Exploratory analysis on the impact of alternative grading in introductory physics courses

Richard L. Pearson III and Caitlin Martin
Department of Physical Sciences & Department of Humanities and Communication, respectively, from Embry-Riddle Aeronautical University, Daytona Beach, FL

Alternative grading systems are becoming more prevalent in general academia, to more accurately, effectively, and equitably represent student learning. These systems include labor-based, specifications, mastery, and ungrading. Other disciplines acknowledge variability in individual student preparation, time commitment, effort, and learning and have used these alternative systems to provide a link between individualized learning and the cemented letter grading system through which students are currently evaluated. Nevertheless, these alternative grading systems have yet to find widespread audience in the physics community, even though the inherent student variability is similarly prominent. These grading systems aim to enhance individualized learning by encouraging remediation of mistakes and reducing grade anxiety. A blend of these grading systems was implemented in multiple introductory physics courses. Concept inventory gains historically highlighted the positive impact of pedagogical alterations; those comparative gains are not present here. Regardless, an argument is made in support of grading system transformation due to instructor involvement, student perceptions, and end-of-term evaluations.

2025 PERC Proceedings edited by Pawl, Zwolak, and Leak; Peer-reviewed, doi.org/10.1119/perc.2025.pr.Pearson_III

I. INTRODUCTION

Until an overhaul is made on the cultural and financial incentive structure of traditional letter grades, instructors and teachers must continue to adapt to this seemingly confined space [1]. As such, a variety of alternative grading systems have been introduced across academia in an attempt to counteract grade inflation, grade deflation, grade misrepresentation, grades that undermine student motivation, and other issues connected to traditional letter grading systems [2]. These alternative systems do not profess to be the sweeping solution in the higher-education classroom. Instead, these systems acknowledge the adverse—and often conflicting—responsibilities of instructors as both the facilitator of learning and the arbiter of grades. They seek to remove the adjudicator role.

The transition to these alternative systems seems widely necessary in light of the decreasing efficacy of graduating high school seniors over the last few years [3], though care must be taken to review impacts of the grades themselves on student success [4]. Additionally, first and second-year university students need more holistic support for how to learn as an individual, what it means to intellectually struggle, and how to overcome adversity and failure (see [5, 6]; and [7, 8] on *threshold* concepts). These are principles and dispositions educators hope to instill in their students regardless of their preparation or previous learning experiences. As such, alternative grading systems attempt to address and imbue these ideals into students regardless of course subject.

Some have made the transition, published their efforts, and should be celebrated for doing so (as with [9–14] to identify a few examples). Yet, there is limited research, tools, and theoretical foundations supporting alternative grading methods and their benefits [15]. Still, scholars interested in alternative grading have surfaced the arbitrary nature of "traditional" grades that many have come to accept uncritically as the status quo [16]. This project uses several pilot attempts at alternative grading in introductory, calculus-based physics courses to try to identify assumptions about grading practices and reveal areas for future research about how learning in physics is evaluated. The presented system evolved out of an interdisciplinary collaboration and exploration between physics and rhetoric/writing instructors.

II. GRADING SYSTEMS

Alternative grading is a recent umbrella term for a variety of practices that seek to eschew traditional letter grades [16, 17]. Some of these approaches grew out of concerns about the arbitrary nature of grades and a recognition that grading can stigmatize learners [18], while others sought to restore academic rigor by ensuring that final course letter grades represent achievement of course learning outcomes [19]. A full exploration of these approaches is beyond the scope of this article, especially given that many instructors have used similar names to describe different grading practices (read [16] for full details).

The following four approaches were discussed by the authors as they piloted alternative grading in physics, humanities, and communication classes. The order described here represents a continuum of the extent to which the alternative grading system differs from traditional grading.

Ungrading: Ungrading is perhaps the most extreme alternative grading approach, as it is intended as "a systematic critique" that includes students in conversations about grading [20]. This approach decouples an instructors' qualitative feedback and puts the onus of evaluation on students, typically through reflection and self-assessment practices. Ungrading recognizes the non-linear nature of learning and centers trust in the student-faculty relationship [20, 21]. Students continue to receive expert feedback, but final grades are assigned based on students' own evaluations of their learning.

