Should I grade or should I comment: Links among feedback, emotions, and performance

Anastasiya A. Lipnevich [a],[*], Dana Murano [b], Maike Krannich [c], Thomas Goetz [d]

[a] Queens College and the Graduate Center, The City University of New York, New York, USA
[b] The Graduate Center, The City University of New York, New York, USA
[c] Institute of Education, University of Zurich, Zurich, Switzerland
[d] Department of Psychology, University of Vienna, Vienna, Austria

A B S T R A C T

This paper explored links among three key constructs pertinent to student learning: performance feedback received, emotions elicited as a result of such feedback, and student performance on a writing task. A sample of 464 university students (age: $M =18.91$, $SD = 2.51$) were asked to write an essay and then were encouraged to revise it based on feedback presented to them. Path mediation models showed that overall negative affect, as well as discrete negative emotions, mediated the relation between receiving feedback and student performance on the final version of the essay. Furthermore, the direct effect of receiving a numeric score negatively predicted students' performance on an essay exam and positively predicted the experience of negative emotions. The indirect effect was positive, suggesting that the experience of negative emotions may have served as a motivational factor in students' desire to improve performance. More research is needed to further explain this relation and the reciprocal causal role emotions play in different feedback mechanisms and performance.

In instructional settings there are a variety of factors that can influence students' learning and performance. Internal and external factors, such as student cognitive ability, attitudes, personality, and school climate, have been well researched and shown to predict student performance (Anderson, 1982; Poropat, 2009; Roth et al., 2015; Stevenson & Newman, 1986). Receiving feedback is undoubtedly a common occurrence in a typical classroom, and performance feedback has been consistently linked to student performance on a task (e.g., Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Lipnevich & Smith, 2018; Shute, 2008). Obviously, feedback can elicit a broad range of both positive and negative emotions, which have also been shown to impact student academic outcomes (e.g., Goetz et al., 2018; Goetz & Hall, 2013, 2020; Ma, 1999; Pekrun, 2006; Zeidner, 2007). There is evidence suggesting that differential feedback elicits a range of variable emotional responses in learners (Lipnevich, Smith, 2009a, 2009b; Peterson et al., 2015; Vogl & Pekrun, 2016), however, the exact mechanisms and links among specific types of feedback and emotions remain to be further examined.

The effect of feedback on emotions can be direct (e.g., positive achievement feedback may elicit enjoyment) or indirect (e.g., feedback can result in high control cognitions that evoke pride) and emotions, in turn, may (indirectly, e.g., via the use of learning strategies) affect learning behaviors and performance outcomes. Interestingly, extensive literature reviews on feedback and various conceptualizations of academic emotions do not typically discuss feedback as an antecedent of emotions, so studies of these relations are quite scarce. The notable exception is Pekrun's control-value theory (CVT) that describes consequences and antecedents of academic emotions. The theory explicitly mentions feedback as a key antecedent of emotions, with emotions subsequently affecting performance outcomes.

According to the CVT (Pekrun, 2006, 2018), teacher feedback about student performance is a powerful precursor of student emotions. This contingency is mediated by students' appraisals of control (e.g., "Can I deal with the feedback given to me?") and value of the feedback ("How important is the feedback to me?"). Further, scholars agree that emotions affect students' achievement, and this link, in turn, may be mediated by how students' deal with the feedback offered to them. The latter combines students' motivation to engage with feedback and availability of self-regulatory strategies needed to effectively process feedback (Black & Wiliam, 1998; Lipnevich et al., 2014). Thus, for example, a cumulative failure feedback that a student receives may undermine his or her sense of control, thus contributing to the development of

---

achievement-related anxiety and hopelessness, low motivation to engage with feedback, and, finally, low achievement.

Similarly to the CVT, Lipnevich et al. (2016) proposed that differential feedback would relate to varying affective and cognitive responses from students. This contingency would inevitably translate into a specific action and, consequently, hinder or enhance performance. That is, a student may experience dismay, joy, pride, or embarrassment; may worry about how his or her parents would react, or be anxious about consequences of feedback on his or her future attainment in class. These affective reactions may lead to students' varied performance on a task. In general, these theoretical frameworks indicate that emotions should mediate the relation between feedback type and performance on a task. To our knowledge, no studies have systematically examined the links between teacher feedback, the achievement emotions they engender, and academic performance. The current study attempted to bridge this chasm and examined the mechanism through which differential feedback provided by teachers (i.e., grades, comments) linked to academic achievement via emotions in the context of a controlled experimental study conducted in a real class setting.

1. Feedback, emotions, and performance

Research has consistently shown that feedback is a key factor in student academic attainment, and when well-constructed, can lead to improved performance on a task. Hattie and Timperley (2007) found that feedback had a large overall effect size on student performance ($d = 0.79$), but that heterogeneity existed depending on the type of feedback students received (e.g., whether in the form of praise, extrinsic rewards, or of task-level feedback). Kluger and DeNisi (1996) also found a great deal of heterogeneity in the effects of feedback; 32% of the feedback effects included in their meta-analysis feedback were negative, with many of the negative effects stemming from studies involving student praise. Hattie and Timperley (2007) recommend that the most effective feedback bridged the gap between where a student currently is, what their intended goal state was, and what steps they needed to take to get from the current state to the intended state. Shute (2008) echoes this recommendation by claiming that feedback should help to reduce uncertainty between performance and goals, in addition to being supportive, timely, non-evaluative, and specific. A large experimental study showed that students who received detailed written comments as a form of feedback outperformed peers who received grades only or comments and grades combined (Lipnevich, Smith, 2009a, 2009b). This supports recommendations made by both Hattie and Timperley (2007) and Shute (2008) that feedback should guide students in improving their work, rather than act solely to communicate summative evaluation.

