

Thematic analysis of students' perceptions of grading practices in physics

J. Caleb Speirs^{*,†}, W. Brian Lane[†], and Naomi Laird

Department of Physics, University of North Florida, Jacksonville, Florida 32224, USA



(Received 28 August 2025; accepted 23 January 2026; published 5 February 2026)

In the world of science, technology, engineering, and mathematics education practice, alternative grading strategies have been implemented and disseminated with the goal of reinforcing student outcomes like productive reasoning, motivation to learn, and sense of identity. However, while PER has successfully improved these outcomes through reformed instructional strategies, much less work has investigated how grading practices influence these outcomes. This dearth leaves PER without a framework readily available to formulate or interpret studies of alternative grading practices. To help address this gap, we surveyed $N = 94$ students from introductory algebra-based and calculus-based physics courses, asking free-response questions about their experiences with and perspectives on grading in physics. We thematically coded their responses to identify themes in what they said grades represent (such as their effort invested in the course or their mastery of course content) and what effects they said grades produce (such as stress or an adversarial relationship with the instructor). We conducted focus groups to member-check these themes with the students. For themes about what they said grades represent, we further identify whether the student framed this representation positively (“Grades should represent my effort”), negatively (“Grades should not represent the features of a particular assignment”), or neutrally or in mixed terms. We particularly attend to whether students indicated that grades produce an adversarial relationship with their instructor, finding that a greater percentage of these students positively spoke of grades as representing improvement in their performance than among students who did not mention an adversarial relationship. We discuss how these themes offer insights into how we might further research alternative grading strategies, how students’ perspectives on grades might influence the broader goals of PER, and implications for physics instruction.

DOI: [10.1103/pyvm-4s9t](https://doi.org/10.1103/pyvm-4s9t)

I. INTRODUCTION

I live by the grade I die by the grade. One of the greatest pleasures in my life is doing better than others. Pass fail isn’t rigorous enough. College is something to be proud of.

While perhaps hyperbolic, this comment from a student illustrates how centrally grades feature in how students think about their physics classes. Grades communicate expectations and can induce a sense of competition, high stakes, and accomplishment or disappointment. They can also motivate students to study harder and they reward

specific study behaviors. In other words, the outcomes of our grading systems can either further or impede our efforts to reform the physics learning experience, and therefore warrant our attention as physics education researchers.

Physics education researchers have studied how teachers and TAs grade [1,2], investigated *post hoc* how changes to grading schemes would have impacted students’ final grades and DFW rates (the percentage of students who receive a D or F, or who withdraw from the course) [3,4], and have begun to implement “grading for growth” alternative grading strategies [5,6]. Researchers have also investigated how gender and other identities impact grades in a physics course [7], but few PER studies have investigated the reverse: how do grades impact students’ experience in the classroom?

In this paper, we consider the topic of how students perceive grades in physics. Through thematic analysis of students’ answers on a free-response survey, we investigate this underexplored topic by addressing the following two research questions.

RQ1: What do students say about the *meaning* of grades in an introductory physics course? For example, when a student receives a grade of B in a physics course, what do they think this letter grade signifies about their learning?

*Contact author: j.caleb.speirs@unf.edu

†Contact author: brian.lane@unf.edu

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

RQ2: What do students say about the *effects* of grades in an introductory physics course? For example, when a student receives a grade of B in a physics course, what impact does this have on their stress level or motivation to learn?

The answers to these questions inform further research into how student experiences of grades in a physics course impact their learning of the physics content, how we can best modify grading practices to highlight the goals of research-based physics instruction, and how best to communicate these efforts to physics students. In Sec. II, we review the existing research literature on grading practices in physics and science, technology, engineering, and mathematics (STEM) more broadly. In Sec. III, we present the survey that we administered to $N = 94$ students and describe how we identified themes in students' responses. In Sec. IV, we present the themes we identified about what meaning students said grades represent and what effects they said grades produce, and a count of the number of students who expressed each theme in their survey responses. We disaggregate this count by the particularly striking theme of how grades contribute to an adversarial student-instructor relationship. In Sec. V, we discuss how our findings offer insights into alternative grading research, explore how students' views of grades might impact the broader goals of PER, and offer suggestions for grading practices. In Sec. VI, we present our conclusions, discuss limitations of our study, and outline future directions that grading-related work in PER might take.

II. BACKGROUND ON GRADES AND LEARNING

Grades influence how students approach learning in a course, with outcomes that run contrary to the goals of improving the learning experience as sought within PER. For example, a grading orientation rather than a learning orientation has a negative effect on student interest and motivation [8,9]. Also, exam grades can contribute to changes in concern about social comparison, which was demonstrated to feed back into future exam scores [10]. An emphasis on grades tends to decrease students' intellectual risk-taking [11,12] and reduces quality of thinking and learning [13–17]. Grades also reduce student collaboration [18,19] while also discouraging student autonomy [11]. Finally, PER has established that instructor values and norms set a tone for the learning environment of the classroom [20,21]. Grades contribute to this tone by promoting a transactional relationship between instructor and students [22,23], working against the instructor's efforts to support learning in their classroom.

Specifically within PER, a study by Elby [24] showed that physics students hold a set of ideal study habits to support learning but engage in a different set of distorted study habits to obtain a good grade in physics. These distorted study habits are typically considered unproductive in PER [25], with more focus on formulas and rote

memorization and less focus on concepts and real-life applications. As we argue below from the data in our study, these influences from grading practices could undercut the impact of adopting research-based instructional strategies (RBISs) that so much PER work has been devoted to promoting [26,27].

In interviews with physics instructors, Henderson *et al.* [1] identified a gap between the grades that physics instructors assign and the goals they have for the graded assignment. While these instructors valued students' demonstrating their reasoning, they also described a process of “project[ing] correct thought processes onto a student solution when the student does not explicitly describe his/her thought processes.” They describe instructors as grading with a “burden of proof” on the instructor or on the student when determining an assignment grade. Such differences can send inconsistent messages to students about what aspects of learning are valued in a physics course.

These outcomes of traditional grading strategies contribute to inequitable outcomes of physics instruction. For example, Malespina and Singh [28] found that while test anxiety is mediated by self-efficacy with an approximately equal effect for female and male physics students, female physics students tend to start with lower self-efficacy, corresponding to higher test anxiety. The course structures within which grades are determined (e.g., percentage scores versus qualitative binning, relative weight assigned to formative and summative assessments) can create or eliminate gaps in grade-related inequities experienced by women or racial minorities [3,29,30]. These observations support an argument for instructors to adjust course grades to give less importance to traditional exams and more importance to formative assessments such as clicker questions and homework [3,28].

In response to these outcomes, many instructors in physics and beyond are adopting alternative grading strategies (AGSs), defined broadly as any approach to grading that expands beyond scoring individual assignments and calculating a weighted average of those scores. In their scoping review of AGS studies in STEM education research, Hackerson *et al.* conclude that research shows that AGSs are promising but the research lags behind adoption [31]. They highlight that studies of AGSs are particularly sparse within PER, and grading studies tend to only cite papers within their own academic discipline. They identify three significant gaps in the literature: A knowledge gap (a “lack [of] necessary empirical research investigating the impacts of alternative forms of grading on learning in undergraduate STEM courses”) in “the impacts of alternative forms of grading on learning in undergraduate STEM courses”; a methodological gap (“little variation in methodological approaches”) in research practices and in implementation of AGSs; and a theoretical gap (“a lack of theory underlying the research in a given area”)

in the frameworks that grading-related papers identify or describe.

These gaps are present in education research in general and in physics education research in particular, where our efforts to improve student learning through reformed instruction could be mitigated by the impacts of grading practices on how students learn and think about physics. In order to close these gaps in the PER context, we first need to understand a baseline of how students think about grades in physics courses. Doing so will not only help those looking to transform their physics courses with AGSs but also give researchers and instructors a basis for understanding how grades transform the way that students think about physics. Here we explore how physics students think about grades. Such considerations give researchers insights into how grading strategies might produce these learning-related outcomes, and give instructors insights into how to effectively communicate their grading strategy to students.

In answering our research questions, this paper addresses each of these gaps by demonstrating the use of reflection, interpretive tasks, and hypothetical grading experiences (methodological gap); giving an initial view of how students might think about grades in physics classes (knowledge gap); and outlining dimensions along which students think about grading practices (theoretical gap).