Labor-based grading: This system distributes final course grades based on the effort, or labor, that one puts into the course [18]. Students and instructor openly discuss the expectations for earning a specific letter grade and students commit to that grade by signing a grading contract or agreement. Instructors provide feedback, but do not assign letter grades or other typical indicators of quantitative evaluation. This system presumes that the assignments and labor expectations provide students with appropriate learning opportunities. Grading contracts have been a common alternative grading system in composition, rhetoric, and writing studies since the 1970s [22] that have seen a recent resurgence (for additional resources, examples, and outlines, see asaobinoue.blogspot.com).

Specification grading: Recursive attempts on pass/fail assessments of focused course learning outcomes centers the achievement in this grading system [19]. Indeed, one of the main goals of "specs" grading is to remove the concept of partial credit, thereby directing student energy and focus to mastery of the material. This shift also introduces and reinforces acceptance that mistakes and failure are requisite in mastering new ideas. Assessments are "bundled" to specific course learning outcomes, such that a final grade correlates to the number of outcomes successfully achieved [23]. The decision about what demonstrates mastery remains with the course instructor rather than being negotiated by students as in the previous two systems.

Mastery and standards-based grading: Here, students must master specific learning objectives, but unlike specs grading, the mastery of these objectives follow linear (or vertical) progressions [24]. Mastery and standards-based grading systems focus on individualized learning with demonstrations of proficiency in identified areas. For instance, before working out derivations of foundational physics equations, students need to recall and replicate the equations themselves (perhaps following along a Bloom taxonomy type of path [25]). In these schemes, students must master some learning outcomes before they are able to move onto the next mate-

TABLE I. Adopted alternative grade expectations for introductory physics courses.

To demonstrate a(n):	**A**	**B**	**C**	**D**
No. of examination questions scored at a "4"	15 of 20	11 of 20	7 of 20	3 of 20
Final exam percentage	$80-100\%$	$70-79\%$	$50-69\%$	$40-49\%$
Completed advanced problem sets	10 of 12	8 of 12	5 of 12	N/A
Completed learning activity sets	3 of 3	2 of 3	1 of 3	N/A

rial or outcome in the sequence. The instructor remains the arbiter of proficiency.

It is crucial to recognize that all of these alternative grading systems seek to decentralize letter grades on exams or course projects and replace it with an emphasis on the process of learning, though some systems place the ultimate evaluation more in students' hands, while others maintain the instructor's authority to evaluate student learning. Specs, mastery, and standard-based grading, in particular, emphasize recursive, practice-oriented learning processes that recognize the crucial role of expert feedback in supporting students as they develop expertise [26].

III. ADOPTED GRADING SYSTEM

A practice-oriented focus is built into the current structure of introductory physics courses in higher education via the litany of available practice problems and worked examples. Therefore, it seems suitable to fit these courses into the previously described grading models. Currently, however, evaluation of student learning in physics courses relies heavily upon examinations—and in some cases, term projects as well. These inherently require evaluation from the instructor, though there are alternatives [27–29].

The traditional course letter grades for introductory, calculus-based physics courses at a private, aeronautical-focused institution consists of the distribution of three exams for 60% of the grade (including the final exam), weekly quizzes for 30%, and 10% for all other work in the course (e.g. homework, preparatory assignments, etc.). Letter grades are heavily dependent on student performance in high-stakes, instructor-graded assessments, providing very limited feedback and space for student revisions, corrections, adjustment, and, perhaps, learning. Nevertheless, physics students must be able to demonstrate conceptual understanding and problem-solving skills in a measurable way and in a timely fashion without the aid of outside resources or artificial intelligence chats. One-on-one, in-person oral assessments would fulfill this need but are not particularly suitable for large class sizes [30]; paper-based exams are, thus, the default and the *expected* assessment. The dichotomy of the introductory physics classroom makes the transition to alternative grading systems precarious to navigate, as its need to ensure student understanding (or learning) seems to collide with its cultural expectation of utilizing high-stakes examinations to assess it.