1.1. Emotions relating to feedback and performance

Emotions are undoubtedly affected in receiving feedback (Goetz et al., 2018; Jonsson & Panadero, 2018). This makes intuitive sense; a student receiving a piece of evaluative feedback will naturally experience some sort of emotional reaction that can vary depending on stakes of the feedback situation, relationship with the feedback-giver, and a variety of other factors (e.g., Lipnevich & Smith, 2018; Panadero, 2016; van Loon & van de Pol, 2019). Both theoretically and empirically, there is some evidence describing direct relations between feedback and emotions. In their qualitative investigations, Brown and Wang (2013) and Harris and Brown (2009) showed that assessment in general and evaluative feedback in particular tended to elicit negative emotions, especially among older students. Emotions, elicited by feedback may also accumulate over time (Pekrun, 2006). So, positive feedback after an exam may induce pride and enjoyment and enhance motivation to do well on the next task. Similarly, failure feedback may induce anxiety and fuel attempts to do better on the next task – or discourage a student from participation (Vogl & Pekrun, 2016).

Goetz and Hall (2013, 2020) suggested that the valence of both feedback and emotions defined the relation between the two constructs. That is, positive feedback would tend to elicit positive emotions, whereas negative feedback would elicit negative emotions (Goetz & Hall, 2013, 2020). Though this relationship is theoretically supported, very little causal evidence exists upon which to support the relationship (Goetz et al., 2018). Further, when feedback is not clearly positive or negative in nature, things become even more muddled. For example, a numerical score or a grade does not inherently carry any valence, be it positive or negative. However, depending on the student, context, type of task, and a number of other variables, grade becomes one of the most emotionally charged pieces of feedback a student receives. So, a score of 85/100 may be received as overwhelmingly positive by one student, and detrimentally negative by another. Students' prior performance, receptivity, feedback-seeking inclinations, teacher characteristics, and subject domain may all result in differential receptivity of grades (see Evans & Waring, 2011; Fong et al., 2016; Lipnevich et al., 2016).

There are volumes of research evidence revealing a strong affective component in students' responses to feedback. For example, Van der Kleij and Lipnevich (2020), showed that out of 164 studies examining student perceptions of feedback, twenty two reported findings that related to students' emotional reactions. Unfortunately, this comprehensive review demonstrated that hardly any of the included studies attempted to link student perceptions and affective responses to feedback to meaningful educational outcomes, such as indicators of performance, course completion, and general indices of well-being. It is, of course, important to know that grades, high or low, make students less likely to experience positive affect and be motivated to carry out revisions (Lipnevich, Smith, 2009a, 2009b). It is even more important to find out how these affective responses link to student performance on a task. In other words, more research is needed to establish the relation between these two constructs; in particular, understanding how different forms of feedback (e.g., written comments, grades, summative feedback versus formative feedback) affect emotional responses in students can be particularly useful, with subsequent examination of links to performance outcomes (Goetz et al., 2018).

CVT provides a strong framework for establishing such relations. According to the theory, feedback may serve as an antecedent of control and value appraisals, which, in turn would instigate different emotions (enjoyment and pride, as positive, activating; anger and anxiety, as negative, activating; relief, as positive, deactivating; and hopelessness, as negative, deactivating). Further, emotions will have a bearing upon academic performance (mainly via motivation, use of learning strategies, and activation of cognitive resources, e.g., Pekrun, 2006). In terms of the latter link, the average relation between discrete positive or negative emotions and performance is about $|0.25|$ (Goetz & Hall, 2013, 2020), an effect size that may be deemed small to medium in classical interpretations (Cohen, 1992) but can be considered meaningful in the realm of educational interventions (Hattie et al., 1996). In particular, due to omnipresence of feedback situations and emotions experienced in academic settings, cumulative effects, even if weak, may have strong effects on performance and other meaningful outcomes (Westphal et al., 2018). In particular, anxiety, a negative activating emotion, shows consistent negative correlations with performance across various academic domains (Ma, 1999; Seipp, 1991).

Goetz et al. (2018) discussed relations among emotions and feedback, presenting a comprehensive overview of moderators and mediators of these links, and call researchers for empirical studies investigating these relations. In the current paper we considered differential feedback and its effect on emotions and subsequent performance on a task. Thus, guided by Pekrun's CVT, Lipnevich et al.'s (2016) model, as well as Goetz et al.'s (2018) representation of relations among feedback, emotions, and performance, we hypothesized that emotions would mediate the effect of feedback on performance.

2. Current study

The purpose of this study is to explore relations among different types of feedback, emotions elicited, and performance, as well as to test the hypothesized mediational role of emotions between feedback and performance. Specifically, the current study explored how different forms of feedback (i.e., in the form of a grade accompanied by written comments or in the form of written comments alone) affected discrete emotions, and, consequently, performance.

The results of this study can help practitioners to design optimal feedback, as well as further our understanding of the role emotions play in performance, both in how emotions are elicited and in how they affect performance. Based on relations described in the literature between feedback and emotions, as well as between emotions and performance, emotion may play a mediational role in the relation between feedback type and performance on an essay task. Hence, we will attempt to answer the following research question:

Do emotions mediate the link between different types of feedback and student performance on an essay?

3. Method

3.1. Participants

Four hundred and sixty four students enrolled in the introductory psychology course at a large public Northeastern university in the USA participated in this study. One of the graded course assignments involved writing an essay, so students were motivated to do well on the task. Informed consent was obtained to use students' written work for research purposes and to administer a series of questionnaires. All students enrolled in the course provided their consent and, hence, agreed to have their responses used for the purposes of our study. APA ethical standards were followed in the conduct of the study, and the study protocol was approved by the institutional review board of the university.

Participants' ages ranged from 17 to 51 ($M = 18.9$, $SD = 2.5$). Two hundred and forty (51.9%) were women, and 223 (48.1%) were men, whereas one participant chose not to report gender. Of the 464 participants, 382 (82.3%) were born in the US, and 82 (17.7%) were not. The majority of the participants self-identified as White (54.7%), followed by 24.6% as Asian, 6.9% as Hispanic/Latino/Latina, 3.9% as Black, and 9.4% as Other. Three hundred and seventy one students (80%) reported to be English-speakers, and 93 (20%) to be native speakers of language other than English. Students represented twenty two different departments and majors.

