Pondering zeros: Uncovering hidden inequities within a decade of grades

Cassandra Paul,[1,2] David J. Webb,[3] Mary K. Chessey,[3] and James Lucas[2]

[1]*Department of Physics & Astronomy, San Jose State University, One Washington Square, San Jose, CA, 95192*
[2]*Science Education Program, San Jose State University, One Washington Square, San Jose, CA, 95192*
[3]*Department of Physics, University of California, Davis, One Shields Avenue, Davis, CA, 95616*

When assessing student work, graders will often find that some students will leave one or more problems blank on assessments. Since there is no work shown, the grader has no means to evaluate the student's understanding of a particular problem, and thus awards zero points. This practice punishes the student behavior of leaving a problem blank, but this zero is not necessarily an accurate assessment of student understanding of a particular topic. While some might argue that this grading practice is "fair" in that students know that they can't receive points for answers they don't submit, we share evidence that different student groups engage in blank-leaving behavior at different rates and are therefore unequally impacted. We analyze 10 years of UC Davis introductory physics course databases to show that different groups of students skip problems and entire exams at different rates. We also share some implications for grading practices.

I. INTRODUCTION

Grades are used by students, teachers, schools, and much of the rest of society as a measurement of a student's understanding of, or skills in, a subject. Given this expectation, it's useful to understand the connection between a grade given to a student and the actual understanding or skill of that student. This connection is difficult to make when a student fails to provide information about their understanding by skipping a question on a quiz or exam, or by missing a quiz entirely. Students may leave answers blank for a variety of reasons, which may or may not be related to the student's overall understanding of a topic. When a student leaves a problem blank it may be because they have run out of time, they don't want to reveal their inadequacies, they don't understand the question, they don't see value in attempting a problem they don't think they can finish, or for any other number of reasons.

One way instructors commonly deal with this problem is by awarding the student a zero, reasoning that if the student doesn't provide any evidence of understanding, no credit can be awarded. While on the surface this instructor practice seems logical and fair ("Equity of Fairness [1]"), it is still a consequence of the student's behavior, and not a measurement of their knowledge of a given topic. Thus a zero, the numerical equivalent of no understanding, is awarded for what instead might be better considered missing data. An instructor might be content with awarding a grade that represents both classroom behavior and understanding (in fact the authors of this paper have done this as instructors), however, if different student populations engage in blank-leaving behavior at different rates or for different reasons, and this practice negatively impacts some student groups more than others on average, then assigning zeros to blanks could contribute to inequity in the course (Rodriguez et al. "Equity of Parity" model [1]).

Various studies [2, 3] over the years have investigated the extent to which a mismatch between the cultural norms of a group of students and the cultural norms of their educational institution may affect learning and the resulting grades of that group of students. For example, Stephens et al. [2] suggest that first-generation college students are perhaps oriented more toward values necessary for a community and less toward the individual independence that is expected by most institutions of higher education. Taking an exam is perhaps one of the most independent activities a student experiences in a course, especially active learning courses, so this particular cultural difference may be exaggerated when it counts the most, during assessment.

Giving a grade of zero for a blank answer [4] can have a very negative impact on a student's overall exam score, especially if their instructor is using a traditional form of the percent scale [5] where earning somewhere around 50 or 60% of the points is considered failing. In this paper we do not draw any conclusions on why students leave questions unanswered. Instead, we evaluate the extent to which different student populations engage in this practice.

For this exploratory study, we consider a few broad groupings of students by gender, race and ethnicity, and first-generation college status. Specifically we ask:

1. How do the behaviors of (a) leaving problems blank and (b) missing an exam vary by student population?

2. Do these behaviors correlate with other measures for understanding?

We examine exam grades in two introductory college physics courses (CLASP) over ten years and show that different populations of students engage in these behaviors at different rates, and provide evidence that this is not necessarily due to lack of student understanding. Finally, we provide suggestions for instructors concerned with increasing grading equity.

II. METHODS

Data and data sources: In a prior study of differences between grade scales [6], we looked up 96 original class databases from 2003-2012 to compile a set of 794,088 grades given on individual parts of 606 quizzes and 76 final exams, to 15,207 students taking either of the first two courses in

the CLASP series [7]. To these grade data we matched student's self-reported demographic data that we received from UC Davis administrative sources. We use this same data set in the current paper. All averages and standard errors were calculated in Excel. We used STATA software for statistical tests. This course utilizes the "Grading by Response Category" [8] method, where student solutions are placed in categories that are later assigned a numeric grade, so the 96 class-level databases include the individual grades that were given for each problem, the total exam scores for each student, the calculations that led to overall exam scores, and (in many cases) the calculations that led to a final grade. We examined copies of the original rubrics for 57 separate exams from these years, corresponding to about 200 exam problems, and found that 95% of the problems for which a score of zero was possible assigned zero only to "blank" or "essentially blank" solutions, meaning that there was no work shown evidencing that the student attempted the problem.

