The Effects of Feedback from Aptitude Tests in College Courses on Students' Academic Self-Perceptions and Planning

By
Emma Julianne Shaffer

Department of Psychology and Neuroscience
University of Colorado Boulder
Boulder, Colorado 80309

Defense Date: Friday, 22 March 2024
Honor Thesis Advisor: Tiffany Ito, Ph.D.

Defense Committee:
Tiffany Ito, Psychology and Neuroscience, Thesis Advisor
Albert Kim, Psychology and Neuroscience, Honors Council Representative
Zak Kopeikin, Department of Philosophy, Outside Member

Abstract

It is well documented that there are gender disparities in science, technology, engineering, and mathematics (STEM) fields. Problems with self-efficacy and belonging are established as potential causes of this gap. Previous literature demonstrated that low self-efficacy correlated with high procrastination and, thus, performance. Therefore, it was hypothesized that manipulating feedback on an aptitude test would cause changes in self-efficacy and expected procrastination in undergraduate students. Four hundred eighty-one undergraduate students at the University of Colorado Boulder were recruited and took an aptitude test for a hypothetical class, received manipulated feedback, saw a syllabus for that class, and then answered questions about expected procrastination, self-efficacy, belonging, expected effort, and anticipated grade. Contrary to the hypothesis, results indicated that negative feedback decreased expected procrastination. Further, expected effort mediated the relationship between feedback condition and procrastination, such that participants who received negative feedback on the aptitude test expected greater effort in the class and then planned to procrastinate less. It was also found that students experienced low self-efficacy and belonging in response to negative feedback, but positive feedback did not boost self-efficacy or belonging as compared to no feedback. These effects were amplified in women. Further research in an ecologically valid environment is necessary to affirm the present study's findings, such as an intervention to remove feedback from aptitude tests administered by professors before they teach course material. This could boost self-efficacy and belonging for students starting with less knowledge of the material, without negatively impacting students who are familiar with the material.

The Effects of Feedback from Aptitude Tests in College Courses on Students' Academic Self-Perceptions

This study seeks to address the problem of procrastination in college students. Procrastination can lead to low performance in classes (Steel, 2007). Thus, it is necessary to understand the cause of procrastination to subdue procrastination behaviors. Procrastination is a familiar behavior to most and involves putting off or postponing a task or decision-making process (Fernie et al., 2017). In students, procrastination usually manifests in students as putting off starting or completing homework assignments, readings, studying for exams, and writing papers. Van Dinther et al. (2011) demonstrated that procrastination behaviors are higher in undergraduate college students than in the general population. This creates the question, what is it about undergraduate classes that increase procrastination behaviors?

This study investigated how self-efficacy affects expected procrastination behavior in a classroom setting. Self-efficacy is defined as the strength of a person's belief in their ability to succeed. Self-efficacy is known to influence behavior, as people fear putting themselves into a situation beyond their coping abilities (Bandura, 1977). Coping for students in academic settings involves the time and effort students put into their classes to receive their expected results. Social modeling, such as social comparison, predicts self-efficacy by comparison among people (Bandura, 1997). Undergraduate college student environments breed competition by providing feedback on works in a format that allows students to compare themselves to others. Common classroom practices, such as grade book features that show mean, median, range, and upper and lower quartile data on each assignment, allow students to compare themselves to others in the class. Students may find this feedback to hurt or help their self-efficacy for a particular class. Solomon and Rothblum (1984) polled students and found that they attributed their procrastination to fear of failure almost half the time. Self-efficacy can logically play a role in

fear of failure. Thus, in the present study, it is expected that participants with low self-efficacy will expect to procrastinate more than participants with high self-efficacy.

Previous literature demonstrated a significant negative correlation between levels of self-efficacy and amount of procrastination (Klassen et al., 2008; Wäschle et al., 2014). For example, Klassen et al. (2008) studied students throughout the course of a semester and found that students with strong academic self-efficacy and strong global self-esteem were less likely to procrastinate than students with low academic self-efficacy and low global self-esteem. Furthermore, Wäschle et al. (2014) also conducted a longitudinal study in which students self-monitored their preparation, goal achievement, cognitive learning strategies, self-efficacy, state procrastination, class preparation, and performance throughout a class. It was found that the effect of perceived goal achievement on procrastination was mediated by self-efficacy. Perceived goal achievement was shown to reduce students' self-efficacy, and from that, procrastination increased. The effects of students' procrastination behavior further reduced their perceived goal achievement. This creates a vicious circle of procrastination where low self-efficacy contributes to high procrastination and low goal achievement, which yields low self-efficacy. Students may find themselves trapped in this cycle without understanding it. These studies offer interesting hypotheses for how self-efficacy impacts procrastination, but the correlational nature of these studies precludes the ability to test for causal relationships. The present study aims to demonstrate a more causal relationship between self-efficacy and procrastination. If more research is conducted to explore this problem further, students and instructors can be educated on the nature of procrastination and interventions can be provided to boost self-efficacy. These studies provide insight into how procrastination is tied to self-efficacy and demonstrate further research is required to find a causal relationship between the two variables. The present study

seeks to fill this gap in previous research by experimentally manipulating feedback to reduce self-efficacy and determine if self-efficacy can affect students' planned procrastination behavior.

It is also necessary to establish the implications of high procrastination and low self-efficacy. Since self-efficacy seems to affect procrastination, it is important to establish what procrastination and, thus, self-efficacy affects. Steel (2007) used a meta-analysis to demonstrate a consistently negative significant but small relationship between academic procrastination and academic performance. Performance variables included overall GPA, course GPA, final exam scores, and assignment grades. This demonstrated that procrastination had a significant role in academic performance. Studying the causes of procrastination can assist students in their ability to learn and increase their grades. Further, Komarraju and Nadler (2013) found that students with high self-efficacy were more likely to subscribe to a growth theory of intelligence; that is, they believe they can change their intelligence through effort. Along with that, students with high self-efficacy self-reported higher GPAs. Thus, the present study will measure variables such as anticipated grade and expected effort to determine if feedback can also manipulate these expected performance results.

Self-efficacy and belonging measure different facets of a student's perceptions of a class, but they both appear to relate to students' overall class experience. While self-efficacy measures a person's belief in their ability to succeed, belonging measures a person's subjective sense of being valued, included, and accepted by a particular group (Lewis, et al. 2016). The present study uses a hypothetical biology class because it is known that self-efficacy and belonging are often low for women in STEM classes. Wilson et al. (2015) studied the relationship between belonging, self-efficacy, and student engagement and found that a person's sense of belonging within a STEM class was strongly correlated with engagement in the class. Further, low

self-efficacy was strongly correlated with negative emotional engagement, defined as negative feelings associated with the experience of being in the class. Engagement and positive emotional valence of a class can assist with performance. This literature demonstrated that self-efficacy and belonging appear to play pertinent roles in a student's experience in a class.

Recent studies have investigated the effects of belonging and self-efficacy on academic performance. Sotardi (2022) studied the relationship between institutional belongingness, self-efficacy in the classroom, learning strategies, and academic performance. It was found that belongingness and self-efficacy were correlated with a greater GPA. They also found that high belongingness and self-efficacy were associated with greater use of learning strategies. The authors predicted that the use of these learning strategies assisted students in interacting with their professors and other academic support staff, which helped students perform better in class. Self-efficacy and belonging are not merely feelings a student has; rather, they appear to impact behavior and, thus, performance in college environments. To continue this research, it is necessary to research variables that impact student performance and develop interventions to address these deficits in learning. Thus, the present study will measure belonging to determine if a feedback manipulation has an effect on students' sense of belonging in the proposed course.