3.2. Procedure

Data for this study came from a larger study that examined effects of differential forms of feedback on student essay performance (see Lipnevich, Smith, 2009a, for detailed description of the study). The procedure was as follows. Students came to a computer lab and were given the task of writing a 500-word expository essay that was part of class requirements, demonstrating their understanding of theories of motivation discussed in class. The prompt for this assignment was chosen from the Educational Testing Service list of topics used on their proprietary *E*-rater platform appropriate for first-year students. The platform was developed to automatically score student essays. The original topic was:

Sometimes we choose to do things that we do not really enjoy — jogging, studying, eating the right foods, and so on. Describe something you do by choice that you really do not enjoy. Explain why you continue to do it. Discuss the changes that might occur in your life if you were to stop this activity.

This prompt was then modified to incorporate a clear reference to theories of motivation, consistent with the course content. The ETS E-Rater team approved the modification. The resulting prompt was:

Sometimes we choose to do things that we do not really enjoy — studying hard, eating the right foods, and so on. Describe something you do by choice that you really do not enjoy. Using theories of motivation, explain why you might continue to do it. Discuss the changes that might occur in your life if you were to stop this activity. Support your claims with specific examples from your life and the course reading.

Students were presented with an extensive rubric describing criteria for evaluation. The rubric was available during the task and could be consulted at any point in the writing or revising process. In order to make sure that students wrote essays of comparable length, an indicator displayed a real-time word count.

One week after the initial session students returned to the lab. They logged into the system and were shown their essays with corresponding feedback that varied depending on the group to which they had been assigned. The amount of time they spent interacting with the feedback was recorded. Prior to moving to the revision screen, students were prompted to report their currently experienced emotions. The participants were then asked to carry out revisions and resubmit their essay. Students had an option of referring to the grading rubric and to their feedback comments at any point during the session by hovering their mouse over hotspots in the feedback text.

During the week between the two sessions, two experimenters scored essays, randomly assigned students to experimental conditions, and provided feedback depending on a specific condition. The original experiment involved a $2 \times 2 \times 3$ design, with score/no score, praise/no praise, and computer, instructor, and no detailed comments, as independent variables. Praise did not reveal a main effect, and there were no differences in average essay scores between the computer and the instructor conditions, hence, for the purposes of the current study we only examined differences in student emotions and their links to performance on the essay-writing task depending on whether or not they received a numerical score (i.e., grade). Hence, for the purposes of this study we looked at (1) a numerical score that reflected student performance in combination with written comments, specific to individual's work, or as (2) written comments alone with no numerical score.

3.2.1. Dependent measures and scoring

Student scores on the initial draft and the final, revised essay, were calculated as follows. E-rater, a proprietary ETS software was used to score the mechanics, grammar, spelling, and stylistic features, and two experimenters scored content using a previously designed rubric (the inter-rater reliability was 0.96 for the draft score and 0.98 for the final essay). The final score was calculated as a weighted average of the two scores and converted into a scale of 100%. The mechanics score accounted for 30% of the final score, and the content score contributed 70% to the total score. This course was 'writing intensive,' which means that the instructor was required to work on developing student writing and subsequently assess it. Table 1 includes examples of comments that students received.

Students were given ample time to revise their essay. The differences from draft to the final score – an outcome variable that reflects change – were used as the outcome. The numerical score was not converted into the letter grade, however, all students had a conversion chart on the syllabus and – as follow up interviews demonstrated – could easily convert percentages into a corresponding grade. Hence, for the purposes of this paper, we will use the terms "score" and "grade" interchangeably.

3.2.2. Emotions

Immediately after receiving feedback, students responded to a self-report questionnaire gauging their emotions. Nineteen discrete

Table 1
Examples of comments on mechanics and content that students received.

Type of comment	Example
Mechanics	*Name*, please break your essay into paragraphs so I can see the structure. *Name*, this sentence is a fragment. Proofread the sentence to be sure that it has correct punctuation and that it has an independent clause with a complete subject and predicate. *Name*, these sentences begin with coordinating conjunctions. Try to combine the sentence that begins with *but* with the sentence that comes before it.
Content	*Name*, a good essay usually contains three main ideas, each developed in a paragraph. Use examples, explanations, and details to support and extend your main ideas. Try to center them around the theories of motivation I discussed in class. Include details and theory-specific terminology. *Name*, discuss all of the components of Atkinson's theory: expectancy, value and the need for achievement. You are missing one of the components.

emotions were measured using the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988) questionnaire, and scores were calculated both for overall positive and negative affect, and for discrete emotions. The items prompted students to report how they felt *right now*, using a five-point scale, bounded by 'Slightly/Not at all' and 'Extremely.' That is, the scale included momentary, or state, instructions for measuring students' current affective state (e.g., Linnenbrink-Garcia & Pekrun, 2011).

The scales were shown to be highly internally consistent, largely uncorrelated, and structurally stable. Watson et al. (1999), however, argue that the dimensions of PA and NA represent the subjective components of the bio-behavioral systems of approach, known as the behavioral inhibition system (BIS), and withdrawal, referred to as the behavioral engagement system (BES; Crawford & Henry, 2004; Watson et al., 1999). These adaptive systems are viewed as separate, but not entirely independent of one another, thus accounting for the moderate correlations reported between the constructs of PA and NA (Watson et al., 1999). In the current study, two additive indices were computed, resulting in separate PA and NA scores for each participant. Alpha coefficients reported in the literature of the positive affect scale range from 0.86 to 0.95, and the negative affect scale from 0.84 to 0.92 (Crawford & Henry, 2004; Von Humboldt et al., 2017; Ilies & Judge, 2005; Jolly et al., 1994; Roesch, 1998; Lipnevich, Smith, 2009a, 2009b). Alphas in this sample were 0.88 for the positive affect scale and 0.86 for the negative affect scale.