We limit our results to the 98% of the overall student population who were either US citizens or had permanent resident status. Our administration considers first-generation students to be those for whom neither parent had earned a bachelor's degree or higher. To identify race and ethnicity groups that are underrepresented in STEM, we use the categorization from our previous work [6, 9]. Students who self-reported belonging to any of the groups African American, Native American, Latina/o American, Mexican American, Chicana/o, and Pacific Island American are included in this group hereafter referred to as URM. Students self-reported binary gender from the options of male or female. Thirty-eight students (0.2%) did not identify as male or as female and are included in this study, but this group is not analyzed independently due to the lack of information about individuals' gender or gender identity.

The CLASP curriculum at UC Davis is spread over three quarters (equivalent to two semesters). We consider only the first two quarters of this 3-quarter introductory physics series because many students were only required to take the first two, and therefore the third quarter represents a different sample of students. We examine data from these two courses offered from 2003 through 2012. Throughout these ten years, students enrolled in a course spent about 4.7 hours per week in an active-learning discussion/lab section and about 1.3 hours per week in lecture. The topics covered and the student activities did not change much over these years so the core of this course series did not change. These courses were usually assigned two instructors for each 300 (or so) students and had a total of 60 different instructors and of order 200-250 Teaching Assistants over these ten years.

III. ANALYSIS AND RESULTS

The number of blank solutions on quiz and final exam items is low overall [10]. Students left blank only about 3% of problems on exams that they took. However, when we break the fraction down by student groups we find significant differences. The average fraction of blank answers left by students from racial and ethnic groups who are underrepresented in STEM (URM) is over 1.4 times the average for non-URM students. The average fraction of blank answers left by first-generation college students is about 1.3 times the average of non-first-generation-college-students. Furthermore, there is a small but significant ($t = 3.06, df = 17,530, p = 0.002$) difference in blank-leaving behavior based on gender; students who identified as female left more problems blank than students who identified as male. These results are shown in Fig. 1.

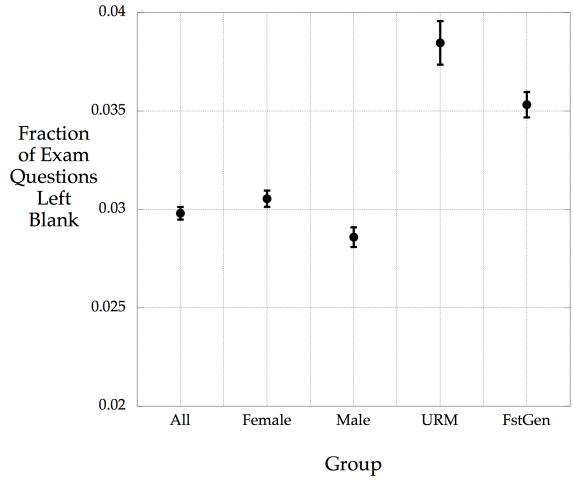

FIG. 1. The fraction of sub-exam-level answers (on quizzes and finals) that are blank and were given zeros (error bars are $\pm$ standard errors). These fractions are plotted for several large groups of students. URM refers to racial/ethnic groups who are underrepresented in STEM, and FstGen refers to students with neither parent a college graduate. $N_{total} = 17,570$, $N_{Female} = 10,856$, $N_{Male} = 6,676$, $N_{URM} = 2,102$, and $N_{FstGen} = 5,120$.

The second way that teachers encounter missing data when evaluating students' understanding is when students skip a quiz completely. The fraction of missed quizzes is also low, at about 4% overall. However, we again see clear differences between different student groups. The average fraction of missed quizzes for students from URM groups is again about 1.3 times the fraction for non-URM students. However, missed-quiz behavior seems to be very different than blank-leaving behavior across the other student groups. For example, while male students left fewer blank solutions than female students left, male students skipped, on average, about 1.3 times as many quizzes as female students skipped. Also, while first generation college students average more blank solutions than the overall average, their average number of missed quizzes is the same as the overall average. These results are displayed in Fig. 2.