III. METHODOLOGY

We administered a survey with free-response questions to $N = 94$ students in algebra- and calculus-based introductory physics at a mid-sized state university. The courses represented here were mostly traditionally taught (separate lecture and lab sections) and were all traditionally graded (weighted average of assignments, participation, and exams). The lectures had varying degrees of active learning implementation: most had some form of “talk to your neighbors” questions, some were partially “flipped” where lecture videos and readings were expected to be digested before class, and class time was devoted to some interactive small-group problem solving. The survey questions (SQs) are presented below.

1. When you think of grades, what comes to mind?
2. What emotions do you think grades make students feel in a physics course? Explain.
3. Think of a course you took where you felt the grading system was unfair and/or did not support your learning. What stands out to you about this course’s grading system?
4. In a physics course, what do you think a course grade of B means? What do you think it should mean?
5. Think of a course you took where you felt the grading system was fair and supported your learning. What stands out to you about this course’s grading system?
6. Suppose you had a choice in how physics instructors weighted students’ understanding of the material

(for example, the degree to which you “got the right answer”) and their effort in the course (for example, the degree to which you “did all the work”) in calculating the final letter grades. How would you prefer the instructor to weight understanding and effort?

7. Suppose you were in a physics class where the assignments, tests, etc., had comments and feedback from the instructor but no numerical scores. How would this impact how you study in a physics course? What sort of comments and feedback would you like the instructor to provide so that you would know how you are doing in the course as well as what you needed to do differently?
8. Suppose you were in a physics class where the assignments, tests, etc., had comments and feedback from the instructor but no numerical scores. What would be a good way for your instructor to determine your final letter grade at the end of the semester?
9. What else would you like physics instructors to know about your experiences with grading?

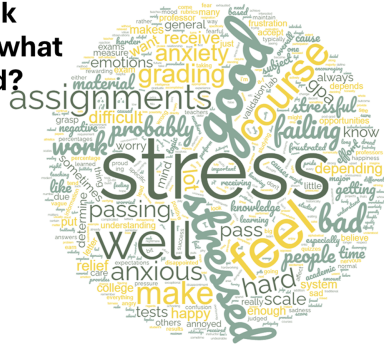
All SQs were free-response, except SQ6, which provided a set of numerical sliders.

We collected survey responses online near the end of the spring semester to ensure that students had sufficient experience with the grading practices employed in their physics courses. We discussed these responses with undergraduate members of our PER group to identify ideas that were frequently expressed, and to understand what common concerns or experiences they represent from the student perspective. Through this discussion, we identified five takeaways that we wanted to discuss with students in focus groups to confirm that we understood their responses: Instructors sometimes fail to grade objectively; Grades provide necessary feedback above and beyond comments; Grades are supposed to measure learning; Grades cause stress; Grades should balance effort and understanding. We prepared two discussion slides (Fig. 1) to help organize the focus groups.

We conducted focus groups two weeks after we collected the survey responses. We held these focus groups during the courses’ associated lab sections so that we could focus discussion on groups of 15–20 students at a time. We began each focus group by presenting the two slides shown in Fig. 1 and asked the students to confirm whether our five takeaways accurately represented their thinking and whether they wanted to add any other ideas we should consider as part of our study. In doing so, we followed the example of Muller [32] in ensuring that “the opinions of individual participants were examined by other members of the group” and of Knaub *et al.* in “encourage[ing] students to be honest and offer alternatives if they felt we were focusing on a nonissue” [33]. We video recorded each focus group and generated transcripts of the recordings.

Upon reviewing the focus group recordings and transcripts and reading the survey responses in greater detail,

When you think
about grades, what
comes to mind?



These are our takeaways from your survey responses. Do you think these themes represent your perspectives on grades?

Instructors sometimes fail to grade objectively.

"Some professors have vague answer keys and when they grade they remove too many points because of their own vague answers."

Grades provide necessary feedback above and beyond comments.

"Comments of what wrong equation or step in an exam I took would be helpful, but not enough."

Grades are supposed to measure learning.

"Sometimes you have bad days... An exam shouldn't be the entire reason why you pass or fail."

Grades cause stress.

In a physics course, students feel "stress or pride, depending on what the grade is."

Grades should balance effort and understanding.

"Personally, I liked this class's grading style. The professor focused more on your effort and less on getting the exact answer."

FIG. 1. Slides used to guide discussion during focus groups, two weeks after survey data were collected.

we concluded that the survey responses adequately captured the students' views and that we understood what the students had communicated in these surveys. We found that the focus groups did not offer additional insights into student thinking about grades beyond what we found in the surveys, so we proceed here with focusing on the students' survey responses, which allows us to code responses with the student as the unit of analysis.

Reviewing the survey responses in greater detail, we thematically coded each student's responses based on what qualities they said grades *represent* and what effects they said grades *produce*. For example, a student might say that grades represent their mastery of course content or their level of effort toward completing coursework, and they might say that grades produce stress or an adversarial relationship with the instructor. For statements about what grades represent, we further identified whether each student described this representation in positive terms, negative terms, or neutrally or in mixed terms. For example, a student might say that grades represent effort, and that this is a fair approach since not everyone is good at physics, or a student might say that grades represent their natural intelligence and identify this as unfair.

As an example of this coding process, consider these two responses one student provided to SQ3 and SQ5. In brackets, we insert the code we identified.

In [an example of an unfair grading system], students had three exams and homework assignments due twice per month. Students were graded on the correctness of their submissions for both exams and homework. However, the homework did not correlate with lecture materials, so it was frustrating being graded on it. The exams were also short answer and based on the subjective judgment of the instructor [Instructors' Internal Expectations—Negative]. No matter how hard you studied, went to tutoring, asked questions, etc. it always felt like you would never be enough [Adversarial Relationship]. The instructor would always find something wrong with the submissions

[Adversarial Relationship] ... [In an example of a fair grading system,] we were expected to respond to discussion prompts, take three brief short answer exams and create a presentation on a resource topic. So long as students participated at all [Effort-Positive], we would receive at least half credit ... The instructor was always very clear on their expectations [Instructor's Externalized Criteria—Positive] and was happy to help students who needed additional guidance.

From these answers, we identified that this student described grades as representing students' effort (positively), instructors' internal expectations (negatively), and instructors' externalized criteria (positively), and that they described grades as producing an adversarial relationship with the instructor.

A student would be coded as neutrally speaking of a theme if they did not seem to attach negative or positive feelings about it, or if the student spoke of the theme positively in one question and negatively in another. For this reason, in our figures we represent this coding as "neutral or mixed" to remind that there are two categories that we have collapsed to one. For instance, one student responded to SQ3 by writing, "All that mattered was the exam. There were 2 exams and each was 40 percent of your total grade" and responded to SQ5 with, "The teacher cared more about you understanding the material and not too much on the exams. The exam only counted to 10 percent of your total grade." Both of these statements spoke to the *instructor's weighting* theme, with the first describing a negative experience and the second a positive experience. So overall, this student's survey responses were coded as "Instructor's Weighting—Neutral or Mixed." The students' survey responses and focus group comments did not show notable differences between a neutral framing and a mixed framing. For example, a student who gave neither positive nor negative framing to a theme might have mixed feelings about the theme, or a student with mixed feelings about the theme might describe themselves as neutral toward it. Therefore, we opted to keep these two framings as one neutral or mixed category.

In our coding scheme, we made the individual student the unit of analysis, rather than individual responses, since a given theme might (and did) occur across responses to different questions. So, for example, in the quote above, there are two statements that speak to an adversarial relationship, but this theme would only be counted once for this student. Additionally, coding at the student level allows us to consider how two or more themes might be related. For example, in response to SQ3, this student described internalized instructor expectations negatively and, in response to SQ5, they described externalized criteria positively, which helps us contrast these themes and consider how they might contribute to or lessen an adversarial relationship with the instructor. Our deidentified textual data with thematic coding are available on request.

IV. RESULTS

Our results include two primary components: the identified themes, which give us a set of dimensions over which students think about grades, and a count of the number of students who expressed each theme, which tells us how prevalent each of these themes was across the sample.

When identifying themes in responses to the survey questions, we noted that students primarily communicated their ideas about what qualities of themselves or the course that grades *represent* and what effects grades *produce*. We group our themes into two large categories accordingly. We first give examples of what students said grades represent, and then examples of what students said grades produce. When presenting quotes of student responses, we indicate which survey question (numbered in Sec. III) the student responded to. We have edited student responses for clarity.