3.2.3. Latency

The software tracked the amount of times students spent reading the feedback provided, and then revising essay based on the feedback provided. We examined this data to determine how much time students spent reading their feedback, how much time students spent revising, and whether there were differences depending on feedback condition.

3.3. Data analysis

First, descriptive statistics were computed. Additionally, Pearson and point-biserial correlations were calculated for all key variables in the study. A multivariate analysis of variance (MANOVA) model was fit to determine if there were differences between various positive and negative emotions based on whether or not the student received a numeric score in addition to written comments on their essay. Multiple regression models were also fit to test the effects of negative and positive affect on the time 2 essay grade, while controlling for scores on the time 1 essay. Then, *t*-tests were conducted to determine if there were mean differences in time spent reading feedback and revising essays based on grade condition. Last, path mediation models were fit to the data with

grade condition entered as an independent variable, emotions considered as possible mediators, and the change scores in students' essay performance from time 1 to time 2 as the dependent variable. Grade condition was entered into the model as a dummy variable with grade = 1, and we controlled for scores on the time 1 essay score. Path models were fit using the lavaan package in R and all other analyses were conducted using SPSS Version 24.

4. Results

4.1. Descriptive statistics

We examined descriptive statistics and correlations between all discrete emotions, positive affect total, negative affect total, and feedback condition (comments with and without grades). Additionally, we included essay score after revisions and difference in essay score between students' initial and revised submission for additional analyses. Table 2 contains descriptive statistics and correlations between all variables of interest, all of which followed a univariate normal distribution. We note high correlations between comments with grades condition and the negative affect scale score, as well as with the difference in student pre- and post-revision scores. Table 3 contains scores for draft and revision for graded and non-graded conditions.

4.2. Multivariate analyses

In addition to examining correlations, we used multivariate analyses to determine whether or not there were differences in each discrete emotion depending on the feedback condition. That is, we checked whether there were differences in final scores for students who received their score and those who did not. We found that receiving a score, as opposed to written comments, increased student ratings on *distressed, upset, scared, hostile, ashamed, nervous,* and *afraid,* all negative emotions. Additionally, we saw a decrease in pride when scores were given compared to written comments. Table 4 shows all results and effect sizes for each discrete emotion between grade conditions. As almost all significant effects were for negative emotions, we focused primarily on these in the following mediation analysis.

To determine the effect of emotions on student essay performance, we performed a multiple regression analysis examining the effects of positive and negative emotions on the difference score on the revised version of the exam. The PANAS positive and negative scores were entered into the model as predictors. After controlling for the Time 1 score, neither positive affect (standardized $\beta = -0.04$, $SE = 0.04$, $p = .36$) nor negative affect (standardized $\beta = 0.08$, $SE = 0.04$, $p = .10$) significantly predicted the difference score in student essays from time 1 to time 2. Time 1 scores were the only significant predictor in the model (standardized $\beta = -0.33$, $p < .05$, $R = 0.35$, $R^2 = 0.12$).

We also examined whether or not feedback condition affected the average time students spent reading their feedback and revising their essay. On average, students spent a total of 3.50 minutes ($SD = 3.93$) reading their feedback. Time spent reading the feedback did not differ between the graded condition ($M = 3.48$, $SD = 3.51$) and the non-graded condition ($M = 3.51$, $SD = 4.32$, $t(461) = 0.10$, $p = .92$). Students who received only written comments, but no numerical grade ($M = 54.64$, $SD = 106.70$) did not spend significantly more time revising their essays than students who received numerical grades ($M = 46.90$, $SD = 79.27$, $t(461) = 0.89$, $p = .38$). Due to non-significant differences in time spent reading and revising essays based on feedback, we did not include this variable in further analyses.

4.3. Mediation analysis

In order to further explore the relation between feedback condition, emotions elicited, and differences in student essay scores, we tested a mediation path model. We hypothesized that emotions played a

Table 2
Correlation matrix of key variables.

	Mean	SD	1	2	3	4	5
1 Grade	0.50	0.50	1				
2 Essay score time 2	79.13	8.62	−0.04	1			
3 Difference in essay scores	4.81	6.27	−0.05	0.38**	1		
4 Positive affect	29.86	7.17	−0.01	−0.02	−0.05	1	
5 Negative affect	24.00	7.51	0.17**	−0.06	0.13**	−0.06	1
Active	2.93	1.05	0.02	−0.08	0.04	0.73**	0.05
Alert	3.14	1.03	0.01	0.01	0.01	0.63**	0.08
Attentive	3.42	1.01	0.05	0.04	−0.03	0.80**	0.03
Determined	3.76	1.04	0.01	0.04	−0.01	0.77**	0.05
Enthusiastic	2.59	1.04	0.00	−0.05	−0.05	0.75**	−0.11*
Excited	2.50	1.04	−0.01	−0.05	−0.08	0.64**	−0.01
Inspired	2.65	1.09	0.01	−0.07	−0.04	0.76**	−0.02
Interested	3.33	0.99	0.01	−0.04	−0.19	0.66**	0.05
Proud	2.54	1.03	−0.13	0.04	−0.13**	0.62**	−0.28**
Strong	2.99	0.94	−0.06	0.04	−0.05	0.62**	−0.26**
Afraid	2.33	1.17	0.12**	−0.10*	0.03	0.03	0.80**
Ashamed	1.95	1.11	0.22**	−0.07	0.10*	−0.06	0.69**
Distressed	3.07	1.16	0.13*	−0.03	0.05	−0.14**	0.69**
Guilty	1.64	0.93	0.09	−0.06	0.08	−0.02	0.56**
Hostile	1.72	0.96	0.10*	−0.04	0.04	−0.19**	0.51**
Irritable	2.70	1.18	0.01	0.01	0.05	−0.14**	0.49**
Jittery	2.57	1.20	0.06	−0.04	0.04	0.16**	0.66**
Nervous	3.04	1.24	0.10*	−0.03	0.06	0.07	0.73**
Scared	2.40	1.17	0.13**	−0.03	0.08	−0.01	0.74**
Upset	2.59	1.19	0.19**	−0.05	0.11*	−0.14**	0.71**

* $p < .05$.
** $p < .01$.