This study will also investigate belonging because it is a possible route to understanding gender differences in STEM. Previous literature has investigated the impact of belonging in underrepresented communities. Tellhed et al. (2017) studied this gap by examining high school senior's interests, self-efficacy, and belonging in STEM majors and HEED (Health care, Elementary Education, and Domestic spheres) majors. It was found that men showed more interest, higher self-efficacy, and higher social belonging in STEM majors than women. Conversely, it was found that women showed more interest and higher belonging in HEED

majors than men. Further, women had low self-efficacy in STEM majors, but men did not demonstrate low self-efficacy in HEED majors. These results demonstrated that before students enter a college environment, they already show gender differences in key factors (self-efficacy and belonging) for their success in a class. Students carry these beliefs into their college classrooms and, as explained previously, can develop vicious circles of low performance, procrastination, and self-efficacy.

At the university level, Blaney and Stout (2017) studied self-efficacy and belonging in first-generation college women in an introductory computing course. These women reported significantly lower self-efficacy and belonging as compared to both other women and men in their class. It was found that first-generation women's self-efficacy and belonging were more strongly correlated to interaction with faculty inside the classroom than other students. These results demonstrate the need to study factors such as belonging and self-efficacy, as they have consequences that impact performance and further contribute to the gender divide in STEM fields, as well as the difficulties first-generation college students face in obtaining a degree.

Herrmann et al. (2016) investigated the effect of a female role model on female STEM students. In an intervention whereby students read a letter written by a female graduate student about their experience in STEM classes, students in the intervention condition had significantly higher course grades and lower drop rates in the course than female students in the control condition. The letter had interventions involving attribution, which encouraged students to attribute poor performance to unstable factors such as study strategy rather than ability. The letter also included an intervention to address concerns about belonging, which encouraged students to normalize their feelings of low belonging and understand those feelings are normal and temporary. These findings demonstrate that research on belonging and self-efficacy can assist in

closing the gender divide in STEM. This further demonstrates the need to conduct the present study.

Study Overview

It is well known that people, especially college students, struggle with procrastination. The problems go beyond individual choices to delay work and are deeply intertwined with people's belief in their abilities and sense of community with those around them. Thus, it is important to determine the causes of procrastination in student populations so students can identify why they procrastinate, and professors can identify how they can assist students in reducing procrastination behavior in their classes. The present study emulates students' experience at the beginning of an introductory molecular biology class and manipulates feedback by providing positive, neutral, or negative feedback on an aptitude test for the class.

Thus, it was hypothesized that the feedback manipulation would be positively related to self-efficacy, whereby participants in the low feedback condition would have lower self-efficacy, participants in the high feedback condition would have higher self-efficacy, and participants in the neutral feedback condition would lie somewhere in the middle. A similar pattern was predicted for measures of belonging and anticipated grade. The relationship between feedback and procrastination, as well as expected effort, was predicted to be the opposite, where negative feedback would yield high expected procrastination and effort, and positive feedback would yield low expected procrastination and effort. Further, it was hypothesized that women would be more negatively affected by negative feedback than men but would not be more positively affected by positive feedback because of preexisting gaps between men's and women's self-efficacy in STEM classes. Thus, women in the negative feedback condition were expected to have lower self-efficacy, belonging, and anticipated grade, and higher planned procrastination

and expected effort than men in the negative feedback condition. However, it was expected that women and men in the positive feedback condition would not experience different self-efficacy, procrastination, belonging, expected effort, or anticipated grade.

Method

Subjects

Participants included 481 undergraduate students taking an introductory psychology course at the University of Colorado at Boulder. Nine participants were excluded because they did not complete the study. Three participants were excluded for extremely short or long study duration times out of concern for potential lack of engagement and attention (less than five minutes or greater than forty minutes). Two participants were excluded for straightlining on components of the study with reverse-coded items. Eight participants were excluded because they were not in the gender binary or did not select a gender, which would preclude analyses involving gender as a predictor. These exclusions leave a final sample size of 460.

The final sample after exclusions was primarily women (337 women, 123 men). Self-reported race among the final sample was as follows: 319 participants identified as White; 62 participants identified with multiple races; 35 participants identified as Latino/a/x/e; 18 participants identified as East Asian; 9 participants identified as South Asian; 7 participants identified as Black; 4 participants identified as Arabic; and 6 participants did not specify their race. One participant selected an age range of 31-35, one selected 36-40, and one selected >50; their data were included in the final analyses. Excluding these participants, ages ranged from 18-30 (M = 18.85, SD = 1.28). 59.13% of the participants were first-year students, 25.87% were second-year students, 10% were third-year students, and 5% were fourth-year students. Participants were not compensated but did receive class credit for completing the survey.

Procedures

Students completed the study during a one-time 30-minute online session. They signed up for a timeslot through SONA, a participant management pool comprised of introductory psychology students. Then, at their selected time, participants joined a Zoom room where they were told that the purpose of the study is to understand students' perceptions of courses. Next, they were given the link to the online survey. In the survey, participants took a 10-question test on topics related to the syllabus they later viewed. They were randomly assigned feedback from the test and then viewed a syllabus for the class related to the test. Then, they were asked a series of questions about their predicted performance and feelings if they were to take the class. Finally, they completed a demographic questionnaire and were debriefed on the deception in the study. The data were analyzed using version 4.3.2 of RStudio. Materials and methods below are described in the order they were presented to participants.

Materials

Aptitude Test

The study included a 10-question molecular and cellular developmental biology (MCDB) test (Figure 1). Questions were pulled from in-class problems evaluating student comprehension of material from an introductory MCDB course offered at the university. The questions that were selected were intentionally confusing or difficult. In some cases, questions offered incorrect answers that seemed correct. This was done to ensure participants would believe the feedback condition they received.

Figure 1

Aptitude Test Questions

When 2 atoms share electrons, what kind of bond are they held together by?
a. Noncovalent bond
b. Covalent bond
c. Ionic bond
d. Hydrogen bond

Which of the following chemical bonds or interactions DO NOT involve electrostatic interactions between molecules or atoms?
a. Ionic bonds
b. Van der Waals interactions
c. Hydrogen bonds
d. hydrophobic/hydrophilic interactions
e. None of the above

At what temperature is water least dense?
a. -5°C
b. 5°C
c. 18°C
d. 38°C

When prions misfold…
a. The primary structure of the protein has been altered
b. The secondary structure of the protein has been altered
c. Both of the above

Can bacteria cells do photosynthesis?
a. Yes
b. No

What membrane is the membrane of the rough endoplasmic reticulum intimately connected to?
a. The inner nuclear membrane
b. The outer nuclear membrane
c. The outer mitochondrial membrane
d. The golgi membrane

A protein lacking a signal sequence will likely reside in which of the following?
a. Outside the cell
b. In the cytoplasm
c. In the nucleus
d. In the endoplasmic reticulum

To be effective, ligand binding to a receptor protein must be:
a. Covalent and irreversible
b. Noncovalent and reversible
c. Covalent and reversible
d. Noncovalent and irreversible

Base pairing is important for which of the following processes?
a. tRNA structure
b. Transcription
c. Protein synthesis
d. All of the above

An enzyme catalyzes a reaction by…
a. Lowering the overall free energy of a reaction
b. Lowering the activation energy of a reaction
c. Lowering the free energy of the reactants
d. Lowering the free energy of the products

Feedback Manipulation

After submitting the test, participants received feedback that was not representative of performance; rather, feedback was randomly assigned. Manipulated feedback was adapted from a self-efficacy manipulation by Bouffard et al. (2005). In that study, feedback was more extensive and provided graphs, allowing participants to compare themselves. The present study shortened the amount of feedback provided. The Qualtrics survey system randomly assigned students to one of three conditions: positive feedback; negative feedback; and neutral feedback. In the positive feedback condition, students received this feedback, “We’ve scored your answers on the aptitude test. You did quite well. Your score would place you in the top group of students who have taken MCDB 1125. This score suggests you would do well if you took this class.” Similarly, students in the low self-efficacy condition received this feedback, “We’ve scored your answers on the aptitude test. Your score was relatively low. Your score would place you in the bottom group of students who have taken MCDB 1125. This score suggests you would have difficulties doing well if you took this class.” Students in the control condition received no feedback on their performance.

