Peer Grading Without Protest: The SPARK Approach to Summative Peer Assessment

Jennifer S. Kay
Computer Science Department
Rowan University
Glassboro, NJ 08028
kay@rowan.edu

ABSTRACT

As Computer Science Professors, we strive to construct courses that maximally support and contribute to student learning through carefully crafted in-class and out-of-class activities. There is evidence that homework enhances student learning, and that students are more likely to do their homework when it affects their grade. Thus, faculty often find themselves seeking the right balance between the increase in student understanding that additional graded homework might offer and the burden of grading that homework if automated grading software is not available. Peer grading may seem like the obvious solution, since it results in only a limited increase in student workload while still incentivizing homework completion. Unfortunately, students tend to distrust their peers' abilities to evaluate their work, and consequently previous approaches to summative peer assessment risk increasing both student frustration and faculty grading load.

This experience report describes SPARK, our unique approach to summative peer assessment that we have successfully used over several semesters for weekly problem sets in a theoretical computer science course as well as on a more limited basis in both undergraduate and graduate robotics courses. Surveys indicate that a majority of students find it easy to use the SPARK approach to grade their peers, believe that the homework assignments helped their learning, acknowledge that they would not have put the same time and effort into their homework had it not been graded, and believe that SPARK's method of computing their final grade for an assignment is fair.

CCS CONCEPTS

• Social and professional topics → Professional topics → Computing education → Student assessment

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
SIGCSE 2022, March 3–5, 2022, Providence, RI, USA
© 2022 Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-9070-5/22/03…$15.00
https://doi.org/10.1145/3478431.3499284

KEYWORDS

Summative Peer Assessment, Peer Assessment, Peer Evaluation, Peer Grading, Peer Review, SPARK, Assessment, CS Education

ACM Reference format:

Jennifer S. Kay. 2022. Peer Grading Without Protest: The SPARK Approach to Summative Peer Assessment. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V.1 (SIGCSE 2022), March 3–5, 2022, Providence, RI, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3478431.3499284

1 Introduction

The types of activities required of students in Computer Science (CS) classes can take many different forms, as can the ways faculty endeavor to motivate students to do those activities. Our undergraduate "Foundations of Computer Science" (FOCS) class (that covers topics such as formal languages and automata along with some formal logic) is one in which we feel strongly that student success is heavily dependent upon regularly working through practice problems. As a result, when we began teaching FOCS, we *required* students to complete and submit weekly problem sets. Unfortunately, grading FOCS homework is extremely time consuming, even when reviewing only a subset of the problems assigned. We experimented with further reducing the number of problems graded per homework to enable us to return them faster, but this left some students feeling that the resulting grades were a poor reflection of their overall performance on the assignment. Removing problem sets from the course grade entirely (changing them to "strongly recommended") left us with the sense that most of our students were spending much less time on their homework and as a result were not learning as much as they had in previous semesters.

We decided to experiment with using summative peer assessment as an alternate approach to incentivize student homework completion, since it could result in faster feedback with only a minimal increase in student workload. We knew that this would only succeed if students were confident that the final grade they received was not dependent on which of their peers graded their assignment. The result was our "Summative Peer Assessment while Rounding Kindly" (SPARK) approach to peer grading, which is the subject of this experience report.

We have used SPARK over several semesters for weekly FOCS homework assignments as well as on a more limited basis in both undergraduate and graduate robotics courses. From the faculty perspective, we are convinced that SPARK is the perfect approach to incentivize students to complete low-stakes assignments that cannot be easily auto-graded. Of equal importance, the majority of our students have favorable opinions of the SPARK process, in stark contrast to student opinion of traditional peer grading.

2 Related Work

2.1 The Impact of Homework on Learning

There are regular debates in both the literature and among the public over the value of homework [2] [23]. While the exact relationship and mechanisms between homework and student achievement remain unclear [22], meta-analyses of homework studies of K-12 students suggest that the impact of homework on student achievement increases with grade-level [7] and offer strong evidence of the positive relationship between homework and student achievement for high school students [8].

Undergraduates tend to believe that doing homework helps their learning [13][16][24]. There are fewer studies of undergraduates than K-12 students, but they do provide insight into some of the many variables at play. For example, Shepard & Law found no significant difference in exam performance between engineering undergraduates who completed graded homework problems and those who completed suggested ones [19]. In contrast, when Koban et al. factored in SAT scores, they found that graded homework significantly improved the exam performance of medium-ability mathematics students, but not the exam performance of low- or high-ability students [13].