A. Themes: What grades represent

In answer to the research question, “What do students say about the meaning of grades in an introductory physics course?” we found several themes in student responses, which are described in Table I. These themes incorporate both what students feel grades represent in their experience and what they feel grades *should* represent in a physics course.

1. Effort, mastery, and improvement

Students identified three main components of what a grade should represent and described the ideal grade as a balanced combination of these three components: effort toward course activities, mastery of course content, and improvement in performance over time. Also of importance is the flexibility to be awarded credit if you are “on the right track” in problem solving, even if you ultimately arrive at the wrong answer. We surmise that this idea might stem from an emphasis on process over answer in the physics courses represented here.

Many students felt that they should not be expected to “master” the content as they were not intending to be

physics majors or pursue physics-centered professions. Instead, they felt their understanding should be evaluated against a baseline of fundamentals rather than a fluency that expert physicists achieve.

For example, one student said (SQ4)

A [physics grade of] B should be someone who understood the basic content but isn't really able to apply it beyond a narrow range of situations. They might be able to reliably solve an angular momentum problem about a spinning disk, but they wouldn't be able to use angular momentum to explain Kepler's second law.

In this quote, the student is categorizing tasks they are expected to carry out in a physics course (solving a quantitative problem and constructing an explanation for a new phenomenon) and connecting those tasks to expectations for a grade.

In response to SQ7 about a hypothetical course that did not have any numerical grades, just written feedback, a student answered,

I feel like if you are not going to ever use physics—i.e., health majors—that it would work, but actual physics or chemistry majors that will use physics later in their life or jobs should probably have a scale of how much they mastered the material, i.e., grades.

Another answered in response to SQ9,

[Graders should] understand that not all of their students are physics majors [and that] not all schools require or even teach physics.

This student could be referring to other universities that offer similar life-sciences programs but do not view physics as integral to those programs. This would indicate that views on what grades should represent are dependent on the student's future goals and associated professional identity. The student might also be referring to high schools that do not offer physics courses. In that case, the student here is seeing their grade in the university physics course as dependent on their prior knowledge, not on any knowledge developed within the course. In either of these cases, however, the student seemed to feel that the expectation of understanding the material needed to be balanced against the relative prevalence of physics in their prior experience and future career.

2. Instructor expectations

Students also felt that, too often, their grade depended on uncommunicated expectations held by the instructor and they expressed a desire for clear and upfront expectations in problem solving using rubrics.

TABLE I. A summary of themes identified among student responses about the meaning of grades.

Grades represent ...		
Theme	Description	Example
Effort made toward learning	The student refers to a grade reflecting good-faith effort to learn material or complete an assignment.	"A good grade means I put in enough focused time working to understand the material. If I don't put in this work, I wouldn't get a good grade."
Improvement over time (e.g., exam retakes)	The student refers to a grading system accounting for an increase in the quality or quantity of their understanding or performance between different points in the semester.	"I got a B in this class because I learned from feedback on the mid-term and improved my performance on the final."
Mastery of course content	The student refers to a grade reflecting their demonstration of sufficient understanding of a topic or fluency with a practice.	"A grade tells you how much physics you understand and how well you can apply it to different problems."
Progress toward the right answer on a specific assignment (e.g., partial credit)	The student refers to the score on an assignment reflecting the presence of some correct elements that could lead to a correct final outcome if mistakes were avoided.	"A good grade on an exam question shows you are on the right track and used the right process to get the answer even if you may not get to the correct answer in the end."
The student's domain-specific abilities, usually with physics	The student refers to a grade reflecting their innate fluency with domain-specific concepts or academic contexts, such as physics content or test taking.	"A high grade in a physics class shows that you are good at physics."
Features of the assignment (structure, problem context)	The student refers to a grade reflecting elements of an individual assignment instead of their understanding or ability.	"I can get a good grade by cramming for each exam without learning, as long as I look out for trick questions."
The student's innate abilities (usually intelligence)	The student refers to a grade reflecting an innate capability they possess that transcends disciplines or academic contexts, usually referring to some form of intelligence.	"Grades are how a professor judges your intelligence, so if you're smart you'll get an A."
The student's circumstances	The student refers to a grade reflecting elements of their life outside the class, usually outside of their control.	"I had a bad day before the test and that made my grade go down."
Timely feedback	The student refers to feedback from the instructor in terms of frequency and/or interval after assignment submission.	"I would want the instructor to clarify their expectations early and frequently."
Numerical feedback	The student refers to a numerical score or rating on an assignment as a means of feedback.	"A numerical scale shows me how close I am to passing or failing. It gives me a scale to show how much I'm picking up from the course. Otherwise, I have no way of knowing."
Feedback about course standing	The student refers to a grade as indicative of their progress in the course.	"My grade shows me how I'm doing in the class and how much progress I've made."
The instructor's externalized criteria	The student refers to grades as shaped by the criteria expressed for the assignment, usually in class or on a rubric.	"Having a detailed rubric makes my grade more fair and helps me know what to do differently."
The instructor's internal expectations	The student refers to grades as shaped by the instructor's internalized expectations for students (usually different than expressed criteria).	"My exam grade was low because I didn't know what the professor was looking for based on what they talked about in class."
The instructor's weighting	The student refers to grades as shaped by the percentage awarded to assignments in the syllabus.	"I appreciate a class where things are equally weighted because when it's just exams I have a hard time recovering from a bad exam score."
A gain or loss of points or standing.	The student refers to a grade as increasing or decreasing over time due to high or low performance, usually recorded in an online gradebook.	"I didn't do well on the first exam and my grade dropped significantly. I did better on the second exam and my grade went up a full letter."

One student articulated this idea (SQ1 and SQ3) as

I think [grades] are completely ok, it's just some professors have vague answer keys and when they grade they remove too many points because of their own vague answers. ... Alg physics with Dr. [X], grading is extremely unfair and also [they] provide no useful practice in the form of homework that relates to the class for [their] students to succeed ... It's not the grading that is the issue in physics classes. It's the professors and them not providing useful homework problems.

Through other responses and the focus groups, we interpreted the statement about providing useful homework problems to be related to exam problems not being similar enough to the homework. The PER community has discussed pedagogical practice (such as what to assign for homework, how to connect it to what goes on in lecture or lab, *etc.*) extensively. Here, we see students indicating that grading is being affected by pedagogical practice. Even though the student above says that the grading is not “the issue,” the student indicates that “professors remove too many points because of their own vague answers,” and that the grading “is extremely unfair,” indicating dissatisfaction with how the grading is interacting with the pedagogical practices (such as vague answers) implemented in the course. We surmise that the tension between the expectations of the student compared to the instructor may in part be coming from the relative newness (to the student) of the active-learning pedagogical technique employed in the course. This student also said (SQ9), “The flipped classroom experience is the worst idea for a class.” For this student, then, the pedagogical technique interacted negatively with the grading experience, and this was perceived as an internalized, uncommunicated expectation from the instructor.

Another student articulated (SQ3)

One course I took where the grading system was unfair was where it was all free response questions, and the professor graded based off what he thought at the time it deserved. There was no rubric, or certain qualities he was looking for, just if he liked your answer or not. I think a fair grading system is where you can easily see why you got points deducted for your answer based off a rubric. When strict rubrics are used for grading, there's no room for subjective deduction.

Another student praised the rubrics used in an introductory physics course (SQ5):

Dr. X's intro to physics course [had fair grading]. We were given a set rubric of items to show [what] we could do rather than just being graded on correct answers. Especially when some physics problems are like 10-step processes.

Another student compared rubrics in a literature course and physics, noting that physics grading felt more subjective than the literature course (SQ5):

[A fair grading system was in] English class. There's always a reference paper that satisfies all the rubric and they get a 100. Physics doesn't do that and the grading feels like arbitrary.

Earlier (SQ3), this student described that “[in] physics, when explaining something I have to talk like the graders are babies. That's hard sometimes.”

3. The instructor's weighting

Another key feature students described in fair physics grading systems was balancing the weighting of the final grade across exam and nonexam components. Students frequently described courses that are “heavily weighted on tests and that's it,” and linked this to “increased anxiety outside of the assessments.”

Wanting a more distributed weighting for a course seems related primarily to anxiety about the final grade. For instance (SQ3):

The thing that stood out about this course's [unfair] grading system is that it all depended on the exams and the quizzes we had in the class, I would have liked some homework assignments since they would have worked as grade “insurance.”