Syllabus

After receiving feedback on the aptitude test, participants viewed a two-page syllabus (Figure 2) that outlined the requirements and procedures of the class participants were asked to consider. The instructor and TA both had traditionally male names, as male instructors are common in STEM classes. This may lead to further gender divides in self-efficacy and belonging results in this study. Items in the syllabus followed standard practices at the university, such as course website, lecture meeting time, required text, course topics, attendance policies, and grading policies. Assignment types presented in the syllabus were chosen to measure predicted procrastination through multiple assignment modalities. Statements were included in the syllabus, such as, “It is strongly recommended that you begin your homework with adequate time to attend office hours if you have questions”, to cue participants to seriously consider when they would begin assignments.

Figure 2

Syllabus Presented to Participants

Introduction to Molecular, Cellular & Developmental Biology for Non-Majors

MCDB 1125

Instructor:	Matthew Harris, PhD Gold 225A Office Hours: Mondays 3 - 4 pm
TA:	David Picard Gold 338D Office Hours: Thursdays 11 am - 12 pm
Course Website:	Course materials, assignments, and other details can be found on the Canvas page
Lecture Meeting Time:	The class meets on Monday, Wednesday, and Friday from 11:10 am - 12:00 pm in Gold A2B40
Required Text:	*Molecular and Cell Biology*, 8th Edition. By Harvey Lodish et al.

Course Topics Overview:
Unit 1: Bonds, proteins, carbohydrates, lipids, nucleic acids
Unit 2: DNA replication; transcription and translation
Unit 3: Protein structure, organelles, protein transport
Unit 4: Microtubules, actin filaments, intermediate filaments, cell cycle

Attendance:
While attendance will not directly be taken in class, clicker questions will be administered in class throughout the semester for participation points. You can miss up to 3 lectures of clicker questions and not be penalized.

Homework:
Weekly homework assignments will be due on Canvas at the beginning of every week. They will be due before class on Mondays at 11:09 am. It is strongly recommended that you begin your homework with adequate time to attend office hours if you have questions. Homework will be graded based on accuracy, not completion.

Readings:
There will be weekly reading assignments throughout the semester. Readings will come from the assigned textbook, *Molecular and Cell Biology*, 8th Edition. While we are not collecting any proof that you have done your readings, doing the readings is vital to understanding class information, completing homework assignments, and doing well on exams. Ensure you complete the readings for the week before the first class of the week.

Midterms:
Over the course of the semester, there will be 3 midterm exams. The midterms will be taken in class during class time. Ensure you are taking the time to adequately study for your midterms.

Final Exam:
A cumulative final exam will be administered during this class's exam period. It will take place in our classroom in Gold A2B40. Ensure you are understanding the material as we go in order to be successful on the final exam.

Final Paper:
The final paper will be written about a topic of your choice from a list of preapproved topics. If you want to write about something that is not on the list provided, you must email me or come to office hours to discuss what you would like to write about. To write this paper, you will be required to research a topic using validated scientific articles and write about your findings in 7-10 pages. More details will be provided later in the semester. It will be due on the last day we meet for class.

Grading Policy:
The due dates for all the assignments are the deadline. If you need an extension, reach out to me by email before the due date.

Clickers: 5%
Homework: 10%
Midterms (3): 15% each for a total of 45%
Final Paper: 20%
Final Exam: 20%

A	100% - 93%	C+	79% - 77%
A-	92% - 90%	C	76% - 73%
B+	89% - 87%	C-	72% - 70%
B	86% - 83%	D	69% - 60%
B-	82% - 80%	F	<60%

Measures

This study measured variables including expected procrastination, self-efficacy, belonging, expected effort, and anticipated grade.

Expected Procrastination

Expected procrastination was assessed with five items based on assignments described in the syllabus. Participants were asked, if they were to take the class, when they would begin: (1) homework assignments; (2) weekly reading assignments; (3) studying for midterm exams; (4) studying for the final exam; and (5) writing the final paper. Items had different timescales because some assignments require less expected time to complete and others more. For items (1) and (2), the options for response were: *more than a week before it is due; one week before it is due; a few days before it is due; the day before it is due; and on the day it is due*. For items (3) and (4), the options for response were: *more than a month before the exam; a few weeks before the exam; about a week before the exam; the day before the exam; and on the day of the exam.* For item (5), the options for response were: *more than a month before it is due; a few weeks before it is due; about a week before it is due; the day before it is due; the day it is due*. Ratings in the scale were averaged to form a single index of procrastination, with higher scores indicating higher expected procrastination behavior. The reliability score for this measure was considered adequate according to the standards used in the field of social psychology ($\alpha = .76$).

Self-Efficacy

The self-efficacy scale was adapted from previous research by Ito and McPherson (2018). Each item was rated on a 6-point scale with labels of *Strongly Disagree, Disagree, Somewhat Disagree, Somewhat Agree, Agree,* and *Strongly Agree*, where higher scores indicate more agreement. Items measure students' expected capabilities in the class presented to them.

Participants were asked to rate how confident they are that they can do the following: *demonstrate what I know on exams in MCDB 1125; complete homework assignments by myself in MCDB 1125; do well on the final MCDB 1125 paper; perform well on exams in MCDB 1125; and learn the basic concepts associated with MCDB 1125*. Ratings in the scale were averaged to form a single index of self-efficacy, with higher scores indicating higher self-efficacy. The reliability score for this measure was considered adequate according to the standards used in the field of social psychology ($\alpha = .86$).

Belonging

The belonging scale was adapted from previous research by Ito & McPherson (2018). It was rated on a 6-point scale with labels of S*trongly Disagree, Disagree, Somewhat Disagree, Somewhat Agree, Agree,* and *Strongly Agree,* where higher scores indicate more agreement. The items for belonging were: *I feel a connection with the community associated with MCDB 1125; I would feel like an outsider in MCDB 1125* (reverse coded)*; I would feel like I belong in MCDB 1125; people in MCDB 1125 would accept me; people in MCDB 1125 seem like they would be a lot like me; I feel similar to the kinds of people who have what it takes to succeed in MCDB 1125; I'm not certain I would "fit in" intellectually in my class* (reverse coded)*; I feel like other students in MCDB 1125 would have skills that I don't have; I would worry that no matter how hard I try, I won't be able to perform successfully in MCDB 1125*. Ratings in the scale were averaged to form a single index of belonging, with higher scores indicating higher belonging. The reliability score for this measure was considered adequate according to the standards used in the field of social psychology ($\alpha = .88$). The last 19 participants did not receive questions about belonging. The sample size for any analysis involving belonging is $n = 441$.

Expected Effort

The expected effort scale was adapted from previous research by Smith et al. (2012). Expected effort was measured with four items, in which the first three were rated on a 6-point scale with labels of *none at all, a low amount, a somewhat low amount, a somewhat high amount, a high amount,* and *a great deal*, where higher scores indicate more expected effort. The items were: How *much effort do you think it would take to succeed in MCDB 1125? How much hard work do you think you would need to put in for MCDB 1125? How much effort do you think it would take to get a passing grade in MCDB 1125?* The final item was *How easily and naturally do you think you would understand class material?* The options for response were: *not at all easily and naturally; not easily and naturally; somewhat easily and naturally; relatively easily and naturally; easily and naturally; very easily and naturally* (reverse coded). Ratings in the scale were averaged to form a single index of expected effort, with higher scores indicating higher expected effort. The reliability score for this measure was considered adequate according to the standards used in the field of social psychology ($\alpha = .83$).