It is not shocking to learn that there is evidence that high school students are more likely to do homework when it impacts their grade [9]. Undergraduate engineering students also report spending more time on graded homework than on suggested problems [19]. In addition, there is evidence that increased incentives correlate to increased homework completion rates at both the high school [9] and undergraduate [18] levels. It is worth noting, however, that larger incentives on homework will likely lead to increases in academic misconduct [16][19][24].

2.2 Summative Peer Assessment

Topping performed a meta-analysis of peer assessment[1] studies in higher education from the 1970s through the 1990s and found that peer assessment can have a positive effect on student learning [20]. In this seminal work, he also introduced a 17-dimensional topology of peer assessment that provides an overarching view of the topic [20]. Shortly thereafter, Falchikov & Goldfinch followed with a meta-analysis of peer assessment in higher education that stretched back as far as 1959 that was more narrowly focused on grades [10]. They found that peer grades generally agree with instructor grades when students are given clear grading criteria [10]. More recently, Li et al. performed a meta-analysis of peer assessment that covers both K-12 and higher education settings that concurs that peer assessment has a positive effect on student learning [15]. Li also found that the most important factor in this is rater training [15]. While no significant differences were found between undergraduate informatics students' assessment of the work of "friends" and "non-friends" [3], gender has been shown to have a slight effect in male undergraduates' assessment of oral presentations [1], and it seems sensible to recommend that peer assessment should be blind whenever possible. Finally, while there is evidence that the act of reviewing a peer's work can have a positive impact on the assessor [21], we see that as potential "icing on the cake" rather than a motivating factor.

2.3 Assigning Grades into Discrete Bins

There is anecdotal evidence that faculty grading using a binary ("satisfactory" vs. "unsatisfactory") scale (together with the option for assignment resubmission) has been successful in several different undergraduate CS classes [4].

In the context of peer assessment, there is some evidence that using a limited set of grading bins rather than a continuous numeric scale may result in better correlation between student and instructor grades. Falchikov & Goldfinch [10] felt it necessary to include a "Cautionary Note" about a study by Burnett & Cavaye [6] of fifth year medical students which found "an almost perfect correlation between peer assessment and final grade (r = 0.99)" despite not giving clear grading criteria [10]. Falchikov & Goldfinch hypothesized that this was this was a result of a grading scheme in which qualitative grades (from "Outstanding" down to "Bad Failure") [6] corresponded with percentage ranges (from "85-100" down to "<35") [6] and thus it was more likely that student and instructor grades would correlate with each other [10]. Kritikos et al. used a similar qualitative grading scheme with five "bins" among groups of undergraduate pharmacy students in an attempt to increase their engagement in class presentations and found that ratings assigned by collaborative student groups generally matched instructor ratings [14].

2.4 Student Satisfaction with Peer Assessment

Students are positive about formative peer assessment, but "highly critical" of summative peer assessment [17]. In fact, student opinions about summative peer assessment are independent of the actual level of agreement between student and instructor grades [21]! This does not bode well for summative peer assessment as a tool to reduce instructor workload since there is the potential that large numbers of students might dispute the grades assigned by their peers. In a study of over 400 undergraduates in a "Communication Skills for

[1] Note that the terms "peer assessment", "peer review", "peer evaluation", and "peer grading" are often used interchangeably. "Peer assessment" is the language that appears to be favored by the seminal literature. However, in the SPARK context, we often choose to use "peer grading" because that is our students' primary task. We also use "peer review" in the context of a Learning Management System (LMS) that uses that language.

Engineers" course, Zrnić et al. found that for each of five summative peer assessments the majority of students felt their grades were within 5% of what they expected [27]. However, almost 30% of students reported expecting a higher grade on one assignment, and almost 20% of the students reported their grade was actually higher than expected on another assignment. Overall, roughly one third of students disagreed with the grades their peers had assigned them, attributing it to assessors being "too subjective or incompetent to assess their work." [27]

Using a limited set of grading bins does not necessarily correlate with student satisfaction. In the pharmacy student study described above, only 43.6% of students agreed that their peers are capable of fairly assessing their work [14].