To find an amenable balance between these current inherent conflicts, a grading system was constructed and adopted utilizing *specifications grading* as its base, and incorporating aspects of *labor-* and *mastery-based* systems along with components of *ungrading*. Table I displays how the coursework divides into four categories and disperses as a column of specifications for each letter grade: exam questions (each question corresponds to a specific learning target), the final exam (a traditional, comprehensive assessment with partial credit on worked problems), advanced problem sets (coursework with complex, higher-order problem-solving), and learning activity sets (self-reflection surveys, tutorials, in-class peer work, quizzes, etc. aggregated with material covered on each exam). All but the comprehensive final exam allowed for repetitive learning via an iterative feedback loop. The advanced problems and learning activity sets were counted as "complete" when students displayed proficiency of the material; proficiency standards were established and evaluated by the instructor. Letter grades differentiated by level of completion and amount of labor. The grading structure shown in Table I is the most recent iteration used in Course 3, as identified in Table II; the previous iterations were improved upon for the latest adaptation.

The learning activities and advanced problems counted as "complete" when submitted work demonstrated proficiency (defined as having little-to-no errors with fully detailed explanations of the solution). If explanations were insufficient, it was returned with feedback; students were expected to discuss with others, make adjustments, provide corrections and resubmit. Resubmitted work was again analyzed and either marked as proficient/complete or returned with additional feedback to repeat the process. Overall, this encouraged engagement with the instructor, fellow peers, and the material to ultimately detach the letter grade's worth from the correctness of the activity.

Instead of increasing the number of high-stake examinations to offset the large weight on grades (as is typically done), the alternative grading system created here considers the cumulative aspect of individual exam questions. Each question is assessed on an integer scale of 0 to 4: "0" indicates no applicable work shown; "1" means a little work was given (and/or approximately less than 25% correct); a "3" signifies errors were present but the answer was nearly complete; a "4" shows proficiency with little-to-no error; and a "2" applies to everything between a "1" and a "3". Exam questions not reaching a "4" were permitted an additional attempt on complementary problems, after students completed remedial interventions (e.g., instructive tutorials, peer-work, etc.). In

TABLE II. Quantitative data from three alternative grading courses as compared to a traditional course

	Course 1	Course 2	Course 3	Traditional
Concept inventory gains and student perceptions of learning				
MWCS2 learning gain [95% CI]	0.14 [0.03,0.26]	0.21 [0.15,0.26]	0.20 [0.11,0.30]	0.16 [0.06,0.26]
CLASS favorable % change (SE)	+5 (8)%	−3 (4)%	−1 (4)%	−4 (7)%
End-of-course (EOC) student survey results as compared to "Traditional" scale of 1 to 4 (with $p < 0.001$ unless otherwise noted)				
Instructor presentation of material	+0.73	+0.31	+0.23	—
Structure of course	+0.96	+0.46	+0.31 (< 0.005)	—
Learning outcomes achieved	+0.70	+0.30	+0.18 (< 0.05)	—
Learning environment	+0.66	+0.34	+0.28	—
Term	2023 Summer	2023 Fall	2024 Fall	2023 Fall
No. of students in course	11	76	73	48
No. of responses for MWCS2, CLASS, and EOC survey	6, 10, 10	29, 36, 57	42, 38, 54	12, 13, 38
Average participation rates	79%	53%	61%	44%
Type of physics course	†	‡	‡	†

† *course material includes fluid mechanics, thermodynamics, waves, and optics for undergraduate engineering majors.*

‡ *same course material as the other but for undergraduate physics, astronomy, and space-physics majors.*

other words, even the high-stake assessments were adjusted so as to encourage practice-focused learning processes and further disentangle the convoluted course letter grade.

As with *specifications grading*, students earn the letter grade for which they have reached the lowest spec. For instance, using the scheme in Table I, a student earns a "C" grade for scoring a "4" on twelve exam questions, scoring a 65% on the final exam, completing eleven advanced problems, and finishing three learning activity sets. Students unable to reach "D" level specs earn a failing grade. Most students in the alternative grading courses were self-motivated enough and aware of their progress such that the final exam became the determinant of their course letter grade.

Upper administrators were concerned that an alternative grading implementation would lower the letter grade distribution. To examine it directly (and to ease administration fears), a traditional exam-quiz-homework weighting was fit onto assessment scores from the alternative courses to calculate letter grades. The alternative grading system positively shifted the distribution. Since the alternative system requires more focused effort in every coursework category, the letter grades are meant to illustrate a more wholistic representation of student learning and effort. The results are compared for one of the courses in Fig. 1.