Anticipated Grade

Anticipated grade was measured via a drop-down menu with grades from A-F, including pluses and minuses as options. In analyses, anticipated grade ratings are reported on a 4.0 scale.

Results

Randomization Checks

A one-way ANOVA was used to determine if there truly was a randomization of participants' ability in the class via the aptitude test across randomly assigned feedback conditions (see Table 1). Aptitude tests were later scored during analyses. As predicted, there was no difference in aptitude test scores between conditions, $F(2,457) = .058$, $p = .943$.

A Pearson's χ^2 test was used to determine whether there was a difference between frequencies of participants' previous experience with molecular biology between randomly assigned feedback conditions (see Table 1). There was not a significant difference between the number of participants who took molecular biology in each condition $\chi^2(2,457) = 0.491$, $p = 0.782$.

A Pearson's χ^2 test was used to determine if there was a difference between frequencies of each gender between randomly assigned feedback conditions (see Table 1). There was not a significant difference in gender distributions between conditions $\chi^2(2,457) = 1.167$, $p = 0.558$.

Table 1

Randomization Check Results

Condition	Aptitude Score	Has Previous Experience	Gender
Positive Feedback	$M = 44\%$	5%	Women: 76.3% Men: 23.7%
Neutral Feedback	$M = 43\%$	4.57%	Women: 71.1% Men: 28.9%
Negative Feedback	$M = 44\%$	4.13%	Women: 72.4% Men: 27.6%

Variable Correlations

Before testing hypotheses, a correlation matrix with all dependent variables was run. Unexpectedly, self-efficacy was not significantly correlated with expected procrastination. However, as predicted, self-efficacy was positively correlated with belonging and anticipated grade and negatively correlated with expected effort.

Table 2

Correlational Relationship Between Variables

Variable	1	2	3	4	5
1. Self-Efficacy	-				
2. Procrastination	-.02	-			
3. Belonging	.61***	-.03	-		
4. Anticipated Grade	.66***	.02	.61***	-	
5. Expected Effort	-.54***	-.15**	-.67***	-.59***	-

Note. *** Significant at $p < .0001$, ** Significant at $p < .001$

Hypothesis Tests

Procrastination

A 3 (condition: positive feedback, neutral feedback, negative feedback) x 2 (gender: woman, man) ANOVA was used to determine if negative feedback increased expected procrastination and positive feedback decreased expected procrastination. There was a statistically significant main effect of condition, $F(2, 454) = 3.02$, $p = .0498$, but no significant results in the subsequent t-test. However, unlike the hypothesis, looking at mean scores, procrastination was higher in the positive feedback condition ($M = 2.56$, $SD = .54$) than in the negative feedback condition ($M = 2.43$, $SD = .57$), $p = .093$. There was a gender main effect wherein, on average, men ($M = 2.56$, $SD = .50$) expected to procrastinate more than women ($M = 2.44$, $SD = .54$), $F(2,454) = 3.02$, $p = .025$. There was no significant interaction between condition and gender, $F(2,454) = 1.281$, $p = .279$.

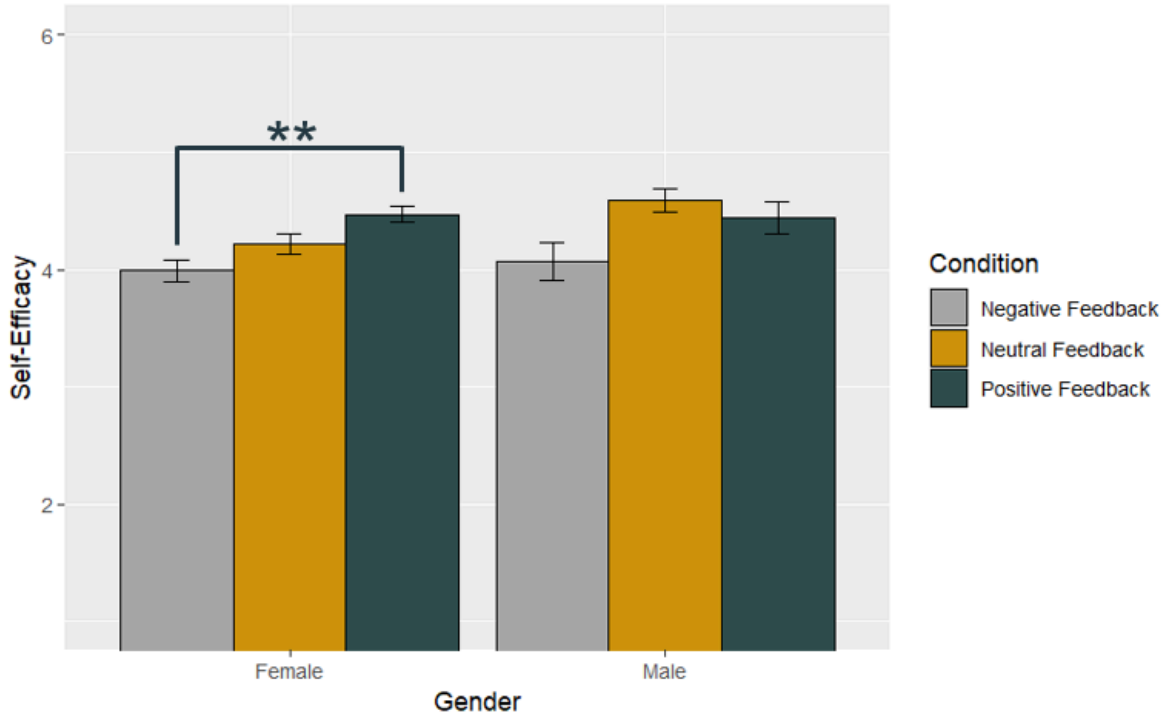
Self-Efficacy

A 3 (condition: positive feedback, neutral feedback, negative feedback) x 2 (gender: woman, man) ANOVA was used to compare the self-efficacy score of the three feedback conditions by both genders. There was a statistically significant main effect of condition in self-efficacy score, $F(2, 454) = 10.42$, $p < .001$. Specifically, participants rated themselves as having higher self-efficacy in the positive feedback condition ($M = 4.46$, $SD = .81$) than in the

negative feedback condition (M = 4.01, SD = 1.00), $p < .001$, and higher self-efficacy in the neutral feedback condition (M = 4.32, SD = .85) than the negative feedback condition ($p < .01$), but there was not a difference between the positive feedback condition and the neutral feedback condition, $p = .498$. Neither the gender main effect, $F(2, 454) = 2.51$, $p = .114$, nor the condition by gender interaction was significant, $F(2, 454) = 1.72$, $p = .180$. Despite the lack of a significant ANOVA interaction, self-efficacy, as a function of the condition by gender interaction, was further explored. A pairwise comparison among the different feedback and gender conditions was run because self-efficacy is a central part of the present study. A simple condition effect within gender was found between women in the negative feedback condition (M = 3.99, SD = .97) and women in the positive feedback condition (M = 4.47, SD = .81), $p < .001$ (see Figure 3).

Figure 3

Interaction between Gender, Feedback Condition, and Self-Efficacy



Note. Error bars represent the standard deviation of the mean. ** Significant at $p < .001$, * Significant at $p < .01$.