Table 3
Scores for draft and revision for graded and non-graded conditions.

		N	Mean	Std. Deviation	Std. Error
Grade for Draft	no grade	233	74.28	8.278	0.542
	grade	230	74.56	8.288	0.546
	Total	463	74.42	8.275	0.385
Grade for Revision	no grade	233	79.25	9.244	0.606
	grade	230	77.62	8.152	0.538
	Total	463	78.43	8.715	0.405

Table 4
Effects of feedback conditions on discrete emotions.

	No grade M(SD)	Grade M (SD)	F	Partial Eta Squared
Interested	3.32(0.98)	3.34(1.00)	0.08	0.000
Distressed	2.93(1.16)	3.22(1.15)	7.53**	0.016
Excited	2.51(0.99)	2.50(1.08)	0.01	0.000
Upset	2.37(1.11)	2.82(1.23)	16.62***	0.035
Strong	3.04(0.91)	2.93(0.96)	1.55	0.003
Guilty	1.56(0.88)	1.72(0.98)	3.4	0.007
Scared	2.25(1.12)	2.56(1.21)	8.81**	0.017
Hostile	1.62(0.94)	1.81(0.98)	4.58*	0.010
Enthusiastic	2.59(1.01)	2.59(1.07)	0.01	0.000
Proud	2.67(0.99)	2.40(1.06)	8.06**	0.017
Irritable	2.69(1.22)	2.71(1.15)	0.03	0.000
Alert	3.14(1.00)	3.15(1.06)	0.02	0.000
Ashamed	1.70(1.00)	2.19(1.16)	23.34***	0.048
Inspired	2.64(1.06)	2.67(1.13)	0.06	0.000
Nervous	2.90(1.31)	3.16(1.1)	4.97*	0.011
Determined	3.76(1.01)	3.77(1.07)	0.02	0.000
Attentive	3.37(0.99)	3.47(1.02)	1.15	0.002
Jittery	2.50(1.24)	2.64(1.16)	1.5	0.003
Active	2.91(1.03)	2.95(1.08)	0.19	0.000
Afraid	2.18(1.14)	2.47(1.19)	7.16**	0.015

* $p < .05$.
** $p < .01$.
*** $p < .001$.

mediational role in the relation between feedback type and the change in essay score from time 1 to time 2 and computed both direct and indirect effects to test this. Feedback type was set as a manifest dummy variable in the model, with written comments only coded as 0 and a grade accompanying written comments as feedback coded as 1; regression coefficients can therefore be interpreted as the effects of receiving numeric scores and written comments as feedback on improvement in essay score. The negative affect total score from the PANAS questionnaire was used as the manifest negative affect variable, and the difference score between essays at time 1 and time 2 was considered as the outcome variable. The model was recursive and satisfied the df ≥ 0 condition, meeting the identification rule for path models (Kline, 2016). With 6 observations and 6 estimated parameters, the resulting model contained 0 degrees of freedom and was therefore considered saturated; although fit statistics are not appropriate to report for saturated models, estimated direct and indirect effects can be valuable in exploratory research projects such as this. Fig. 1 shows regression coefficients for each path and the standardized effects are as follows: indirect effect = 0.24, $SE = 0.12$, $p = .04$, direct effect = −1.13, $SE = 0.58$, $p = .05$, and total effect = −0.90, $SE = 0.57$, $p = .11$.

Receiving a grade as feedback, as expected, showed a direct increase in negative affect experienced. Additionally, receiving a numeric grade as opposed to written feedback alone has a negative effect on the change in essay scores. This suggests that receiving a grade as opposed to comments alone may not be as effective of a mechanism. The effect of negative emotion on the difference score is positive, albeit small. Although the emotion elicited may be unpleasant, it could potentially serve as a triggering factor in motivating students to want to improve their scores.

Additionally, we conducted exploratory path analyses with three discrete emotions that were highly correlated with feedback condition and differences in student scores (see Table 5). We found a similar pattern with discrete negative emotions as we did with the combined negative affect scale. That is, receiving a grade as feedback instead of written comments alone significantly predicted discrete negative emotions *upset* and *ashamed*. Additionally, we observed negative coefficients (with small effects) between the negative emotions elicited and the time 2 essay score. As expected, receiving a grade also resulted in a decrease
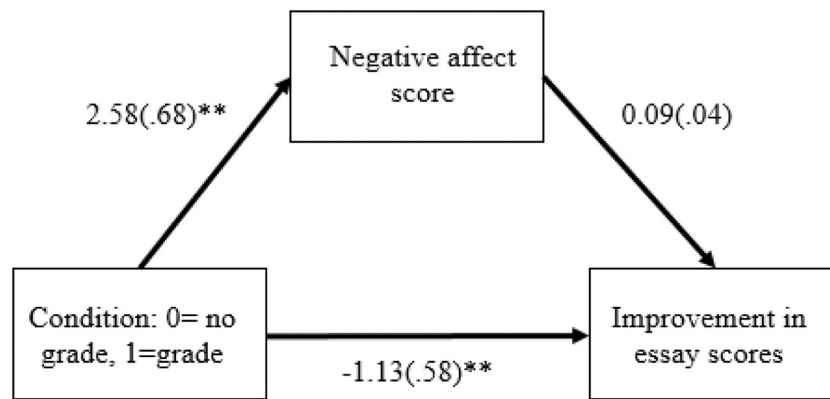
Fig. 1. Negative affect as a mediator of feedback condition and improvement in essay scores from Time 1 to Time 2.

Table 5
Mediator analyses with discrete emotions.