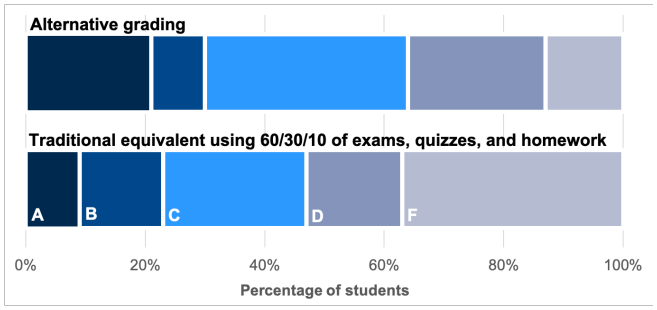

FIG. 1. The distribution of letter grades from Course 3 when calculated using both the traditional and alternative grading approaches. The traditional grading assumed a weight distribution of exams and other coursework given as a 60-30-10 split.

IV. DATA & DISCUSSION

The alternative grading system presented in Section III was implemented across three introductory, calculus-based physics courses at a private, aeronautical-focused Southeastern United States institution from 2023 to 2024, as detailed in Table II. A fourth course was a part of a larger conglomerate of sections covering the same material and using the same traditional grading scheme; this is the comparative course identified as "traditional". All four courses covered introductory material from fluid mechanics, periodic motion, thermodynamics, and optics. Course 1 took place across a summer term with a low student population but was an important testbed for the initial implementation. Courses 2 and 3 were larger in numbers, but varied with the other two in that the former are presented for physics majors, rather than engineers; at this university, some of the in-classroom approaches, comprehension expectations, and learning objectives vary slightly.

Course activities such as preparatory assignments, homework, and quizzes were consistent between courses. In-person class experiences were also similar, with lectures utilizing think-pair-share questions, peer-learning activities, and group problem solving. The only activities that differed between the alternative and traditionally graded courses was the exam question feedback process.

Quantitative results. Both the concept inventory (the Mechanical Waves Conceptual Survey 2, MWCS2) and the perception survey (the Colorado Learning Attitude about Science Survey, CLASS) were administered in a

pre/post format during the first and last weeks of the course [31, 32]. The student learning gain (calculated as $[< \text{post} > - < \text{pre} >] / [100 - < \text{pre} >]$) and the 95% confidence interval are listed for the MWCS2 in Table II. All courses had low (< 0.2) but significant learning gains; overlapping CIs, however, suggests no measurable variance within the classroom.

The change of favorable ratings in the CLASS, shown in Table II, are not statistically different either, though all alternative grading courses trended less negative than the traditional. It should be noted that a depreciation of favorable responses is not uncommon (even the "typical" data in the validating CLASS paper displays this [32]).

The end-of-course (EOC) survey collected by the University had high participation rates. Survey results (based on a scale of 1 to 4) are divided into four categories: instructor presentation of material, structure of the course, learning outcomes achieves, and learning environment. All courses averaged scores above a "3" on this scale. Table II shows the differences between the average responses of the alternative grading system and the traditional course, with each category displaying a positive, statistically significant change (with all differences with $p < 0.05$).

Wholistically, these are not particularly astonishing results. But they do paint a promising portrait that is subtly enticing and inviting. For instance, if one presumes that more learning is expected within the alternative grading environment (for all the reasons previously discussed), then these pilot courses insinuate limits to the concept inventory's capacity to measure learning differences between grading systems. If so, perhaps usage of new or other instruments are necessary.

Qualitative feedback. Student feedback was collected throughout the entirety of the course. Midterm adjustments to the grading framework were made in conjunction with the on-campus faculty teaching and learning center directors. As stated earlier, the specs published in Table I are the end-result of collaborative feedback and discussion between the authors and the students. Students also provided comments at the course end, prior to the administration of their final exams.