Belonging

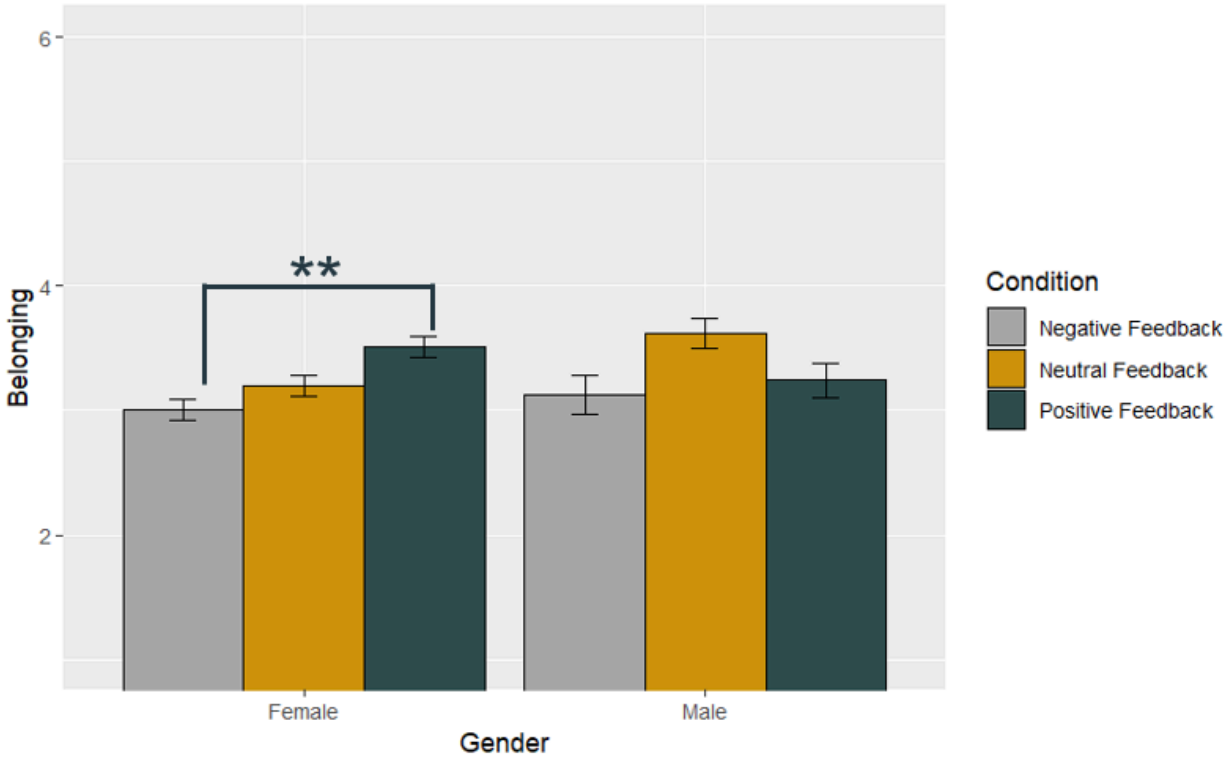
There was an error in data collection, wherein the last 19 participants did not receive any questions related to belonging. As a result, for any tests that involve belonging, analyses are based on $n = 441$. Expected effort and anticipated grade were the two dependent variables collected after belonging. The relationship between feedback conditions and both of those variables was compared between the full data set and the data set with missingness. The significance of relationships between conditions and variables is the same in both data sets. Thus, the belonging missingness did not appear to impact results.

A 3 (condition: positive feedback, neutral feedback, negative feedback) x 2 (gender: woman, man) was used to compare the belonging score of the 3 feedback conditions by both genders. Supporting the hypothesis, there was a statistically significant main effect of feedback condition on the belonging score, $F(2, 435) = 8.72$, $p < .001$. Specifically, participants in the positive feedback condition ($M = 3.46$, $SD = .86$) reported higher belonging than those in the negative feedback condition ($M = 3.04$, $SD = .91$), $p < .001$. Further, participants in the neutral feedback condition ($M = 3.31$, $SD = .84$) reported higher belonging than those in the negative feedback condition, $p < .05$. There was no difference between the positive feedback and neutral feedback condition, $p = .478$. There was no gender main effect, $F(2,435) = 1.32$, $p = .251$. However, it was found that there was a significant interaction between condition and gender, $F(2, 435) = 4.27$, $p < .05$ (see Figure 4). A pairwise comparison was used to understand the interaction. A simple condition effect within gender was found between women in the negative

feedback condition (M = 3.01, SD = .88) and women in the positive feedback condition (M = 3.51, SD = .88), $p < .001$.

Figure 4

Interaction between Gender, Feedback Condition, and Belonging



Note. Error bars represent the standard deviation of the mean. ** Significant at $p < .001$, * Significant at $p < .01$.

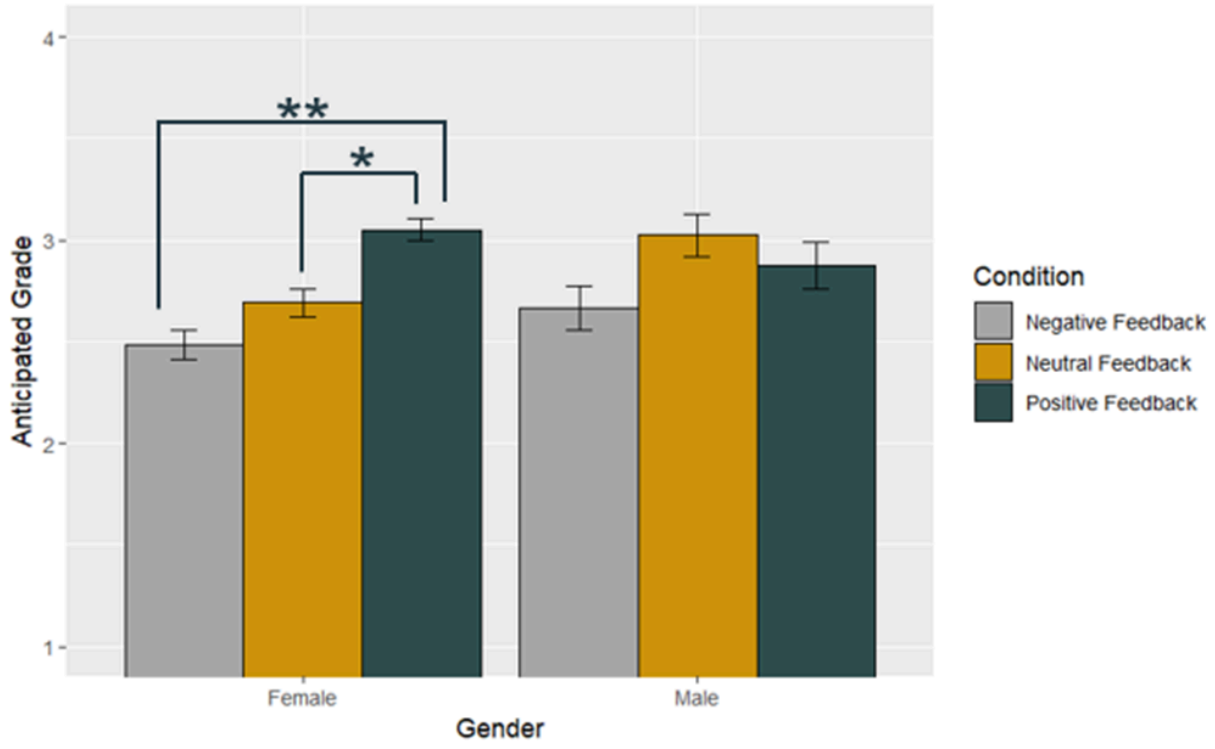
Anticipated Grade

A 3 (condition: positive feedback, neutral feedback, negative feedback) x 2 (gender: woman, man) ANOVA was used to determine if the feedback condition had an effect on anticipated grade. Note that anticipated grade was scored on a 4.0 scale. As predicted, feedback condition affected participants' anticipated grade, $F(2, 454) = 18.18$, $p < .001$. Participants in the positive feedback condition (M = 3.01, SD = .62) expected higher grades than the neutral

feedback condition (M = 2.79, SD = .74), $p < .05$, and the negative feedback condition (M = 2.53, SD = .72), $p < .001$. Also, participants in the neutral feedback condition anticipated higher grades than participants in the negative feedback condition, $p < .01$. There was no significant gender main effect, $F(2,454) = 2.77$, $p = .097$. However, there was a significant interaction between condition and gender, $F(2,454) = 4.20$, $p < .05$. A pairwise comparison was used to understand the interaction. There were no condition differences among men; however, as predicted, there were condition differences among women. For women, positive feedback (M = 3.05, SD = .59) and negative feedback (M = 2.48, SD = .73) conditions were significantly different, $p < .001$. Positive feedback and neutral feedback (M = 2.69, SD = .74) conditions were also significantly different, $p < .01$ (see Figure 5). There was no significant difference between women in neutral and negative feedback conditions, $p = .409$.