Some students, however, also articulated that they felt exams did not represent their understanding. For example (SQ3):

I feel that courses that rely heavily on test grades are unfair. I think it should be more about the effort and the understanding instead of only focusing on how well the person performs on a test.

B. Themes: What grades produce

In this section, we discuss several themes found among the students' responses in answer to the research question, “What do students say about the *effects* of grades in an introductory physics course?” In other words, what do grades produce?

1. Stress

We found that students primarily talked about the stress of grades in responding to our survey questions. While stress could be talked about in a neutral sense (that is, without necessarily attaching a negative emotion to the word) and two students even described grade-related stress positively (an important motivation to study), we found that the majority of mentions of stress carried a negative connotation.

For instance, in response to the question “What emotions do you think grades make students feel in a physics course?” (SQ1), one student wrote simply

Stressed, pressured, angry.

Students related this stress in physics courses to the perception that physics is difficult to master. For instance, “*Physics is not an easy subject for me so grading is especially stressful*” (SQ2) or “*Stressed out. Physics is hard and people feel like they fail*” (SQ2). One student linked this to exam scores in physics courses (SQ2):

I think students learned to feel intimidated by physics, especially when we hear that the average grade on exams is in the range of 50%–60%.

The distributed weighting between exam and nonexam components in the course also can play a role in students’ emotions about grades:

Anxiety, just having to try and plug in different numbers to see how my lab will affect my overall [physics] grade as well as a little bit of relief because the lab can help my grade. (SQ2)

2. A positive (and negative) learning environment

Despite the stress, some students (24%) also articulated that grading helps them learn and creates a positive learning environment. Students articulated that grades motivate them to study harder for physics, and a grading system without numerical scores would cause them to study less. It seems that this effect is drawn from two main sources: (1) the desire for a higher grade motivates them to study, and (2) numerical scores give students feedback that they find clearer than written comments. For example, these three students said (SQ7):

[A class without assignment grades] would make it very difficult to study at all. I tend to base study effort on where I need to be to get a B or A depending on the course. Comments of what wrong equation or step in an exam I took would be helpful, but not enough.

I feel like scores help drive me to do better, I do not want to wait to see whether I pass or fail, I want to see how much I have to do to get a certain grade.

I may actually put in less effort, because regardless of my actions as long as I do meet expectations and am at least competent, I will just need a confirmation I did okay to be satisfied. It wouldn’t be worth it anymore to go above expectations to get an A.

Students also described their desire for numerical grades as driven by interactions with other courses. For example, one student wrote (SQ7):

I would worry I may not prioritize studying for a test that would not receive a grade if there were conflicts with other classes [such as exams in close temporal proximity] as much as I would if it were graded. If it were just a test being handed to use without a grade associated to it, I feel it would be treated more as a learning tool as opposed to a performance evaluation. This could be a good thing in terms of learning but the added motivation/stress of having to perform would increase the necessity to study more.

In this comment, this student seems to understand the difference between formative and summative assessment, and also seems to draw a distinction between *learning* and *studying to perform* as separate activities.

Others (34%) indicated that numerical grades contribute to a *negative* learning environment for them. When asked about a course where no numerical grades were given, one student wrote (SQ7)

Yes please do it. Oh my god do it. It would alleviate so much stress, numerical grades have ambiguity. Pass/Fail with a good professor does not have ambiguity. I’d study more if it meant the learning would be on my own accord.

Here are some other student quotes that indicated grades as producing a negative learning environment.

(SQ5): I enjoyed that there was a kind of curve and I was not stressed about failing. This made it easier to actually learn the information rather than just memorizing it.

(SQ7): [Taking away points from grades and just having comments] would take away the pressure of getting a certain “number” and let you focus on learning and understanding.

(SQ2): Grades foster negative emotions and lead to a disdain with physics rather than leading to learning.

(SQ5): [An unfair grading system was one where I] didn’t learn anything since you focus on trying to get a good grade.

From these comments and others, we saw that students that said grades produced a positive learning environment focused on the effort of studying, whereas those that said grades produced a negative learning environment focused on their learning. It appears that these two ideas, *learning* and *studying*, are different for students, even if they are interrelated activities (as in “*I’d study more if it meant the learning would be of my own accord*”). It also seems apparent that a student’s views on whether grades are helpful to learning depends on (as of yet) unknown characteristics of the individual student—i.e., our sample did not present a strong majority opinion on this issue.

3. An adversarial relationship with instructor

In addition to a negative learning environment (which describes the course overall without necessarily pointing to the instructor), we also noted when students described grades as contributing to an adversarial relationship with the instructor. Some adversarial language was about classroom behavior, as in the following response to what types of comments the students would want from instructors.

(SQ8): I think with feedback and comments that are not pounding on students because of their performance should be enforced.

The metaphor of “pounding” used by this student indicates the view of the professor as physically attacking them.

As another example, one student wrote (SQ3):

In my Physics I course, the homework only allowed 3 attempts and then locked out the problem ... This is basically a hostile design that disincentivizes learning and mastery, it only rewards students who already understand.

In this comment, the student attributes a motivation (hostility) to the instructor behind the grading practices. They describe an intentionality that goes beyond an unfortunate byproduct for the course environment.

Other students articulated an adversarial relationship using terms more explicitly about the instructor and in the form of grievances.

TABLE II. A summary of themes identified among student responses about the effect of grades.

Grades produce ...		
Theme	Description	Example
Pride	The student refers to grades as eliciting a sense of pride, accomplishment, or satisfaction.	“Students are proud of their good grades because it means they made progress.”
Stress	The student refers to grades as eliciting a sense of stress, pressure, or concern.	“I think in most courses, students are probably stressed when thinking about grades.”
Stress (positively framed)	The student refers to grade-induced stress as a benefit to or necessary element of the learning process.	“Having stress about grades makes me study more.”
Stress (negatively framed)	The student refers to grade-induced stress as a hindrance to the learning process. Sometimes framed as anxiety.	“I have a hard time learning when I am stressed so I just try to memorize the answers for a test.”
Embarrassment	The student refers to grades as eliciting a sense of embarrassment, guilt, or shame.	“If a grade is a C or lower, students could feel stupid.”
A positive learning environment	The student refers to grades as part of a beneficial or positive learning experience.	“Because this grading system was fair and the professor was helpful, you could focus on improving your performance instead of memorizing for the test.”
A negative learning environment	The student refers to grades as part of a detrimental or negative learning experience. Beyond stress, about the course overall.	“Once I got a second bad test grade, I stopped trying in the class.”
An adversarial student-instructor relationship	The student refers to grades as creating or exacerbating a detrimental relationship between student and instructor.	“The exams had a lot of trick questions because the professor wanted to see how many of us would fail.”
Motivation through comparison	The student describes being motivated by comparisons between their grades and classmates’ grades, possibly in competition.	“I don’t like pass fail systems because I need to know how I’m doing compared to my classmates to know if I need to study more.”
Motivation through signaling to third parties	The student describes being motivated by needing a certain grade to communicate a proficiency to a third party, such as a graduate admissions office.	“Grades in general make me think of my future and opportunities. If my GPA isn’t up to the expectations of employers then my work opportunities drop into the realm of things I’m no longer interested in.”
Motivation through reward	The student describes being motivated by receiving a certain grade as a fundamental reward.	“I want to receive a score with the comments and feedback rather than just words to work harder on the score.”
Interactions between courses	The student describes a trade-off in time or effort between concurrent courses or concern about subsequent course prerequisites.	“I need to prioritize studying for classes that are harder or where the grade is more important.”

- (SQ3): They did not leave any room for communication and were not open to hearing out students if they need help in class.
- (SQ3): The teacher would take no accountability for the class average being low.
- (SQ3): The teacher took off points from an exam because I didn't write the answer on Canvas even though I had submitted the answer on the paper where we were required to do all the work. She purposely told us there wasn't enough time to do so but still took points off.

Again, these comments speak to how grades specifically affect the students' relationship with the instructor rather than the general climate of the course. Given how significantly this experience might affect a student's learning, sense of identity, and well-being, we consider in the next subsection what else the students who expressed this theme tended to say about grades.

Overall, students say that grades in physics produce stress due to difficulty performing well on exams. However, the students are divided somewhat evenly into whether this stress is a motivator or demotivator for studying and learning.