	Upset				Ashamed				Proud			
	b	SE	Std.	z	b	SE	Std.	z	b	SE	Std.	z
Grade condition on score difference	−1.19	0.58	−0.10	−2.05*	−1.22	0.58	−0.10	−2.08*	−1.13	0.57	−0.09	−1.98*
Grade condition on emotion	0.44	0.11	0.19	4.09***	0.49	0.10	0.22	4.45***	−0.27	0.09	−0.13	−2.85**
Emotion on score difference	0.65	0.24	0.13	2.68**	0.65	0.26	0.12	2.49**	−0.87	0.28	−0.15	−3.13**
Indirect effect	0.29	0.13	0.02	2.24*	0.32	0.14	0.03	2.21*	0.23	0.11	0.02	2.10*

* $p < .05$.
** $p < .01$.
*** $p < .001$.

in the experience of pride, a positive emotion.

Furthermore, for changes in score from the draft to revised version of student essays, experiencing pride positively predicted the change directly. Whereas experiencing negative emotions directly increased students' change in essay scores, experiencing pride negatively predicted change in essay scores. That is, experiencing negative emotions may elicit change more so than experiencing positive. Though these effects are small in magnitude, the trend is nonetheless notable.

5. Discussion

This study aimed to examine links among different types of feedback, emotions feedback elicited, and student performance on an essay task. The results showed that there were effects of feedback on student emotions, as well as of emotions experienced on student performance on a task. Receiving grades as a form of feedback, as compared to written comments alone, directly predicted students' experience of negative emotions. We found that receiving a grade, as opposed to written comments alone, was associated with increased student ratings on *distressed, upset, scared, hostile, ashamed, nervous,* and *afraid*, all negative emotions. Additionally, we saw a decrease in pride when grades were given compared to written comments. This suggests that feedback in the form of written comments may be more beneficial to students, echoing recommendations made by feedback researchers in the past (e.g., Hattie & Timperley, 2007; Lipnevich, Smith, 2009a, 2009b; Shute, 2008). Grades have a heavy evaluative and normative component that may result in students' experience of negative emotions, compared to a more neutral, descriptive commentary that offers specific suggestions on how they might improve their work.

Interestingly, our examination of indirect effects revealed that experiencing negative emotions may actually help students to improve their grades on revised essays. According to Pekrun's (2006, 2018) CVT, the effects of negative activating emotions, such as *anxiety, anger,* and *shame,* on achievement are varied. In other words, the effects of these emotions on student performance can be both beneficial (e.g., due to

positive effects on extrinsic motivation) and detrimental (e.g., due to non-optimal strategy use). However, the evidence for the hypothesized link between negative emotions and achievement is still scarce and in need of further exploration. The only notable exception is *anxiety*, with compendiums of studies investigating this emotion and its relations to meaningful educational outcomes.

However, even though here is a lot of empirical evidence demonstrating links between anxiety (especially test anxiety; Zeidner, 2007) and achievement, it still remains controversial whether the anxiety-performance relation follows an inverted U-shaped form for academic performance (Yerkes-Dodson law, e.g., Teigen, 1994). With respect to the negative activating emotions of *upset* and *ashamed*, our study supported the assumption that these emotions may also, at least in the short run, be beneficial for achievement outcomes (see also Goetz & Hall, 2013, 2020; Pekrun et al., 2017). The negative activating emotions elicited as a result of receiving grades as feedback could have served as a short-term motivational stimuli to drive the individual to improve task performance, but the negative impact on performance might only be seen in the long run (e.g., superficial behavior which does not lead to sustainable learning). Future studies are needed to examine long-term effects of feedback on emotions, performance, and indices of well-being. It is possible that negative emotions may serve a basic, short-term motivational function, but frequent negative feedback loops may have unfavorable effects in the long run (Vogl & Pekrun, 2016).

The effects on positive emotions were significant also, with *pride* being higher for students who did not receive grades in addition to comments. Interestingly, unlike negative emotions, the positive emotion of pride negatively predicted change in scores. Possibly, pride may result in lower motivation, eliciting students' perception of their performance being "good enough" (Lipnevich, Smith, 2009a, 2009b). This finding also extends previous research showing differential links of negative and positive emotions and outcomes. For example, Kaufmann et al. (2019) revealed that although mitigating the experience of negative emotions, illusions of personal control did not foster positive emotions. Perceiving a loss of illusory control, however, significantly reduced the experiences

of positive emotions, and had no effect on negative emotions. Similarly, we should closer investigate separate mechanisms for positive and negative affect and its links to various types of feedback and achievement. Clearly, these links are not always intuitive.

To further consider grades, prior research suggested that grades may undermine students' interest relative to comments, to short-circuit students' thinking, and prevent them from using detailed feedback to improve their work. Research findings suggest that grades are perceived as potent sources of control over learning (Narciss & Huth, 2004; Stipek, 2002). Students may perceive personalized comments as useful sources of information, but presentation of a grade proclaims the instructor's control over their learning, refocuses their attention on the self and the quantitative aspect of learning, and reduces their desire or ability to improve their learning. Guskey and Link (2019) suggests, however, that grades are not inherently good or bad. When based on clearly articulated learning criteria, grades can provide information by providing a detailed summary of how well students performed. Hence, there are certain instructional situations in which grades may work well. Our study revealed that, indeed, grades are not all bad, and we need to be more nuanced when discussing their effects on meaningful outcomes. That is, based on the results of our study we can conclude that the presentation of grades increased student negative affect, an outcome that is certainly not desirable. At the same time, we did not observe a decrease in performance, but, rather, a slight increase in it. Considering the omnipresence of grades this is a rather optimistic finding, but a finding that needs further consideration. Exploring long-term accumulated effects of grades on emotions and performance is certainly in order.

Our study design allowed us to examine latency indicators that can serve as proxies for behavioral consequences of feedback. The results revealed no differences in either the time spent reading feedback or the time revising essays among graded and non-graded conditions. Hence, time on task did not appear to be affected by the presentation of a numerical score, the way student affective reactions did. It is possible that the quality of student revisions predicted differences in final scores. In other words, it is not the time invested into revisions but the quality of changes is what gets influenced by grades. Future studies should systematically address this question by examining the types of changes that students carry out when presented or not presented with grades.