Figure 5

Interaction between Gender, Feedback Condition, and Anticipated Grade



Note. Error bars represent the standard deviation of the mean. ** Significant at $p < .001$, * Significant at $p < .01$.

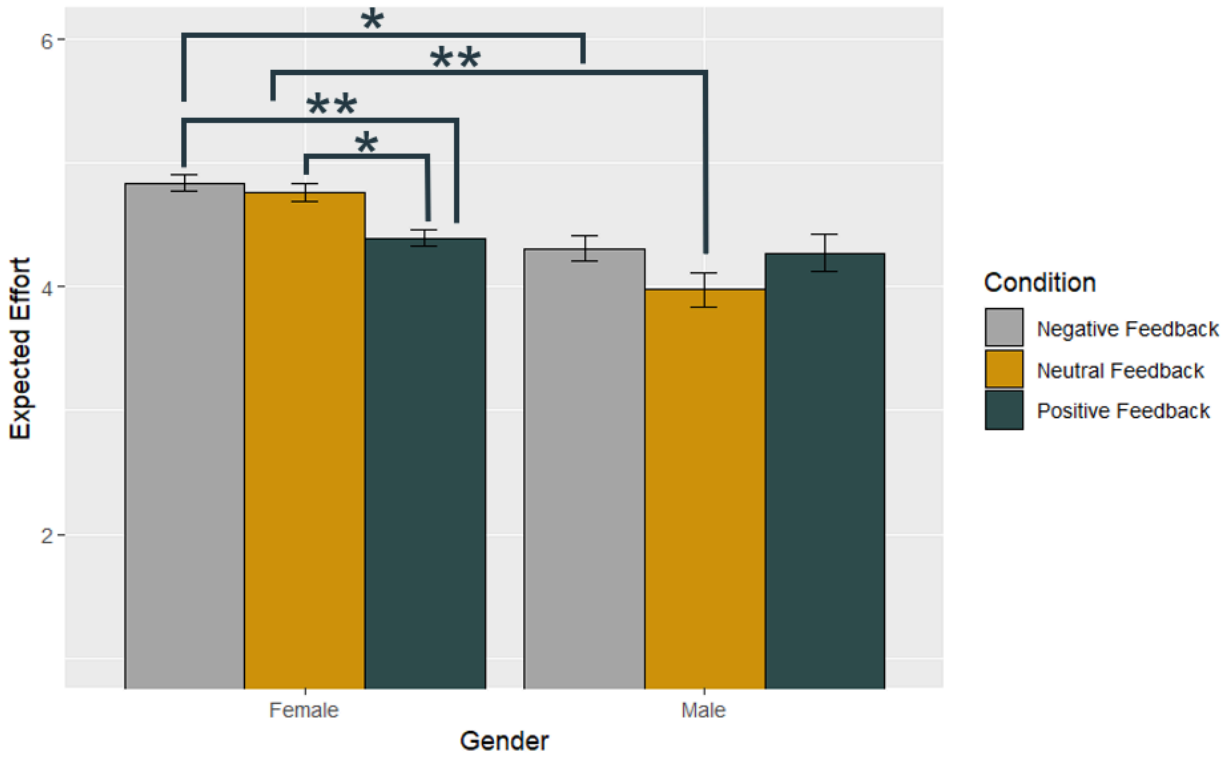
Expected Effort

A 3 (condition: positive feedback, neutral feedback, negative feedback) x 2 (gender: woman, man) ANOVA was used to determine if feedback conditions had an effect on expected effort. As predicted, the feedback condition affected expected effort, $F(2, 454) = 7.13$, $p < .001$. Consistent with the hypothesis, participants expected the class to be more effort in the negative feedback condition ($M = 4.69$, $SD = .74$) than the neutral feedback condition ($M = 4.54$, $SD = .87$) $p < .01$ and the positive feedback condition ($M = 4.36$, $SD = 79$), $p < .0001$. There was also a gender main effect wherein women ($M = 4.66$, $SD = .76$) expected the class to be more effort than men ($M = 4.18$, $SD = .85$), regardless of condition, $F(2, 454) = 37.06$, $p < .001$. There was

also a significant interaction between gender and condition, $F(2,454) = 5.66$, $p < .01$. A pairwise comparison was used to understand the interaction. It was found that there were no condition differences among men, but there were condition differences among women. For condition effects within women, the positive feedback condition ($M = 4.39$, $SD = .76$) and the negative feedback condition ($M = 4.84$, $SD = .72$) were significantly different, $p < .001$. The positive feedback and the neutral feedback condition ($M = 4.76$, $SD = .73$) were also significantly different for women, $p < .01$. There was no significant difference between neutral and negative feedback conditions for women, $p = 1.00$. Two within-condition gender effects were found. For the neutral feedback condition, men ($M = 3.98$, $SD = .93$) and women ($M = 4.76$, $SD = .73$) significantly differed, $p < .001$. In the negative feedback condition, men ($M = 4.31$, $SD = .67$) and women ($M = 4.84$, $SD = .72$) significantly differed, $p < .01$. In the positive feedback condition, there was no difference by gender, $p = 1.00$ (see Figure 6). These results support the hypothesis that women were more negatively affected by the negative feedback condition than men, but women were not more positively affected by the positive feedback condition than men.

Figure 6

Interaction between Feedback Condition, Gender, and Expected Effort



Note. Error bars represent the standard deviation of the mean. ** Significant at $p < .001$, * Significant at $p < .01$.

Tests of Mediation

To test the main hypothesis that self-efficacy affects procrastination, a mediation model (bootstrap $n = 5000$) using PROCESS macro (Hayes, 2017) was run. The analysis did not support the hypothesis; the results demonstrated that self-efficacy did not affect procrastination. First, feedback condition significantly predicted self-efficacy, ($b = .23$, $SE = .05$, $t = 4.46$, $p = .0000$, 95% CI = [.13, .33]). Second, when both feedback condition and self-efficacy were in the model, feedback condition did significantly predict procrastination ($b = .07$, $SE = .03$, $t = 2.30$, $p = .022$, 95% CI = [.01, .13]), but self-efficacy did not predict procrastination, ($b = -.02$, $SE = .03$, $t = -.84$, $p = .390$, 95% CI = [-.08, .03]). Finally, the indirect effect of feedback condition on procrastination through self-efficacy was not significant (Standardized Indirect Effect = -.005, *SE*