C. Theme counts

We coded each student's survey responses with the themes in Tables I and II to determine the prevalence of each theme across our sample. Figures 2 and 3 show a visualization of the percentage of students who expressed each theme in their free responses. In Fig. 2, expressions about what grades represent were further coded as whether the student spoke of the theme positively (green in the figure), negatively (red in the figure), or neutrally or in mixed terms (gray in the figure).

As seen in Fig. 2, roughly half (46%) of students mentioned effort toward learning as a desirable criterion

for a grade. Improvement over time and mastery were secondary criteria, with roughly 30% of students speaking of each theme in their survey responses. As detailed in another manuscript [34], in SQ6, we asked students to rate their ideal percentages for how effort toward learning and understanding of course content should factor into a student's grade in a physics course. Just over half (51%) of students weighted effort more heavily than understanding, 14% of students weighted effort and understanding equally, and 35% of students weighted understanding more heavily than effort. These numerical responses support our observation of students describing effort as a desirable criterion for a grade.

Figure 3 shows the frequency of themes regarding what effects grades produce. Stress is mentioned in 78% of students' responses, and the majority of students (53%) frame it negatively. There is a roughly equal number of students who describe grades as producing a positive physics learning environment (24%) compared to a negative physics learning environment (32%).

While the number of students indicating that grades produce an adversarial relationship with their instructor was low (19%), we were particularly concerned with how this theme affects students' learning and well-being, so we filtered the student responses to look at the other themes those students articulated (see Figs. 4 and 5). Among those who mentioned an adversarial relationship, the three main contributors to a grade (effort, mastery, and improvement) are strongly represented, with improvement (mentioned by 50% of the "adversarial-relationship" students) being a stronger component than effort and mastery. They also mention improvement at twice the rate of students who did not mention an adversarial relationship (25%, see Fig. 5). It could be, then, that an adversarial student-instructor experience might be related to the extent that the instructor

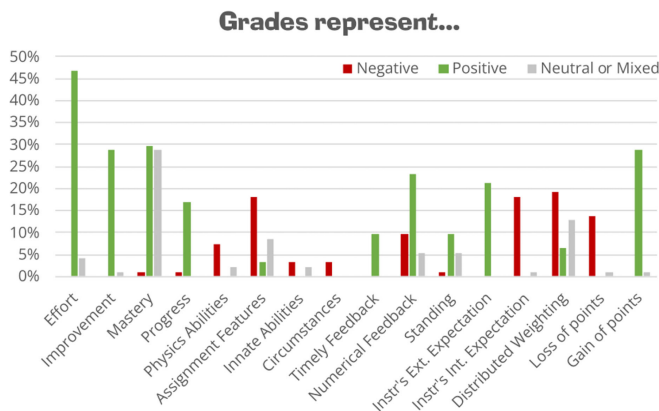


FIG. 2. What do grades represent to students? Percentage of students ($N = 94$) who articulated each theme from Table I with negative framing (red), positive framing (green), and neutral framing (gray).

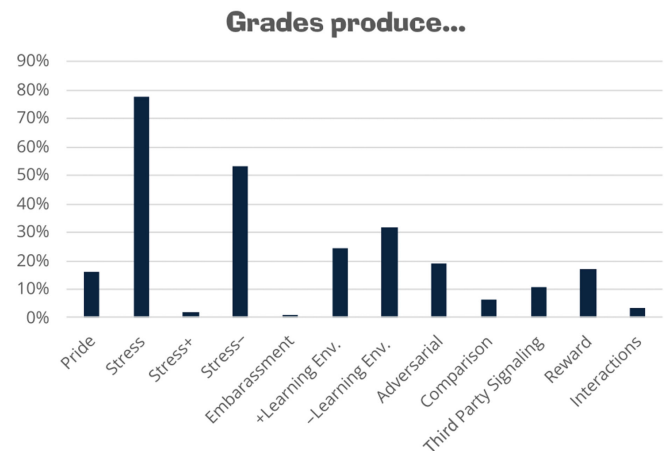


FIG. 3. What do grades produce, according to students? Percentage of students ($N = 94$) who articulated each theme from Table II.

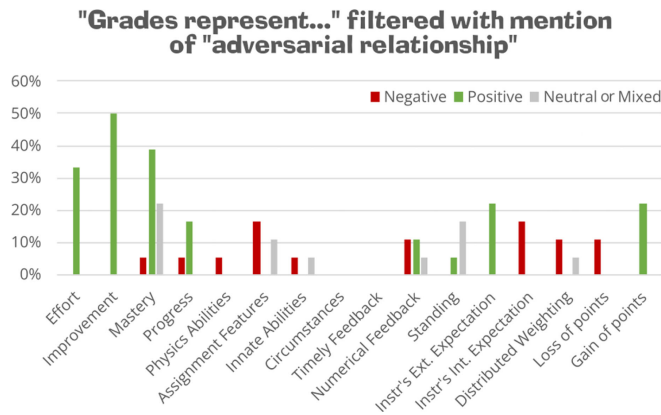


FIG. 4. What do grades represent to students who mention an adversarial relationship with the instructor? Percentages of students ($N = 18$) who mentioned an adversarial relationship and articulated each theme in the figure with negative framing (red), positive framing (green), and neutral or mixed framing (gray).

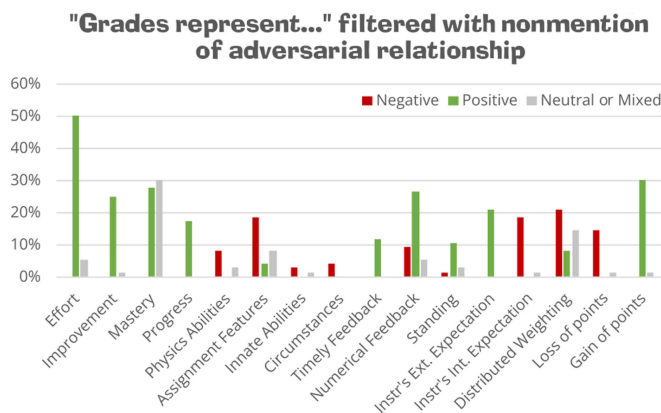


FIG. 5. What do grades represent to students who did not mention an adversarial relationship with the instructor? Percentages of students ($N = 76$) who did not mention an adversarial relationship and articulated each theme in the figure with negative framing (red), positive framing (green), and neutral or mixed framing (gray).

acknowledges a student's improvement over time in their grading scheme.

V. DISCUSSION

In answer to our research questions, we have found that students in an introductory physics course said that their grades in that course signify a combination of their effort, improvement, and mastery; their grades numerically encode feedback about their standing in the course; their grades can reveal a conflict between instructors' internalized and communicated expectations; their grades are subject to instructors' weighting decisions; and that students think of grades as a time-evolving gain or loss of points. These students also said

that grades in their physics course can produce stress (usually negatively), can influence the class environment, and can create or contribute to an adversarial relationship with the instructor. Here, we discuss how these perceptions can inform efforts to study grading strategies, how they might undercut the goals of PER, and how they can inform traditional and alternative grading practices in the classroom.

A. Insights for research into alternative grading

We contextualized this study using a recent call from [31] to address a knowledge gap, a methodological gap, and a theoretical gap in the research literature about AGSs. Although the physics courses that we surveyed these students in did not use AGSs, the themes in students' comments can help inform studies of AGSs in an introductory physics context.

First, we contribute a methodology in the form of various question types (reflection on prior grading experiences, interpretive tasks, responding to hypothetical grading experiences) centered on students' perceptions and experiences. The hypothetical scenarios we posed seem particularly helpful since they can explore students' perspectives on AGSs without running semester-long studies on incremental changes to grading practices. Our survey questions have yielded important insights into students' perceptions of grading practices and can be adapted to specifically explore experiences with a given AGS. The contrasting student perspectives on effort, progress, improvement, and mastery would be particularly of interest in AGSs that favor one or another (such as standards-based grading favoring mastery).

The themes we found in students' responses contribute to the knowledge gap by giving an initial view of how students might think about grades in physics classes. We found a mixture of opinions about grades representing effort, progress, improvement, and mastery. We gained insight into how grades play into a student's relationship with the instructor, particularly when communicated standards do not match internalized criteria. We found that some students value numerical scores as quantitative feedback apart from written feedback, and we saw how students perceive numerical scores as contributing to a gain or loss of points in the course overall.