3 Peer Assessment: Potential Points of Failure

While the research community is still working to gain a better understanding of the impacts of homework on student achievement, its ubiquity throughout the undergraduate curriculum is evidence that many faculty (ourselves included) believe it can play a big part in students' understanding of course material. Since the research suggests that linking completion of the homework to the final grade will incentivize our students to do their homework problems, peer assessment had the potential to solve our FOCS grading conundrum. However, the peer assessment process has multiple potential points of failure, many of which might result in significant increases in faculty load:

- Assessors may be motivated to do peer grading …
  - … but not understand how to do so
  - … but find the submission illegible
- Assessors may not be motivated to do peer grading …
  - … and assign grades haphazardly, with a focus on quick completion rather than accurate assessment
  - … and skip the peer grading process altogether
- Assessors may be biased if the review process is not blind
- *Assessees* may disagree with the grade assigned by their peers and ask the instructor to regrade their assignment

4 SPARK: Peer Grading Without Protest

4.1 Process Overview

4.1.1 Bootstrapping: Facilitating Anonymity. At the start of the semester, students pick an anonymous ID by selecting the name of a color from a long list [25]. We ask them to include their color name at the beginning of the filename for each assignment. This is not required by the Learning Management System (LMS), but gives students a means to identify assignments they are assessing if they have a question for the instructor. We also ask them to be sure that their real name is not included in the filename or in the body of the assignment so that the submissions are anonymous. Unfortunately, some LMSs list the assessee's name next to their written feedback, precluding double blind reviewing.

4.1.2 Bootstrapping: Starting With a Practice Assignment. We familiarize the students with the mechanics of the LMS assignment submission and peer assessment process by giving them a practice assignment early in the first week of class that involves:

- Hand-drawing a smiley face on a piece of paper
- Handwriting a sentence about the importance of homework on a second piece of paper
- Uploading a scan of the two pages as a single pdf

Much of the peer grading for the practice assignment is concerned with ensuring that the students followed the directions (e.g., "Is it easy to read (handwriting legible, good contrast, etc.)?", "Is everything hand-drawn and handwritten?") For this practice assignment, we also share the grading rubric in advance; on the regular problem sets, students are not told which problems will be graded until after submission.

4.1.3 Encouraging Legibility In Handwritten Assignments. Problem sets in FOCS often involve special symbols and diagrams that can take significant time to render electronically. Our feeling is that this may actually detract from learning the content, so we encourage students to do many homework problems with a paper and pencil and then convert multiple pages of handwritten solutions into a single pdf document. More often than not students do this using a phone app, which can lead to variable results. We hope that by routinely including "legibility" as one of the grading criteria from the very beginning (with the "smiley" practice assignment) the students will endeavor to provide their peers with readable work.

4.1.4 Establishing A Regular Routine. In the case of FOCS (where we have weekly assignments), we try to keep the class routine the same from week-to-week so that students become accustomed to the standard cycle. Figure 1 shows a typical timeline for a Monday/Wednesday class.

4.1.5 Getting Started With The Problem Set. At the beginning of week w, we introduce new content C(w) in class. The C(w) problem set is due at the end of Monday of week w+1.

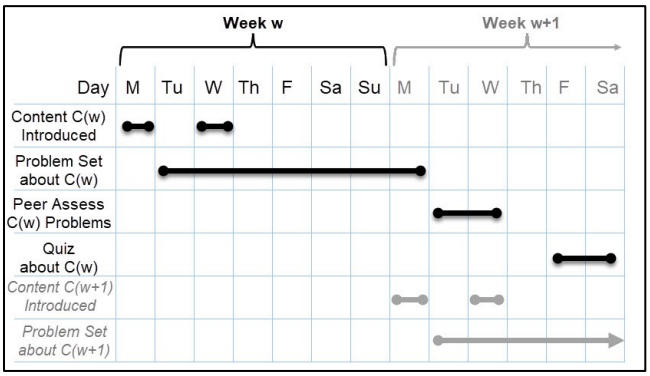

Figure 1: A typical timeline for the FOCS class. Peer assessment of week w's problem set & a quiz on week w's content occur in week w+1, overlapping with the new content & problem set introduced in week w+1.

4.1.6 Assigning The Reviewers. Immediately after the problem set due date, the LMS automatically shuffles the submissions in such a way as to assign three different students to grade each submission, and so that each individual has three assignments to grade. Three peer graders per assignment seems to be a good balance between ensuring that each assignment will be graded (it seems unlikely that all three peers would skip doing their grading) and not overwhelming students with too much grading. The LMS ensures that there is no self-grading.

