescu, I., and Rees, R. (2021). Check the grade, log out": students' engagement with feedback in learning management systems. *Assess. Eval. Higher Educ.* 46, 631–643. doi: 10.1080/02602938.2020.1787331

Wisniewski, B., Zierer, K., and Hattie, J. (2020). The power of feedback revisited: a meta-analysis of educational feedback research. *Front. Psychol.* 10:3087. doi: 10.3389/fpsyg.2019.03087