Finally, these themes present the grounds for a framework that helps the theoretical gap by outlining dimensions along which students think about grading practices. Our themes encourage researchers to consider what an AGS communicates that grades mean and to look for what effects an AGS will produce among students. For example, one might consider how a mastery-heavy focus like in standards-based grading or an effort-heavy focus like in labor-based contract grading might shape the ways in which students engage with physics content. Or one might examine how an AGS could deemphasize students'

attributes, capabilities, and sense of competition with peers and instead support student identity. Or one could try to isolate which aspects of the grading process are most stress-inducing for students and explore how to maintain the assessment process while promoting student well-being.

We especially think research into AGSs should follow up on how grading practices can contribute to a negative learning environment (mentioned by nearly one-third of our sample) and an adversarial relationship with the instructor (mentioned by nearly one-fifth). While AGSs offer promise for student benefits and generate excitement among adopters, researchers and instructors should avoid increasing the degree to which grades are already emotionally charged for students.

B. Implications for the goals of PER

Reflecting on these themes and comments, we offer here the concern that some grading practices might be undercutting some of the goals of reforming physics education pursued in PER. If this is the case, then efforts to introduce RBISs in a course or program could suffer a ceiling effect mitigated by the grading practices in that course or program. Specifically, we highlight potential threats to supporting student reasoning and conceptual understanding; developing student identity; and promoting expertlike learning motivation, learning attitudes, and epistemologies.

A long-standing goal of PER has been to promote student reasoning and conceptual understanding [35–37]. This goal typically leads researchers and instructors to prompt students to explain their reasoning in step-by-step detail. However, our findings show that when this granularity is reflected in grading practices and impacts student scores, students might feel overly burdened to prove their understanding [1], and that this burden might interfere with their ability to receive the feedback that their reasoning and argumentation skills need to be improved. Recall the student who described their difficulty “when explaining something I have to talk like the graders are babies ... the grading [in physics] feels arbitrary.” This student had connected the instructors’ grading focus on logically coherent arguments with a sense of arbitrary point taking rather than with a motivation to improve their argumentation skills. Such comments highlight how assessing students’ reasoning might trigger defensive postures and point-seeking around reasoning rather than help them to develop a transferable skill.

Much work in PER focuses on how students develop physics- or STEM-related identities, and what practices in the classroom best support that development [28,38,39]. Particularly, students’ beliefs about their own competence inform the extent to which they form an identity around physics and an identity as central members of the course community [40]. Student beliefs about their own competence can be impacted by the grades they receive and the

recognition that they feel from their instructors and peers [41,42]. In our study, students reported feeling that grading practices in physics did not always reflect their identity. They felt that expectations of mastery ran contrary to not being physics majors or not pursuing physics-centered professions. They seemed to feel that physics majors and nonphysics majors should be graded differently based on how they expect to use their physics learning after the course concludes. It could be that the expectation of mastery reflected in the grading system could reinforce a nonphysics identity for these students and promote feelings that they do not belong in the course. Again, a course or initiative that seeks to promote student identity could run into a ceiling effect when grading practices induce this sort of response from students.

Finally, PER studies related to student motivation, learning attitudes, and epistemology emphasize the importance of how students think about physics, and not just the amount of knowledge they acquire [25,24,43]. These findings have expanded the ways physics instructors structure learning activities and assessments [44,45]. However, it may be that grading practices send a different message to students. Recalling student comments that grades should be structured based on students’ future aspirations, we see that a grading system misaligned with these aspirations may seem unfair to students, thereby decreasing their motivation to learn in the course. Our students also expressed that written comments without a score attached are insufficient to motivate them to study, and that numerical scores help them to quantify an amount of work they expect to complete. This seems contrary to desired behaviors for determining the value of work (for instance, probing our work for inconsistencies). Also, the mere presence of heavily weighted exams in the course can interfere with formative learning opportunities because, as one student said, students feel “increased anxiety outside of the [exams].” We can likely attribute this disposition to the students’ prior 13+ years of experience with grades, and there may be little a single physics course can do to change it. However, understanding the impacts of grades on student thinking is crucial to fully understanding our students’ experience of learning physics.

One insight our results give is that the students seemed to differentiate between studying and learning (as in “This could be a good thing in terms of learning but the added motivation and stress of having to perform would increase the necessity to study more”). One could interpret this as a distinction between formative and summative knowledge, or process versus product. But especially in light of Elby’s 1999 study [24] showing that students distorted their study habits due to the presence of grades, it is likely that our physics students are having to choose between two options: study to learn or study to perform. Grades would then represent the motivation to perform, rather than the motivation to learn.

C. Implications for grading practices

These findings have implications for both traditional and alternative grading practices in physics, since all grading practices prompt responses from students based on their prior experiences with grading and their expectations of what their physics class will be like.

First, these themes provide a set of dimensions along which instructors can reflect on their grading practices and communicate those practices to their students. It seems particularly important for instructors to reflect on and communicate how their grading practices actualize their values of effort, progress, improvement, and mastery. These values can be difficult to balance in a grading system, and while some popular alternative grading strategies tend to emphasize one of these values above the others, students might be reluctant to embrace such a singular grading focus. Like RBISs, AGSs require buy-in from students to be most effective [46] and these themes give instructors talking points they can use to support buy-in.

Students in this study described experiencing a misalignment between instructors' communicated evaluation criteria (usually expressed in a rubric) and their internalized expectations (usually revealed in feedback returned on an exam). This misalignment suggests instructors should reflect on the degree to which they are sufficiently signaling their grading-relevant expectations to students. Such signaling can be challenging: An instructor wants to fairly and openly communicate expectations to students to inform their preparations for an exam but is obligated to avoid communicating the detailed contents of an upcoming exam. One recommendation is that instructors or TAs publish a detailed rubric that is to be used for assessing student performance on the exam. Such a rubric could focus on aspects of problem solving (such as the Minnesota Assessment of Problem Solving developed by Docktor *et al.* in Ref. [47]), or it could be aligned with particular aspects the instructor is looking for (for instance, logical coherence of arguments).

Students who reported grades contributing to an adversarial relationship with the instructor had a higher rate of positively describing grades as representing improvement in understanding. This trend suggests that adversarial student-instructor relationships could be addressed or avoided by incorporating improvement into one's grading strategy, such as through retesting or awarding points for improved exam performance.

The recommendation to honor improvement through retesting was echoed in Beatty's report of an instantiation of standards-based grading in an introductory physics course [6]. In particular, Beatty reported that "students generally appreciated the 'no history' aspect of SBG, where later evidence completely overwrote earlier evidence of mastery for each standard." He goes on to quote a student

saying, "The standards-based system takes the high-school 'point system' mentality away from the learning experience so that you can focus more on getting the material," and another student as saying "I like that we can reassess if we do poorly on an exam. A lot of the stress of grades comes from doing poorly on one exam and then having to hit a home-run on subsequent exams." Both of these comments echo the issues raised here.

Finally, these findings give insight into how grades are an integral part of a course's design. It is well known that changing a course's design structure can change the course's outcomes [30]. Instructors should be aware that adopting AGSs will impact course outcomes, just like adopting RBISs will.

VI. CONCLUSIONS, LIMITATIONS, AND FUTURE DIRECTION

We find that students describe physics grades as representing their effort, mastery, and improvement. They generally feel positively about effort and improvement contributing to their grade but have mixed feelings about mastery, particularly as it relates to their backgrounds and aspirations. We also find that grades influence students' stress levels and their relationships with instructors. Our findings suggest that grading practices can either support or undermine the goals of physics education reform such as reinforcing positive identities and productive epistemologies. The students seem aware of this impact, describing how they would learn differently without grades as a motivation or pressure. Some of our students' reflections echo motivations in education literature that favor alternative grading, indicating they might be receptive to alternative grading practices if framed appropriately for them. Even without a change to one's grading practices, it seems that both traditional and alternative grading can be better presented to students by focusing on how these practices balance effort, improvement, and mastery, thereby affirming students' values in the course context [48,49].

There are some limitations to our findings. First, we are drawing on a population of students that includes both algebra-based and calculus-based physics courses. Thus, there could be patterns of thought that correspond to one or the other population, and they are currently mixed in our results. For instance, it could be that the calculus-based students showed a preference for effort over mastery, while the algebra-based students indicated more adversarial relationships. Another related limitation is that our study was completed at a single institution, thus limiting the generalizability of our results.