4.1.7 Revealing The Assessment Solutions. Three hours after the assignment is due, the LMS automatically reveals the assessment solutions. This short break offers us the flexibility to accept late submissions (should we choose to do so) in response to the inevitable "the system froze as I was uploading" appeals.

The assessment solutions specify which subset of problems to grade and provide model solutions for each. Students are instructed to use the 0-1-2 grading scale described in Figure 2.

SPARK grading in FOCS typically asks students to grade ten problems for each peer and also assign an 11th grade in which they rate (again on a 0-1-2 scale) the legibility of the submission.

4.1.8 From Raw Score To Final Grade. After the students have submitted their peer grades on the LMS, the average (mean) grade is computed for each submission. We refer to this average as a student's "raw score" for the assignment. The final grade for the assignment factors in both the raw score as well as student participation in peer assessment as shown in Figure 3.

We consider this computation to be the key factor in students' acceptance of SPARK for two reasons. Most importantly, our liberal approach to "rounding up" the raw score reduces the impact that a difference of a few points in the raw score would have on a student's final grade. This has virtually eliminated grade disputes. Of course, for any type of peer grading to work it is essential that students participate in the grading process, and so that is strongly incentivized by SPARK's final grade computation.

4.2 Discussion

4.2.1 Early Experimentation With Peer Assessment In FOCS. We began experimenting with peer assessment in FOCS in 2016, asking students to provide both written feedback and 0-1-2 grades for each of the problems in the assessment guidelines. Students seemed to pick up the 0-1-2 grading immediately without much additional explanation beyond what is given in Figure 2. In contrast, the result of our lack of instruction in written feedback was (as predicted by the research) very lackluster. The first few semesters we assigned three different types of grades for the FOCS homework: weekly grades for each problem set using the SPARK computation described above, together with occasional grades for instructor "spot checks" of homework sets and "spot checks" of peer assessment. After experimenting with different permutations of spot checks over time, we grew to believe that spot checks were of little or no benefit to the students and did not influence their behavior. By 2019, we had entirely eliminated spot checks, deemphasized written feedback, and focused on 0-1-2 grading using the SPARK approach described above.

Peer Evaluation: Assigning Numeric Grades:

0 points: Either of the following is true:

- No answer.
- Really bad answer.

1 point: Either of the following is true:

- It's a problem from the textbook and the answer is just what is written in the back of the book with no other work shown.
- The answer has some significant errors in it.

2 points:

- The answer is essentially correct. More information is given than simply copying the answer from the back of the book. There may be minor nitpicky problems, but not enough to drop it down to 1 point.

Figure 2: Student Instructions for 0-1-2 Peer Grading.

Did you complete all the peer evaluations assigned to you?	"Raw Score" (== Average (mean) grade you received from your peers)	Your final grade on the assignment
YES	avg >= 70%	100
	50% <= avg < 70%	70
	avg < 50%	50
NO	avg >= 70%	50
	avg < 70%	0

Figure 3: SPARK Final Grade Computation. SPARK both incentivizes peer grading and minimizes the impact of small changes in raw score on the final grade.

4.2.2 How Much Should SPARK Grades Contribute To A Student's Overall Course Grade? Since 2019, SPARK grades on the weekly problem sets in FOCS have constituted 10% of a student's final grade. This may seem like a very low percentage, particularly in light of the fact that the homework problem sets likely represent 75% of the class workload. However, we feel it is important to limit the contribution of SPARK grades to a student's final grade for three reasons. First, we want to avoid making students feel that the homework grade is so important that they should simply copy the solutions that appear on various websites with annoying frequency. Second, we recognize that assessors sometimes assign grades that have little relationship to their assessee's work. Typically, this is in the form of assigning 2 points (i.e. full credit) to each of their assessee's problems, presumably in an attempt to get credit for doing peer assessment without actually doing any work (but at the same time not attracting the ire of their assessees). This behavior tends to come to the attention of the instructor when the scores assigned by the other two assessors are dramatically

lower. Third, by its very nature, SPARK's "generous rounding" does not do a good job of differentiating between students' levels of understanding. Thus, we strongly recommend that SPARK only be used on low-stakes assignments that represent a small proportion of a student's overall grade for the semester.