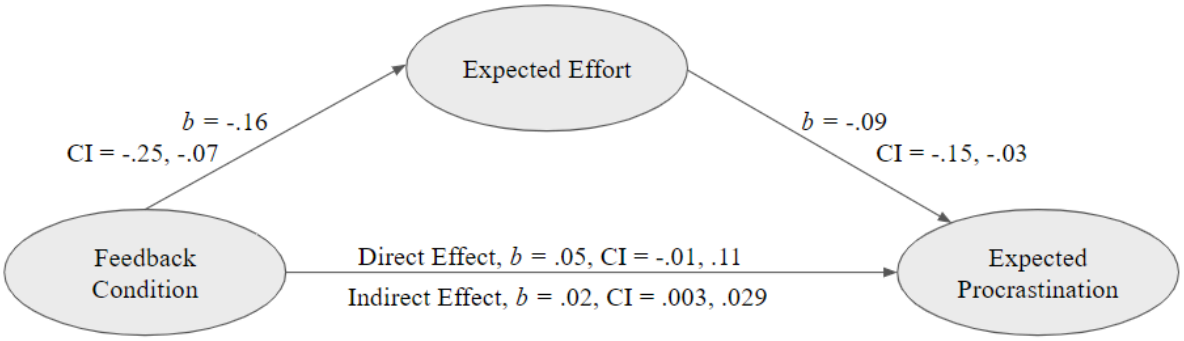
= .0074, 95% CI = [-.02, .01]). Thus, the hypothesis that self-efficacy affects procrastination at the start of a course was rejected.

While changes in efficacy were not the route through which feedback affected expected procrastination, expected effort was tested as a possible route for this relationship. Results showed that feedback condition affected procrastination and expected effort. Further, expected effort was the only variable that was significantly correlated with procrastination, and it was a negative correlation. It would make sense for participants in the positive feedback condition to assume the class was less effort, which led them to assume they would procrastinate more. Since feedback conditions had a different effect on procrastination than predicted, it brought up the question: does expected effort mediate the relationship between feedback condition and procrastination?

Based on these considerations, a mediation model (bootstrap n = 5000) using PROCESS macro (Hayes, 2017) was run to determine if expected effort was a mediating variable between feedback condition and procrastination. The analysis supports this mediation model. First, feedback condition significantly predicted expected effort, (b = -.16, SE = .05, t = -3.60, p = .0003, 95% CI = [-.25, -.07]). Second, when both feedback condition and effort were in the model, feedback condition did not significantly predict procrastination (b = .05, SE = .03, t = 1.66, p = .098, 95% CI = [-.01, .11]) but expected effort did significantly predict procrastination, (b = -.09, SE = .03, t = -3.04, p = .0025, 95% CI = [-.15, -.03]). Finally, the indirect effect of feedback condition on procrastination through expected effort was significant (Standardized Indirect Effect = .02, SE = .01, 95% CI = [.003, .029]) (Figure 7). Thus, expected effort fully mediated the effect of feedback condition on expected procrastination behavior.

Figure 7

Expected Effort Mediates the Relationship Between Feedback Condition and Expected Procrastination



Note. Results demonstrated that feedback condition had a significant effect on expected effort, and expected effort had a significant effect on expected procrastination. Feedback condition did not directly affect expected procrastination, but feedback condition had an indirect effect on expected procrastination.

Discussion

Procrastination, self-efficacy, belonging, expected effort, and anticipated grade were all successfully manipulated by feedback, allowing for further explorations of gender interactions and mediating effects. In the present study, it was hypothesized that participants in the negative feedback condition would predict higher procrastination behavior than those in the neutral and control feedback conditions. The results did not support this hypothesis. While there is a significant difference between all 3 groups, pairwise comparisons showed there were no significant differences when 2 of any group were compared to each other. Additionally, mean expected procrastination scores demonstrated that those who received positive feedback predicted higher procrastination behavior than those in the negative and neutral feedback

conditions. In contrast to the hypotheses, these results demonstrated that students, when told they were not successful on an aptitude test for a class, would predict lower procrastination behavior throughout the course. This is likely due to compensating for a poor score on the aptitude test. On the other hand, students who were told they did well may feel they can procrastinate a little more on assignments since they already had high aptitude in the class.

Wäschle et al. (2014) reported the opposite results, wherein students found themselves in a vicious circle of procrastination. For students in that study, participants self-reported low self-efficacy and high procrastination, which led to low goal achievement. Similarly, Klassen et al. (2008) demonstrated results similar to those of Wäschle et al. (2014). In Klassen et al. (2008), participants with high self-efficacy were less likely to procrastinate than students with low self-efficacy. The differences in these results and the results of the present study could be explained by the differences in methods for these studies. The method in Wäschle et al. (2014) and Klassen et al. (2008) was a longitudinal study that took place throughout the semester. The present study was a one-time study about expectations for a theoretical class. Because of this, participants in the present study were more likely to report idealistic representations of behavior rather than what they would actually do in a class. Alternatively, this study could demonstrate a different relationship between self-efficacy and procrastination. The present study emulated how students feel right at the beginning of a class. In many cases, especially in STEM courses, professors ask students to complete an aptitude test before the professor has taught course material. So, the results of this study demonstrated how students felt at the beginning of a class, not throughout the course of a class. It is possible self-efficacy does not have a relationship with procrastination at the beginning of a class. But, as course content is taught, a negative

relationship emerges, such that students with low self-efficacy procrastinate more. Further research should be done to explore this hypothesis.

In the present study, it was expected that the feedback conditions would manipulate self-efficacy. This was supported by the results. Participants in the negative feedback condition had significantly lower self-efficacy than those in the neutral and positive feedback conditions. However, self-efficacy was not different between the neutral and positive feedback conditions. These results demonstrated that participants did not have any differences in self-efficacy when they were told they did well on the aptitude test compared to when they were told nothing about their results on the aptitude test. In contrast, negative feedback had an adverse effect on self-efficacy.

While the feedback manipulation did affect participants' self-efficacy, the lack of gender main effects demonstrated that, on average, women's and men's self-efficacy did not differ. This was contrary to the expectation that women's self-efficacy would be lower than men's from previous findings of gender differences in STEM, such as Tellhed et al. (2017), who demonstrated that women reported lower self-efficacy than men when they thought about STEM courses. This was a similar method to that of the present study. A possible explanation for the discrepancy in findings could be that the sample pool for this study was a psychology class, which is a type of STEM class, so participants who took the present study may have already self-selected as having confidence and self-efficacy in STEM-related classes. A majority of participants in the present study were STEM majors (62.61%). On the other hand, participants in Tellhed et al. (2017) were high school students without affiliations to a particular major, so there was likely more variability in the students' interests in that study.

Results of the interaction between condition and gender demonstrated that there were between group differences for women but not for men. It appeared that women were more affected by the feedback manipulation than men. This result repeated throughout multiple variables in this study. It is important to note that there were more women who took this study than men, meaning women's results have more statistical power. However, even though there were fewer men in the study than women, results still indicate that the feedback manipulation affected women more than men. Further research on the effects of aptitude tests on men and women can affirm this finding.

It was hypothesized that belonging would be lower in the negative feedback condition and higher in the positive feedback condition. This hypothesis was supported by the results. Belonging also followed the same significance patterns as self-efficacy, wherein negative feedback was significantly different from positive and neutral, but neutral and positive feedback conditions were not different. This demonstrated that belonging was not boosted by positive feedback on the aptitude test but was hurt by negative feedback. This is important because it informs professors that they can remove feedback from aptitude tests in their classes without diminishing belonging or self-efficacy for students high in aptitude.

When results were analyzed by gender, it was found that women's belonging was more affected by the manipulation than men's, as there was a difference between women in the positive feedback condition and women in the negative feedback condition, but that difference was not present in men. As discussed above, Herrmann et al. (2016) created an intervention that could be applied to the present study to boost belonging in women. When women graduate students gave women undergraduate students letters that discussed and normalized feelings of low belonging, participant's grade increased in their STEM classes, and dropout rates were

lower. An intervention like this could help to increase belonging in STEM classes and possibly reduce the potential downstream effects of low belonging. Further, Wilson et al. (2015) demonstrated that increased belonging and self-efficacy within a particular class were associated with increased engagement in the class. An intervention, such as the one demonstrated by Herrmann et al. (2016), could result in boosted academic engagement for those students experiencing low belonging and self-efficacy resulting from gendered effects.