Second, we collected our survey results and conducted the focus groups at the end of the semester but before finals, when grades are particularly salient and stress levels tend to be higher. The focus groups also presented our preliminary findings on a slide for further discussion. While we

explicitly asked students if they had insights that were not represented, displaying these findings could have constrained the emergence of new themes.

Finally, our use of “neutral or mixed” themes was a combination of students who either indicated the theme without a positive or negative framing *and* students who indicated both a positive and a negative framing about the same theme in their responses. It could be that students who feel both positively and negatively about a theme hold a qualitatively different position than a student who is indeed neutral about the same theme, but our coding scheme did not capture these differences. The students’ survey responses and focus group comments did not offer insight into such potential differences, so there may be more unexamined nuance in the data than what we are able to illuminate here.

One key takeaway is that students’ feelings and perspectives about grades are dependent on their prior experiences and values, and their expectations for future experiences such as qualifying for internships or employment opportunities. For instance, while some students praised numerical grades for their objectivity, others argued that grades and grading strategies needed to be more accommodating of students’ backgrounds and, in particular, their strengths.

More nuance can be brought to our understanding of students’ perspectives on grades in physics classes, and one avenue for that research is interrogating the origins and development of students’ perspectives on grades. We surmise that there is a combination of prior experiences, expectations about the course learned from fellow students, and how well the student’s own epistemologies fit those of the physics instructor (*i.e.*, are they expecting to provide detailed arguments for their answer or just give the correct answer?).

One proposal is to create a better fit between the grading scheme and the instructors’ values and epistemologies (as in Ref. [1]), thus harnessing grades to provide cohesive and consistent feedback and motivation toward better physics epistemologies (or put another way, to try to get studying, performance, and learning to be aligned in the student’s view). Since each student will be starting that process in different locations, it would be helpful to study the dynamics of how perspectives about grades in physics courses change over repeated cycles of instruction, assessment, and feedback. We also have found evidence that using grades to motivate epistemologies may backfire (as in “we have to talk to the graders like they are babies”). It is therefore also possible that we will need to reduce the emphasis on grades in general in our physics courses to achieve wide-spread adoption of desired epistemologies.

It seems that on the question of whether grades increase or decrease motivation to learn or study, the students are divided based on personal values and prior experiences, and also seemed to view these activities differently. In the

process of adopting AGSs, then, an instructor would need to attend to the various value systems their students have to create buy-in and understanding of how to learn in the new grading system. Future research would incorporate existing frameworks for identifying and acknowledging students’ values in the classroom. Control-value theory of achievement emotions is one framework that appears helpful in this regard [50,51]. More research needs to be done in how students’ values interact with physics grading systems, particularly for generating models explaining whether and why they are successful within specific grading systems.

Finally, we acknowledge that high levels of stress around grading can potentially create an unhealthy environment for students [52,53,54]. While the majority of students in our study framed stress in the negative, some of the students in our study framed stress as a positive. Frameworks teasing apart productive stress and unproductive stress would be invaluable to the study of the impact of grading on physics students, particularly in cases (such as premed majors or in upper-level physics courses for physics majors) where grades may impact future career opportunities. Many physics instructors wish for a transformative experience for our students, one in which they are challenged to grow. This type of transformation does not come without stress or discomfort. However, we need to be aware of how to produce *productive* stress and discomfort. To differentiate the two would likely require additional research and listening. Hammer’s “meta-affective learning” [55] seems to be a good starting point for this vein of research, but one might also explore Transformative Learning Theory [56].

This work has created an initial framework of dimensions along which students think about grading practices in physics. We outline how this framework can inform research into alternative grading practices in physics, the formulation of alternative grading practices in the classroom, and how grading practices are presented to students. Future studies will explore in greater detail how our grading systems impact students’ perceptions of physics content and physics reasoning, and their sense of identity within that realm of knowledge.

To conclude, we invite readers to consider the following quote from a physics student in relation to the interaction of grades and stress, identity, and physics learning. Our grading practices, if well researched, might be crafted to better appreciate the differences in background and ability our students have, while promoting robust physics learning.

I have had a rough upbringing where my parents associated my worth with my academic performance, and unfortunately working hard didn’t always yield an A like they wanted. I’m getting better with accepting grades for what they are, but I think there’s a disconnect for grading people the same when the said people have different strengths and perform differently regardless of

the effort placed. It isn't the same for everyone and grading should reflect that.

ACKNOWLEDGMENTS

This work was supported by a University of North Florida faculty publishing grant. We are grateful to Suzanne Ehrlich, Terrie Galanti, Elizabeth Gregg, and members of the PER@UNF group for helpful conversations about this

project, and to the physics lab instructors who granted us time with their students to conduct the focus groups.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available because they contain sensitive personal information. The data are available from the authors upon reasonable request.

-
- [1] C. Henderson, E. Yerushalmi, V. H. Kuo, P. Heller, and K. Heller, Grading student problem solutions: The challenge of sending a consistent message, *Am. J. Phys.* **72**, 164 (2004).
 - [2] C. Henderson, E. Marshman, R. Sayer, C. Singh, and E. Yerushalmi, Graduate teaching assistants use different criteria when grading introductory physics vs. quantum mechanics problems, [arXiv:1701.01414](https://arxiv.org/abs/1701.01414).
 - [3] A. B. Simmons and A. F. Heckler, Grades, grade component weighting, and demographic disparities in introductory physics, *Phys. Rev. Phys. Educ. Res.* **16**, 020125 (2020).
 - [4] D. J. Webb, C. A. Paul, and M. K. Chessey, Relative impacts of different grade scales on student success in introductory physics, *Phys. Rev. Phys. Educ. Res.* **16**, 020114 (2020).
 - [5] J. M. Wachter, Labor-based grading practices in the physics classroom, *Am. J. Phys.* **92**, 336 (2024).
 - [6] I. D. Beatty, Standards-based grading in introductory university physics, *J. Scholar. Teach. Learn.* **13**, 1 (2013).
 - [7] S. Cwik and C. Singh, How perception of learning environment predicts male and female students' grades and motivational outcomes in algebra-based introductory physics courses, *Phys. Rev. Phys. Educ. Res.* **17**, 020143 (2021).
 - [8] C. Pulfrey, C. Darnon, and F. Butera, Autonomy and task performance: Explaining the impact of grades on intrinsic motivation, *J. Educ. Psychol.* **105**, 39 (2013).
 - [9] A. Kohn, The case against grades, *Educ. Lead.* **69**, 28 (2011), <https://www.alfiekohn.org/article/case-grades/>.
 - [10] S. Suresh and A. F. Heckler, Evolution of grades and social comparison concern within an introductory physics course, *Phys. Rev. Phys. Educ. Res.* **19**, 010129 (2023).
 - [11] C. Pulfrey, C. Buchs, and F. Butera, Why grades engender performance-avoidance goals: The mediating role of autonomous motivation, *J. Educ. Psychol.* **103**, 683 (2011).
 - [12] J. Condry, Enemies of exploration: Self-initiated versus other-initiated learning, *J. Pers. Soc. Psychol.* **35**, 459 (1977).
 - [13] K. D. Hjelmstad and A. Baisley, A novel approach to mastery-based assessment in sophomore-level mechanics courses, in *Proceedings of the 2020 ASEE Virtual Annual Conference Content Access* (2020), <https://peer.asee.org/a-novel-approach-to-mastery-based-assessment-in-sophomore-level-mechanics-courses>.
 - [14] J. Langdon, D. T. Botnaru, M. Wittenberg, A. J. Riggs, J. Mutchler, M. Syno, and M. C. Caciula, Examining the effects of different teaching strategies on metacognition and academic performance, *Adv. Physiol. Educ.* **43**, 414 (2019).
 - [15] A. Moeller and C. Reschke, A second look at grading and classroom performance: Report of a research study, *Mod. Lang. J.* **77**, 163 (1993).
 - [16] S. Page and L. R. Alexitch, Learning- and grade-orientation, sex, and prediction of self-reported academic performance, *Psychol. Rep.* **92**, 320 (2003).
 - [17] M. Meloth and P. Deering, Effects of two cooperative conditions on peer-group discussions, reading comprehension, and metacognition, *Contemp. Educ. Psychol.* **17**, 175 (1992).
 - [18] A. Hayek, C. Toma, D. Oberle, and F. Butera, Grading hampers cooperative information sharing in group problem solving, *Soc. Psychol.* **46**, 121 (2015).
 - [19] A.-S. Hayek, C. Toma, S. Guidotti, D. Oberlé, and F. Butera, Grades degrade group coordination: Deteriorated interactions and performance in a cooperative motor task, *Eur. J. Psychol. Educ.* **32**, 97 (2017).
 - [20] C. Turpen and N. D. Finkelstein, The construction of different classroom norms during peer instruction: Students perceive differences, *Phys. Rev. ST Phys. Educ. Res.* **6**, 020123 (2010).
 - [21] E. Yerushalmi, C. Henderson, K. Heller, P. Heller, and V. Kuo, Physics faculty beliefs and values about the teaching and learning of problem solving. I. Mapping the common core, *Phys. Rev. ST Phys. Educ. Res.* **3**, 020109 (2007).
 - [22] H. Pollio and H. Beck, When the tail wags the dog: Perceptions of learning and grade orientations in, and by, contemporary college students and faculty, *J. Higher Educ.* **71**, 84 (2000).
 - [23] S. Blum, *Ungrading: Why Rating Students Undermines Learning (and What to Do Instead)* (West Virginia University Press, Morgantown, WV, 2020).
 - [24] A. Elby, Another reason that physics students learn by rote, *Am. J. Phys.* **67**, S52 (1999).
 - [25] J. Tuminaro and E. F. Redish, Elements of a cognitive model of physics problem solving: Epistemic games, *Phys. Rev. ST Phys. Educ. Res.* **3**, 020101 (2007).
 - [26] M. Dancy, C. Henderson, N. Apkarian, E. Johnson, M. Stains, J. R. Raker, and A. Lau, Physics instructors'