In terms of theory, we found support of the CVT (Pekrun, 2006, 2018) as well as Lipnevich et al. (2016), and Goetz et al. (2018) models. According to the CVT, feedback is a powerful antecedent of student cognitive appraisals and subsequent affective responses. Our study showed that indeed feedback predicted students' affective responses, which, in turn, related to their performance on a task. Per Lipnevich et al. (2016), feedback is received in context and has a number of characteristics that interact with those of a student. For example, it may be honest but delivered in a supportive fashion, or be judgmental. It may match what the student is expecting, or be entirely discrepant. Feedback will inevitably produce affective and cognitive responses, and, as a result of those, the student will act adaptively or maladaptively. This study showed that indeed, seemingly subtle differences in feedback elicit differential emotions in students, that, in turn, link to changes in performance. Goetz et al. (2018) described a number of mediators and moderators of feedback-emotions links as well as their potential relations to performance. We found that emotions do serve as mediators of the relation between feedback and indicators of achievement, but further research is needed to explore alternative paths.

5.1. Limitations and future directions

This study is not without limitations. The sample of this study comprised college students, with the majority of the participants being in the first year of studies thus limiting generalizability of our findings. Future studies should downward extend these results and examine whether students in K-12 settings have similar emotional reactions to comments and grades. We must note, however, that recent studies show

that studies conducted in samples of university students reveal lower measurement error; hence, the use of these samples is defensible.

The essay writing task employed in the study involved students working on an essay and then coming back a week later to revise their work based on the provided feedback. It allowed for the isolation of the immediate effects of feedback on students' emotions and performance. Exploring the cumulative effect of feedback on students' affective states and learning by providing them with various kinds of feedback during the course of a semester or a year would aid in developing a more complete view of the potential effects of feedback on important outcomes. Additionally, it would be useful to consider whether presentation of a certain type of feedback leads to differences in performance in those situations, in which students do not receive a chance to revise and resubmit, but are asked to complete a similar assignment later.

Additionally, although the study accounted for time 1 essay performance in evaluating the effects of emotion on the change in essay scores, there were additional learner characteristics that the analyses did not account for. A potential confound is student motivation. That is, some students were more likely to be more motivation than others to work to improve their essay performance. This is a variable that could be explored in future studies, along with students' expectation of feedback. That is, it is possible that students who expected a high grade but received a low one, would have different emotional reactions to feedback. Studies could also explore the effects of rubrics – both holistic and analytical – on students' processing of grades and comments.

Our study provided initial support for CVT tenets that suggests links among feedback, emotions, and performance, as well as components of the model discussed by Lipnevich et al. (2016). Future studies are needed that would examine other variables included in the model, more specifically, cognitive appraisals of control and value. Additional mediators and moderators of links among feedback, emotions, and performance should also be investigated. Future studies should consider integrating variables delineated in Goetz et al. (2018) to obtain a comprehensive picture of the complex relations among different types of feedback, student affective reactions, as well as their performance. For example, Shu and Lam (2016) revealed moderating effects of culture and regulatory focus when examining success and failure feedback. Including culture is of key importance in future investigations. Similarly, Weidinger et al. (2016) and Pat-El et al. (2012) investigated mediators and moderators of links between feedback and motivation. Further examining self-concept, self-regulatory focus, as well as gender and ethnicity could be of value to the filed.

In conclusion, researchers have noted lacking intervention studies that would focus on the provision of feedback with the purpose of eliciting differential emotions and also with respect to emotions having a potentially positive effect on academic outcomes (Goetz et al., 2018). This study provides initial evidence supporting the viability of such studies. We do not in any way encourage teachers to intentionally elicit negative emotions in their students. We do encourage researchers to keep exploring the complex nature of interrelations among feedback, emotions, and performance, so that clear guidelines can be formulated. Obviously, we need more investigations into complex relations among feedback conditions, student characteristics, emotions, and their joint effects on outcomes.

References

Anderson, C. S. (1982). The search for school climate: A review of the research. *Review of Educational Research, 52,* 368–420.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5,* 7–74.

Brown, G. T. L., & Wang, Z. (2013). Illustrating assessment: How Hong Kong university students conceive of the purposes of assessment. *Studies in Higher Education, 38,* 1037–1057.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large

non-clinical sample. *British Journal of Clinical Psychology, 43*, 245–265. https://doi.org/10.1348/0144665031752934.

Evans, C., & Waring, M. (2011). Student teacher assessment feedback preferences: The influence of cognitive styles and gender. *Learning and Individual Differences, 21*(3), 271–280.

Fong, C. J., Warner, J. R., Williams, K. M., Schallert, D. L., Chen, L. H., Williamson, Z. H., & Lin, S. (2016). Deconstructing constructive criticism: The nature of academic emotions associated with constructive, positive, and negative feedback. *Learning and Individual Differences, 49*, 393–399.

Goetz, T., & Hall, N. C. (2013). Emotion and achievement in the classroom. In J. Hattie, & E. M. Anderman (Eds.), *International guide to student achievement* (pp. 192–195). New York, NY: Routledge.

Goetz, T., & Hall, N. C. (2020). Emotion and achievement in the classroom. In J. Hattie, & E. M. Anderman (Eds.), *Visible learning guide to student achievement* (pp. 145–152). New York, NY: Routledge.

Goetz, T., Lipnevich, A. A., Krannich, M., & Gogol, K. (2018). Performance feedback and emotions. In A. A. Lipnevich, & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback* (pp. 554–575). New York, NY: Cambridge University Press.

Guskey, T. R., & Link, L. J. (2019). Exploring the factors teachers consider in determining students' grades. *Assessment in Education: Principles, Policy & Practice, 26*, 303–320.

Harris, L. R., & Brown, G. T. (2009). The complexity of teachers' conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy & Practice, 16*(3), 365–381.

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research, 66*, 99–136.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81–112.

Ilies, R., & Judge, T. A. (2005). Goal regulation across time: The effects of feedback and affect. *Journal of Applied Psychology, 90*, 453–467.