A small, representative sampling of student comments pertaining to the structure and operational aspects of the alternative grading system are given below:

"*What has helped [the most] is . . . the level of redemption instead of punishment in the class . . . I may get frustrated with the material and fumble, but I'm allowed to try again.*"

"*[The instructor] used assignments that we were able to submit and receive feedback on and then try again. This cyclical approach is one of the best ways I have ever learned.*"

"*The grading system didn't encourage mastery in my opinion. It encouraged repeating and obsessing over your score as many of my classmates did which distracted them from the actual material they are learning.*"

"*The grading system was difficult to get used to, but it was actually helpful in the end.*"

The first two student comments explicitly describe the individual success of one of the main goals for transitioning to the alternative grading system: remove the incompatible stress between grading and learning through correcting mistakes. Though possibly hyperbole, it is uplifting to read a student express a breakthrough in their learning journey as being "one of the best ways [to] have ever learned." Many expressed positive views of the feedback loop as relief from not having to understand it perfectly the first time.

On the contrary, some students expressed their frustration with the new system as they felt the feedback loop focused solely on the grade rather than the learning. This displays a disconnect between the introduction of the system, the student's personal application of the prescribed methodology, and the need for instructors to schedule individual follow-up interviews to ensure sufficient guidance. The final student comment touches on that: it took time for the grading system dynamic to become a part of their learning, but then recognized it was indeed a worthwhile pursuit. Some of that transition was evident from the midterm discussions and visible adjustments made together as a whole class.

Future refinements. Student focus groups and follow-up surveys are planned for deployment in longitudinal studies of alternative grading systems impact on student academic performance in physics. Additional adjustments to the adopted grading system are also planned to address the concerns from students and to ensure a focused effort on learning and mastery rather than the grade itself.

V. CONCLUSIONS

The results of a validated concept inventory and perception survey in a handful of introductory physics courses add data to a growing pool of research on alternative grading systems [15]. They are encouraging, yet underwhelming as evidence that a transition to alternative grading is beneficial to student learning. The end-of-course surveys—for which upper administration at many higher education institutions place great stock—do show statistically significant improvements of student perceptions towards the instructor, the course structure, and overall learning experience when transitioned to an alternative grading model. This is a significant part of justifying the growth and implementation of systematic change across additional courses, departments, and institutions. It is imperative that modern-day students learn to learn without the ominous weight of grades.

ACKNOWLEDGMENTS

The authors would like to thank the continual expertise of the Directors of the Center for Teaching and Learning on the Daytona Beach campus; namely Drs Teha Cooks, Claudia A. Cornejo Happel, and Chad M. Rohrbacher, without whom these ideas would still be laying lifeless. The authors would also like to thank administrators for permitting such excursions, as well as the reviewers for their insightful feedback.

[1] A. Kohn, The case against grades, Counterpoints **451**, 143 (2013).

[2] L. B. Nilson, D. Clark, and R. Talbert, Grading for growth 10.4324/9781003445043 (2023).

[3] V. Irwin, K. Wang, J. Jung, E. Kessler, T. Tezil, S. Alhassani, A. Filbey, R. Dilig, and F. Bullock Mann, Report on the condition of education 2024 (nces 2024-144) (2024).

[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] R. Maymon, N. C. Hall, and J. M. Harley, Supporting first-year students during the transition to higher education: The importance of quality and source of received support for student well-being, Student Success **10**, 64 (2019).

[6] C. Johnson, R. Gitay, A.-S. G. Abdel-Salam, A. BenSaid, R. Ismail, R. A. Naji Al-Tameemi, M. H. Romanowski, B. M. Kazem Al Fakih, and K. Al Hazaa, Student support in higher education: campus service utilization, impact, and challenges, Heliyon **8**, e12559 (2022).

[7] E. Wardle, Creative repurposing for expansive learning: Considering "problem-exploring"' and "answer-getting"' dispositions in individuals and fields, Composition Forum (2012).

[8] P. Felten, On the threshold with students, in *Educational Futures* (SensePublishers, 2016) pp. 3–9.

[9] V. L. Dickinson and L. B. Flick, Beating the system: Course structure and student strategies in a traditional introductory undergraduate physics course for nonmajors, School Science and Mathematics **98**, 238 (1998), https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1949-8594.1998.tb17297.x.

[10] A. Rundquist, Standards-based grading with voice: Listening for students' understanding, in *Physics Education Research Conference 2011*, PER Conference Invited Paper, Vol. 1413 (Omaha, Nebraska, 2011) pp. 69–72.

[11] I. D. Beatty, Standards-based grading in introductory university physics, Journal of the Scholarship of Teaching and Learning , 1 (2013).