- knowledge and use of active learning has increased over the last decade but most still lecture too much, *Phys. Rev. Phys. Educ. Res.* **20**, 010119 (2024).
- [27] S. Levy, A. Noga, and E. Yerushalmi, Long and winding road toward change in teachers' goals and orientations in the instructional lab, *Phys. Rev. Phys. Educ. Res.* **21**, 010145 (2025).
- [28] A. Malespina and C. Singh, Gender differences in test anxiety and self-efficacy: Why instructors should emphasize low-stakes formative assessments in physics courses, *Eur. J. Phys.* **43**, 035701 (2022).
- [29] C. Paul and D. Webb, Percent grade scale amplifies racial or ethnic inequities, *Phys. Rev. Phys. Educ. Res.* **18**, 020103 (2022).
- [30] D. J. Webb and W. H. Potter, Gender-grade-gap zeroed out under a specific intro-physics assessment regime, [arXiv:2102.10451](https://arxiv.org/abs/2102.10451).
- [31] E. Hackerson, T. Slominski, N. Johnson, J. Buncher, S. Ismael, L. Singelmann, A. Leontyev, A. Knopps, A. McDarby, J. Nguyen, D. Condry, J. Nyachwaya, K. Wissman, W. Falkner, K. Grieger, L. Montplaisir, A. Hodgson, and J. Momsen, Alternative grading practices in undergraduate STEM education: A scoping review, *Discip. Interdiscip. Sci. Educ. Res.* **6**, 15 (2024).
- [32] D. Muller, *Designing Effective Multimedia for Physics Education* (University of Sydney, Sydney 2008).
- [33] A. V. Knaub, J. M. Aiken, and L. Ding, Two-phase study examining perspectives and use of quantitative methods in physics education research, *Phys. Rev. Phys. Educ. Res.* **15**, 020102 (2019).
- [34] T. O'Brien, W. B. Lane, J. C. Speirs, and N. Laird, Student preferences for grade weighting in introductory physics, presented at PER Conf. 2025, [10.1119/perc.2025.pr.O-brien](https://doi.org/10.1119/perc.2025.pr.O-brien).
- [35] N. D. Finkelstein and S. J. Pollock, Replicating and understanding successful innovations: Implementing tutorials in introductory physics, *Phys. Rev. ST Phys. Educ. Res.* **1**, 010101 (2005).
- [36] L. C. McDermott, Oersted medal lecture 2001: Physics education research—The key to student learning, *Am. J. Phys.* **69**, 1127 (2001).
- [37] E. F. Redish, Oersted lecture 2013: How should we think about how our students think?, *Am. J. Phys.* **82**, 537 (2014).
- [38] Z. Hazari, G. Sonner, P. M. Sadler, and M.-C. Shanahan, Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study, *J. Res. Sci. Teach.* **47**, 978 (2010).
- [39] Z. Y. Kalender, E. Marshman, C. D. Schunn, T. J. Nokes-Malach, and C. Singh, Gendered patterns in the construction of physics identity from motivational factors, *Phys. Rev. Phys. Educ. Res.* **15**, 020119 (2019).
- [40] P. W. Irving, D. McPadden, and M. D. Caballero, Communities of practice as a curriculum design theory in an introductory physics class for engineers, *Phys. Rev. Phys. Educ. Res.* **16**, 020143 (2020).
- [41] N. M. Holincheck and T. M. Galanti, Applying a model of integrated STEM teacher identity to understand change in elementary teachers' STEM self-efficacy and career awareness, *School Sci. Math.* **123**, 234 (2023).
- [42] J. L. Rosenberg, N. Holincheck, K. Fernández, B. W. Dreyfus, F. Wardere, S. Stehle, and T. N. Butler, Role of mentorship, career conceptualization, and leadership in developing women's physics identity and belonging, *Phys. Rev. Phys. Educ. Res.* **20**, 010114 (2024).
- [43] M. Kryjevskaja, P. R. L. Heron, and A. F. Heckler, Intuitive or rational? Students and experts need to be both, *Phys. Today* **74**, No. 8, 28 (2021).
- [44] D. McPadden, E. Brewé, C. Monsalve, and V. Sawtelle, Productive faculty resources activated by curricular materials: An example of epistemological beliefs in University Modeling Instruction, *Phys. Rev. Phys. Educ. Res.* **16**, 020158 (2020).
- [45] D. Hammer and A. Elby, Tapping epistemological resources for learning physics, *J. Learn. Sci.* **12**, 53 (2003).
- [46] A. J. Cavanagh, O. R. Aragón, X. Chen, B. A. Couch, M. F. Durham, A. Bobrownicki, D. I. Hanauer, and M. J. Graham, Student buy-in to active learning in a college science course, *CBE Life Sci. Educ.* **15**, ar76 (2016).
- [47] J. L. Docktor, J. Dornfeld, E. Frodermann, K. Heller, L. Hsu, K. A. Jackson, A. Mason, Q. X. Ryan, and J. Yang, Assessing student written problem solutions: A problem-solving rubric with application to introductory physics, *Phys. Rev. Phys. Educ. Res.* **12**, 010130 (2016).
- [48] B. Gutmann and T. Stelzer, Values affirmation replication at the University of Illinois, *Phys. Rev. Phys. Educ. Res.* **17**, 020121 (2021).
- [49] A. Miyake, L. E. Kost-Smith, N. D. Finkelstein, S. J. Pollock, G. L. Cohen, and T. A. Ito, Reducing the gender achievement gap in college science: A classroom study of values affirmation, *Science* **330**, 1234 (2010).
- [50] R. Pekrun, The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice, *Educ. Psychol. Rev.* **18**, 315 (2006).
- [51] T. Lê, V. Retzer, G. Deerwester, J. C. Speirs, and S. Erhlich, Student experiences with labor-based grading in physics, presented at PER Conf. 2025, Washington, DC, <https://www.per-central.org/perc/2025/detail.cfm?ID=19503>.
- [52] B. Högberg, J. Lindgren, K. Johansson, M. Strandh, and S. Petersen, Consequences of school grading systems on adolescent health: Evidence from a Swedish school reform, *J. Educ. Pol.* **36**, 84 (2021).
- [53] A. Linder, M. Nordin, U.-G. Gerdtham, and G. Heckley, Grading bias and young adult mental health, *Health Econ.* **32**, 675 (2023).
- [54] A. Kjærgaard, J. Buhl-Wiggers, and E. N. Mikkelsen, Does gradeless learning affect students' academic performance? A study of effects over time, *Stud. Higher Educ.* **49**, 336 (2024).
- [55] J. Radoff, L. Z. Jaber, and D. Hammer, "It's scary but it's also exciting": Evidence of meta-affective learning in science, *Cognit. Instr.* **37**, 73 (2019).
- [56] J. Mezirow, An overview on transformative learning, in *Lifelong Learning* (Routledge, Oxfordshire, England, 2008), pp. 40–54.