Jolly, J. B., Dyck, M. J., Kramer, T. A., & Wherry, J. N. (1994). Integration of positive and negative affectivity and cognitive content-specificity: Improved discrimination of anxious and depressive symptoms. *Journal of Abnormal Psychology, 103*, 544–552.

Jonsson, A., & Panadero, E. (2018). Facilitating students' active engagement with feedback. In A. A. Lipnevich, & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback*. New York, NY: Cambridge University Press.

Kaufmann, M., Goetz, T., Lipnevich, A. A., & Pekrun, R. (2019). Do positive illusions of control foster happiness? *Emotion, 19*(6), 1014.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford Press.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254–284.

Linnenbrink-Garcia, L., & Pekrun, R. (2011). Students' emotions and academic engagement: Introduction to the special issue. *Contemporary Educational Psychology, 36*, 1–3. https://doi.org/10.1016/j.cedpsych.2010.11.004.

Lipnevich, A. A., Berg, D. A., & Smith, J. K. (2016). Toward a model of student response to feedback. In G. Brown, & L. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 169–185). New York, NY: Routledge.

Lipnevich, A. A., McCallen, L. N., Miles, K. P., & Smith, J. K. (2014). Mind the gap! Students' use of exemplars and detailed rubrics as formative assessment. *Instructional Science, 42*, 539–559.

Lipnevich, A. A., & Smith, J. K. (2009a). Effects of differential feedback on students' examination performance. *Journal of Applied Experimental Psychology, 15*, 319–333.

Lipnevich, A. A., & Smith, J. K. (2009b). "I really need feedback to learn:" Students' perspectives on the effectiveness of the differential feedback messages. *Educational Assessment, Evaluation and Accountability, 21*, 347–367.

Lipnevich, A. A., & Smith, J. K. (2018). *Cambridge Handbook of Instructional Feedback*. New York, NY: Cambridge University Press.

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education, 30*, 520–540.

Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multimedia learning. In H. M. Niegemann, D. Leutner, & R. Brunken (Eds.), *Instructional design for multimedia learning* (pp. 181–195). Munster, NY: Waxmann.

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G. T. L. Brown, & L. R. Harris (Eds.),

Handbook of social and human conditions in assessment (pp. 247–266). New York, NY: Routledge.

Pat-El, R., Tillema, H., & van Koppen, S. W. (2012). Effects of formative feedback on intrinsic motivation: Examining ethnic differences. *Learning and Individual Differences, 22*(4), 449–454.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315–341.

Pekrun, R. (2018). Control-value theory: A social-cognitive approach to achievement emotions. In G. A. D. Liem, & D. M. McInerney (Eds.), *Big theories revisited 2: A volume of research on sociocultural influences on motivation and learning* (pp. 162–190). Charlotte, NC: Information Age Publishing.

Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement emotions and academic performance: Longitudinal models of reciprocal effects. *Child Development, 88*, 1653–1670. https://doi.org/10.1111/cdev.12704.

Peterson, E. R., Brown, G. T. L., & Jun, M. C. (2015). Achievement emotions in higher education: A diary study exploring emotions across an assessment event. *Contemporary Educational Psychology, 42*, 82–96. https://doi.org/10.1016/j.cedpsych.2015.05.002.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135*, 322–338.

Roesch, S. C. (1998). Validity studies the factorial validity of trait positive affect scores: Confirmatory factor analyses of unidimensional and multidimensional models. *Educational and Psychological Measurement, 58*, 451–466.

Roth, B., Becker, N., Romeyke, S., Schafer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence, 53*, 118–137.

Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety Research, 4*, 27–41.

Shu, T. M., & Lam, S. F. (2016). Is it always good to provide positive feedback to students? The moderating effects of culture and regulatory focus. *Learning and Individual Differences, 49*, 171–177.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*, 153–189.

Stevenson, H. W., & Newman, R. S. (1986). Long-term predication of achievement and attitudes in mathematics and reading. *Child Development, 57*, 646–659.

Stipek, D. (2002). Good instruction is motivating. In A. Wigfield, & J. S. Eccles (Eds.), *A Vol. in the educational psychology series. Development of achievement motivation* (pp. 309–332). San Diego, CA, US: Academic Press.

Teigen, K. H. (1994). Yerkes-Dodson: A law for all seasons. *Theory & Psychology, 4*, 525–547.

Van der Kleij, F. M., & Lipnevich, A. A. (2020). Student perceptions of assessment feedback: A critical scoping review and call for research. In *Educational assessment, evaluation and accountability* (pp. 1–29). https://doi.org/10.1007/s11092-020-09331-x.

van Loon, M., & van de Pol, J. (2019). Judging own and peer performance when using feedback in elementary school. *Learning and Individual Differences, 74*, 101754.

Vogl, E., & Pekrun, R. (2016). Emotions that matter to achievement. In G. T. L. Brown, & L. R. Harris (Eds.), *Handbook of social and human conditions in assessment* (pp. 111–128). New York, NY: Routledge.

Von Humboldt, S., Monteiro, A., & Leal, I. (2017). Validation of the positive and negative affect scale (PANAS) for use with cross-national older adults. *Review of European Studies, 9*, 10–19.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology, 54*, 1063–1070.

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology, 76*, 820–838.

Weidinger, A. F., Spinath, B., & Steinmayr, R. (2016). Why does intrinsic motivation decline following negative feedback? The mediating role of ability self-concept and its moderation by goal orientations. *Learning and Individual Differences, 47*, 117–128.

Westphal, A., Kretschmann, J., Gronostaj, A., & Vock, M. (2018). More enjoyment, less anxiety and boredom: How achievement emotions relate to academic self-concept and teachers' diagnostic skills. *Learning and Individual Differences, 62*, 108–117.

Zeidner, M. (2007). Test anxiety: Conceptions, findings, conclusions. In P. A. Schutz, & R. Pekrun (Eds.), *Emotion in education* (pp. 165–184). San Diego, CA: Academic Press.