Fig's 1 and 2 clearly show that different student groups engaged in these two behaviors at different rates, but one still might value assigning zeros to blank answers or skipped ex-

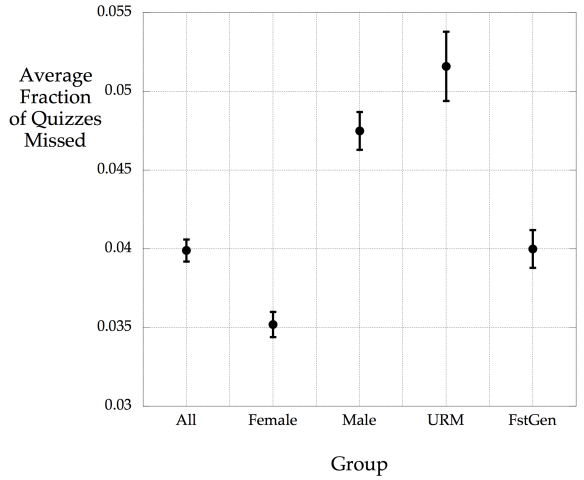

FIG. 2. The fraction of quizzes missed averaged over each of several large groups of students (error bars are ± standard errors). Students took between 4 and 9 quizzes during a course.

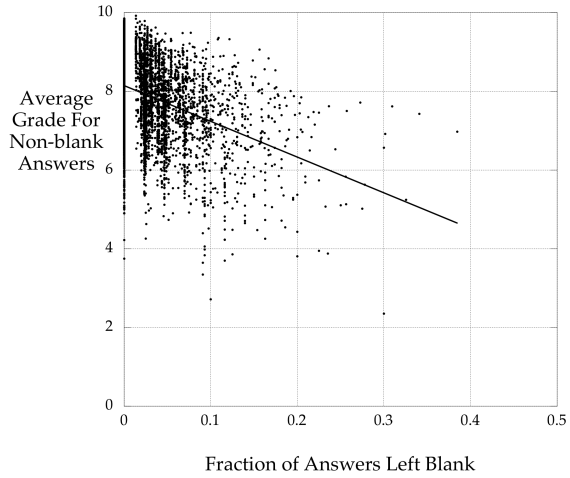

FIG. 3. For courses graded using a 10-point scale we plot students' unweighted average grade for all answers that were not blank against the fraction of answers left blank. The line is a linear fit ($R^2 = 0.13$) showing the trend.

ams as a reasonable assessment of academic strength. In other words, students may leave answers blank because they could not have written anything worth a higher score, or skip exams because they would have received an F anyway. While it's impossible to make a definite statement about this issue, we can use the dataset to look for correlations between blank-leaving behavior and other student work. When we plot average grade for non-blank-solutions against the overall fraction of answers left blank, we see that there is a slight negative correlation. The relationship means that students who leave more blanks are less likely to earn high scores on submitted work. However, an $R^2 = 0.13$ suggests that the variable "Fraction of Answers Left Blank" is a poor predictor of scores on submitted answers. Even the group of students leaving 10 to 15% of their answers blank includes a broad cross-section (SD = 0.8) of grades on the rest of their work, covering all letter grades A through F. See Fig. 3.

While the correlation is weak, Fig. 3 still shows that on average, there is at least some relationship between grades on submitted work and blank-leaving behaviors. Therefore, we wanted to check if blank-leaving was primarily a habit of low-scoring students. For our final analysis of these data, we only consider students who have earned an average of B (8.5 on a 10-point scale or 3.0 on a 4-point scale) or better on all the problems they attempted to solve (see Fig. 4). While there is less blank-leaving behavior overall among this subset of students, we still find similar differences between student groups that we saw in Fig. 1, which included all students.

While a complete quantitative analysis of the impact of these zeros on student course grades is beyond the scope of this paper, with the data shown you can begin to interpret consequences of behavioral zeros. For example, in Fig. 3, we see that there are a large number of students who have earned above 85% (a "B") on their submitted work, but have left between 5 and 15% of their solutions blank. If we assume that

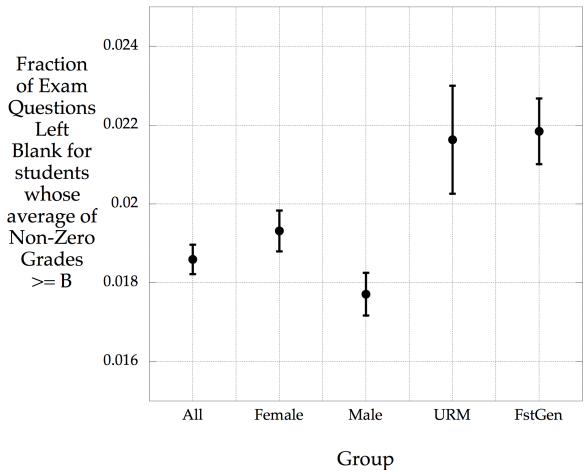

FIG. 4. For students whose non-zero answers average to a B or better, the fraction of sub-exam-level answers that were given zeros are shown as in Fig. 1.

each quiz item is weighted the same, we find that these students likely earned a course grade between 72% ("C-") and 81% ("B-"). Similarly, students who earned a 75% on submitted work (a "C") but left 10% of their answers blank, likely earned about 68% (a "D+").