Anticipated grade was predicted to be high in the positive feedback condition and low in the negative feedback condition. This hypothesis was confirmed by the results. Students who were told they did well on an aptitude test before they engaged with course materials expected a higher grade than those who did not receive feedback and those who received negative feedback. Further, negative feedback made students believe they would receive a lower grade in the class than those who did not receive feedback. This demonstrated that feedback on an aptitude test manipulated students' anticipated grades in a class. This expands on results demonstrated in Sotardi (2022) and Komarraju and Nadler (2013), wherein students with high self-efficacy had greater GPAs than students low in self-efficacy. Not only were these variables correlated at the end of a semester when grades came out, but the present study demonstrated that feedback changed anticipated grade before a class even began. Further research could determine if anticipated grade has an effect on actual grade, where students perform at a level that sufficiently satisfies their expectations without reaching their full potential for performance. If this is true, then removing feedback on aptitude tests would be an important intervention for increasing grades.

The present study demonstrated that women were more affected by feedback on aptitude tests than men. In measures of self-efficacy, belonging, expected effort, and anticipated grade,

there were significant differences in scores between feedback condition in women but never in men. This follows the literature that demonstrated women experience lower self-efficacy and belonging in STEM (Blaney & Stout, 2017). It is known from the present study that feedback affects self-efficacy, belonging, expected effort, and anticipated grade. So, a possible way to understand this is that women's academic self-perceptions were changed when they received feedback; then, they felt the class would require more effort and anticipated a lower grade. On the other hand, men did not experience these effects after receiving feedback on the aptitude test. These results may demonstrate that professors who administer beginning-of-semester aptitude tests may be discouraging women in their classes, contributing to the gender divide in STEM. A future study with an intervention to remove feedback on class aptitude tests should be conducted to determine if this is the case in a class setting.

The present study had strong internal validity, as it used a between-subjects design, so participants were unaware of the various feedback conditions until the debrief. Along with that, questions were designed to measure the intended outcome directly, such as the expected procrastination questions and anticipated grade. Further, validated scales were used to measure other dependent variables, such as self-efficacy, belonging, and expected effort.

Limitations and Future Directions

As previously discussed, the present study was limited by its ecological validity. The study described a hypothetical class that participants likely did not anticipate taking unless they were already interested in the subject. Because of this, results demonstrated what students think they may do or feel rather than what they would actually do and feel. Further research must be conducted in an ecologically valid environment to confirm the results of this study. A future study could involve an intervention wherein students in one section of a STEM class receive

feedback on an aptitude test for the class, and one section does not. Then, procrastination behavior, self-efficacy, belonging, expected effort, and anticipated/actual grade will be measured at the course's beginning, middle, and end. A study with this methodology will yield results affirming or refuting those of the the present study. Students in the intervention condition are expected to have increased belonging, self-efficacy, and anticipated grade, and decreased expected effort. It would be interesting to see if expected effort continues to mediate the relationship between feedback and procrastination after a class has begun.

Implications

Despite the limitations of this study, results created insight into the relationship between student's self-perceptions within a course and their expectation of performance in that course. Specifically, feedback from an aptitude test had an effect on students' expected effort, which altered their planned procrastination behavior. This study also demonstrated that receiving the results of an aptitude test at the beginning of a class, before the professor has taught the material, can result in an altered perception of a student's self-efficacy, belonging, anticipated grade, and expected effort. All of these variables, in theory, should not be affected by the results of a test that does not reflect any learning a student has done in a class, but these data suggest that aptitude tests do impact outcomes. These results can lead to new interventions to boost academic self-perceptions and performance in classes, specifically for students at increased risk for low self-efficacy and belonging.

References

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review, 84, 191–215.

Bandura, A. (1997). Self-efficacy: The exercise of control. New York: W.H. Freeman and Company.

Blaney, J., Stout, J. (2017). Examining the relationship between introductory computing course experiences, self-efficacy, and belonging among first-generation college women. *Proceeding of the Conference on Integrating Technology into Computer Science Education*, 69-74.

Bouffard, T., Bouchard, M., Goulet, G., Denoncourt, I., Couture, N. (2005). Influence of achievement goals and self-efficacy on students' self-regulation and performance. *International Journal of Psychology*, 40(6),373-384.

Fernie, B., Bharucha Z., Nikčević, A.V., Spada, M. M. (2017). The Unintentional Procrastination Scale. *Journal of Rational-Emotive and Cognitive-Behavior Therapy*, 35(2), 136-149.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*, 2nd Edition. New York: The Guilford Press.

Herrmann, S., Adelman, R., Bodford, J., Graudejus, O., Okun, M., Kwan, V. (2016). The effects of a female role model on academic performance and persistence of women in STEM courses. *Basic and Applied Social Psychology*, 38(5), 258-268.

Ito, T. A., McPherson, E. (2018). Factors Influencing High School Students' Interest in pSTEM. *Frontiers in Psychology*, 9: 1535.

Klassen, R. M., Krawchuk, L.L, Rajani, S. (2008). Academic procrastination of undergraduates: Low self-efficacy to self-regulate predicts higher levels of procrastination. *Contemporary Educational Psychology*, 33(4), 915-931.

Komarraju, M., Nadler, D, (2013). Self-efficacy and academic achievement: why do implicit beliefs, goals, and effort regulation matter? *Learning and Individual Differences*, 25, 67–72.

Lewis, K. L., Stout, J. G., Pollock, S. J., Finkelstein, N. D., Ito, T. A. (2016). Fitting in or opting out: A review of key social-psychological factors influencing a sense of belonging for women in physics. *Physical Review Physics Education Research*, 12(2), 020110(10).

Smith, J. L., Lewis, K. L., Hawthorne, L., Hodges, S. D. (2012). When Trying Hard Isn't Natural: Women's Belonging With and Motivation for Male-Dominated STEM Fields As a Function of Effort Expenditure Concerns. *Personality and Social Psychology Bulletin*, 39(2), 131-143.

Solomon, L. J., Rothblum, E. D. (1984). Academic Procrastination: Frequency and Cognitive-Behavioral Correlates. *Journal of Counseling Psychology*, 31(4), 503-509.

Sotardi, V. (2022). On institutional belongingness and academic performance: mediating effects of social self-efficacy and metacognitive strategies. *Studies in Higher Education*, 47(12), 2444-2459.

Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin, 133*(1), 65-94.

Tellhed, U., Bäckström, M., Björklund, F. (2017). Will I Fit in and Do Well? The importance of social belongingness and self-efficacy for explaining gender differences in interest in STEM and HEED. *Sex Roles*, 77(1-2), 86-96.

Van Dinther, M., Dochy, F., Sergers, M. (2011). Factors affecting students' self-efficacy in higher education. *Educational Research Review*, 6(2), 95-108.

Wäschle, K., Allgaier, A., Lachner, A., Fink, A., Nückles, M. (2014). Procrastination and self-efficacy: Tracing vicious and virtuous circles in self-regulated learning. *Learning and Instruction,* 29, 103-114. 10.1016/j.learninstruc.2013.09.005

Wilson, D., Jones, D., Bocell, F., Crawford, J., Kim, M., Veilleux, N., Floyd-Smith, T., Bates. R., Plett, M. (2015). Belonging and Academic Engagement Among Undergraduate STEM Students: A Multi-institutional Study. *Research in Higher Education*, 56(7), 750-776.