[12] T. Zimmerman, Grading for understanding - standards-based grading, Phys. Teach. **55**, 47 (2017).

[13] M. Richard, J. Delgado, S. LeGresley, and C. J. Fischer, arXiv:2206.10574 [physics.ed-ph] (2022).

[14] J. M. Wachter, Labor-based grading practices in the physics classroom, American Journal of Physics **92**, 336 (2024), https://pubs.aip.org/aapt/ajp/article-pdf/92/5/336/19897270/336_1_5.0121625.pdf.

[15] E. L. Hackerson, T. Slominski, N. Johnson, J. B. Buncher, S. Ismael, L. Singelmann, A. Leontyev, A. G. Knopps, A. McDarby, J. J. Nguyen, D. L. J. Condry, J. M. Nyachwaya, K. T. Wissman, W. Falkner, K. Grieger, L. Montplaisir, A. Hodgson, and J. L. Momsen, Alternative grading practices in undergraduate stem education: a scoping review, Discip Interdscip Sci Educ Res **6**, https://doi.org/10.1186/s43031-024-00106-8.

[16] D. Clark and R. Talbert, *Grading for growth: A guide to alternative grading practices that promote authentic learning and student engagement in higher education* (Routledge, 2023).

[17] C. for Grading Reform, Resources (2025), retrieved 2025 May 14 from www.CenterForGradingReform.org.

[18] A. B. Inoue, *Labor-Based Grading Contracts: Building Equity and Inclusion in the Compassionate Writing Classroom, 2nd ed.* (The WAC Clearinghouse, University Press of Colorado, 2022).

[19] L. B. Nilson, *Specifications Grading: Restoring Rigor, Motivating Students, and Saving Faculty Time* (Routledge, 2014).

[20] J. Stommel, *Undoing the grade: Why we grade, and how we stop* (Hybrid Pedagogy Inc., Denver, CO, 2023).

[21] S. Blum, *Ungrading: Why rating students undermines learning (and what to do instead)* (West Virginia University Press, Morganstown, 2020).

[22] J. Danielewics and P. Elbow, A unilaterial grading contract to improve learning and teaching, College Composition and Communication **6**, 244 (2009).

[23] C. Martin, Implementing specs grading to encourage student success (2025), forthcoming 2025.

[24] R. Campbell, D. Clark, and J. OShaughnessy, Introduction to the special issue on implementing mastery grading in the undergraduate mathematics classroom, PRIMUS **30**, 837 (2020), https://doi.org/10.1080/10511970.2020.1778824.

[25] B. S. Bloom, *Taxonomy of Educational Objectives, Handbook: The Cognitive Domain* (David McKay, New York, 1956).

[26] M. C. Lovett, M. W. Bridges, M. DiPietro, S. A. Ambrose, and M. K. Norman., *How learning works: Eight research-based principles for smart teaching, 2nd edition* (Jossey-Bass, 2023).

[27] S. Freeman and J. W. Parks, How accurate is peer grading?, CBE-Life Sciences Education **9**, 482 (2010).

[28] J. Cain, M. Medina, F. Romanelli, and A. Persky, Deficiencies of traditional grading systems and recommendations for the future, Am. J. Pharm. Ed. **86**, 8850 (2022).

[29] L. Zhang, Y. Zong, T. Chen, Y. Ma, S. An, and L. Zhang, A peer review approach to grading projects in computer courses, in *2023 IEEE Frontiers in Education Conference (FIE)* (2023) pp. 1–7.

[30] J. Monks and R. M. Schmidt, The impact of class size on outcomes in higher education, The B.E. Journal of Economic Analysis Policy **11**, doi:10.2202/1935-1682.2803 (2011).

[31] P. Barniol and G. Zavala, Mechanical waves conceptual survey: Its modification and conversion to a standard multiple-choice test, Phys. Rev. Phys. Educ. Res. **12**, 010107 (2016).

[32] W. K. Adams, K. K. Perkins, N. S. Podolefsky, M. Dubson, N. D. Finkelstein, and C. E. Wieman, New instrument for measuring student beliefs about physics and learning physics: The colorado learning attitudes about science survey, Phys. Rev. ST Phys. Educ. Res. **2**, 010101 (2006).