IV. DISCUSSION

For any particular student, leaving a blank or skipping an exam may not change their grade. In our statement that leaving blanks or skipping a quiz will lead to a lower grade we are simply asserting that neither of these actions will ever raise

a grade; these actions can only lower the grade distribution for the relevant group. We find that some groups of students are more likely to leave blanks than others, and these differences seem to be at least somewhat independent of overall course knowledge as evidenced by the presence of group differences even among otherwise high-scoring students. Future work will use the "Equity of Parity" [1] model to examine the extent to which these two behaviors relate to differences in students' overall course grades, whereby incoming differences in students' approaches to the class are either addressed and accounted for ("Parity") or overlooked and maintained ("Fairness"). Student course performance is important, but is not the only reason to consider zeros. There are also other negative consequences we have yet to consider. For example, behavioral zeros can harm student self-concept, self-efficacy and motivation [11], which may affect course retention or STEM identity. These relationships are also important to consider when determining grading policies.

We don't have any direct evidence that the reasons that different groups of students leave blanks or skip exams are different. However, the fact that different groups engage in these behaviors at different rates indicates that there are likely equity implications to consider. Further, the fact that female students, the group least likely to skip exams, are not also the group least likely to skip problems, suggests that these types of behaviors happen for different reasons and perhaps should not be awarded indistinguishable zeros as a consequence.

To illustrate the importance of thinking about zeros, we share a quote from an upper division undergraduate physics major explaining how they felt about attempting to get partial credit on exams when unsure about solving a problem completely accurately [12].

"*Physics is not bullshit. You shouldn't be writing bullshit answers, but you have to. I don't know, it feels a little dirty but at the same time I want a good grade on an exam. . . . Physics is not based on bullshit but here I am.*"

This student needed to sacrifice their value of submitting only high-quality solutions about which they felt confident because they realized that they needed to write something, even what they considered "bullshit" or guesswork, to earn credit on the exam. Indeed, researchers would not want to submit papers for peer-review based on possibly fraudulent assumptions or analysis, so why do we expect this of our students in test-taking scenarios? Furthermore, playing this "dirty" academic game created a conflict for this student between physics ideals and classroom experience, and could influence their choice to pursue physics in the future.

As instructors we want students to demonstrate the full extent of their understanding on exams so that each student can receive a grade that best represents their knowledge and skill in the subject, but we can also see the issue from the point of view of a student who does not want to pretend to possess knowledge or skill that they feel they do not have. This is only one possible reason students may leave blanks, but it illustrates the need to examine grading student behaviors that are not necessarily representative of their knowledge, especially since blank-leaving behavior may be more common for some groups of students than for others.

Instructors who wish to eliminate the use of behavioral zeros have some options. One possible solution is to adopt standards-based grading [13]. While this may be challenging for large lecture courses, it is possible. Another option is "minimum grading" [5], the practice of never assigning grades lower than 50%. An option that allows graders to assign behavioral zeros but weighs their impact more similarly to other letter grades is to use a GPA-like 4-point scale.

ACKNOWLEDGMENTS

We thank the Registrar at UC Davis for providing us with demographic data and the UC Davis and SJSU PER groups for useful feedback on this work. We want to specifically thank Benedikt Harrer for reviewing and providing feedback an early draft of this paper.

[1] I. Rodriguez, E. Brewe, V. Sawtelle, and L. Kramer, Phys. Rev. ST Phys. Educ. Res. **8**, 020103 (2012).
[2] N. M. Stephens, S. A. Fryberg, H. R. Markus, C. Johnson, and R. Covarrubias, J Pers Soc Psychol **102**, 1178 (2012).
[3] C. Stathopoulou and F. Kalabasis, Educ Stud Math, **64**, 231 (2006).
[4] D. Reeves , Phi Delta Kappan **86**, 324 (2004).
[5] T. Guskey, Educ Leadership, **71**, 68 (2013).
[6] C. Paul, D. Webb, M. Chessey, and W. Potter, Proc Phys Educ Res Conf 2017, <https://www.compadre.org/Repository/document/ServeFile.cfm?ID=14628&DocID=4805>.
[7] W. Potter, D. Webb, C. Paul, E. West, M. Bowen, B. Weiss, L. Coleman, and C. D. Leone, Am. J. Phys. **82**, 153 (2014) .
[8] C. Paul, W. Potter, and B. Weiss, Phys Teach **52**, 485 (2014).
[9] APS website defining URM is https://www.aps.org/programs/minorities/index.cfm
[10] We only include courses where we have a full set of exam scores and only include students who have scores on at least half of the exam items.
[11] A. Bandura, Educ. Psychol. **28**, 117 (1993) and ref's therein.
[12] M. K. Chessey, Ph.D. Dissertation, University of California Davis, 2018.
[13] I. Beatty, J. Scholarship of Teach. Learn., **13**, 1 (2013) and ref's therein.