4.2.3 Limiting Student Bias Through Anonymous Submission. It is worth noting that when a class of 30 students select their secret ID from the list of roughly 1000 colors there are often some duplicates. This has no actual impact on the process within the LMS, but students find it disconcerting if, for example, they notice someone else with their ID or discover they have been given two different assignments with the same label. We recommend quickly reviewing their color choices at the start of the semester and ask any duplicates to pick new colors.

As discussed above, the research suggests that asking students to pick an anonymous identifier is unlikely to have a large impact on their grade [1][3]. Nevertheless, we believe that some students might be more comfortable with the anonymity, and it also may serve to indicate the value we place on fairness in our system. Whether or not the process will be double blind is LMS dependent. Canvas [11], for example, allows the instructor to specify that the reviewee's name is hidden from the reviewer, but the reviewer's name is visible to reviewees in the comments.

Of course, students may choose to share their anonymous identifier with their friends in an effort to inflate their grades, although as noted above, research suggests this is unlikely [3]. Anecdotally, our experience is that most of the time when we see a reviewer give one peer uncharacteristically high grades, that reviewer does the same for their other peer assessments. We no longer perform "spot checks," but when we notice this sort of behavior, we inform the student that we are not giving them credit for peer grading that week.

4.2.4 LMS Issues: Converting Raw Scores To Final Grades. We have used this approach with both the Blackboard LMS [5] and the Canvas LMS [11] and each has its own eccentricities. Blackboard has built in functionality that can compute the mean of peer-assigned grades automatically. In contrast, Canvas requires the instructor to either manually check every peer grade for each assignment or to use a non-trivial method that utilizes the Canvas API to extract the data [12].

4.2.5 LMS Issues: Ensuring Everyone Can Peer Grade. Canvas permits students to submit peer assignments late, but those who do are not automatically assigned peers to grade, which means that their maximum SPARK final grade cannot exceed 50%. Thus in our early experiments with SPARK using Canvas, it was not uncommon to receive emails from students whose assignment submissions were "literally 30 seconds late" [sic] asking to participate in the peer assessment process. Facilitating this as the instructor takes time, and so we developed a workaround that we call the "pre-upload". Students are encouraged to upload *anything* (e.g. a picture of their cat) on the day the assignment opens so that the LMS will be sure to include them in the peer assessment process, even if they experience an Internet failure as they try to upload their final assignment at the last minute. The three-hour delay between due-date and sharing the solutions offers an opportunity to complete the upload.

5 Student Attitudes

5.1 Student Surveys

In 2019 and 2020, we used SPARK in eight sections of FOCS in face-to-face and remote (both asynchronous and synchronous) modes. Towards the end of each semester, students completed an anonymous survey that included 8 questions using a Likert-type scale (Strongly Agree (5), Agree (4), Neutral (3), Disagree (2), Strongly Disagree (1)). Several questions were inspired by [26]. The questions are shown in Figure 4; the data are summarized in Table 1 and shown graphically in Figure 5.

5.2 Discussion of Student Responses

5.2.1 Students See Value in Both Doing the Homework Problems and in Reading the Solutions. The vast majority of students (87.42%) agree or strongly agree that doing the problem sets helped their learning (Q1), and more than three quarters (75.95%) agree or strongly agree that reading the solutions also helped their learning (Q2).

5.2.2 Students are Split on Whether the Peer Review Process Helps Them Learn the Material. So are we! 40.25% of students agreed or strongly agreed that the peer grading helped their learning, while 30.82% disagreed or strongly disagreed (Q3). Perhaps student ability had an impact on the results similar to the math students discussed above and in [13]. It is certainly plausible that students could learn from doing their peer

Q1: Doing the assigned homework problems helped me to learn the class material

Q2: Reading the homework solutions helped me to learn the class material

Q3: Reviewing my peer's work helped me to learn the class material

Q4: It was easy to decide what grade to assign to my peers for the problems I graded

Q5: If the professor had assigned the same homework problems, but had not required me to submit them, I would still have put the same amount of time and effort into completing them.

Q6: I spent enough time doing the assigned homework problems

Q7: I spent enough time reviewing my peers' work

Q8: I feel as though the approach to computing my overall grade on a peer review assignment (by combining multiple factors, including whether I've done my reviewing and rounding up my peers' grades) is fair

Figure 4: Student Survey Questions

	N	Mean	SD
Q1 (HW Helped Learning)	159	4.35	.879
Q2 (Reading Solutions Helped Learning)	158	4.00	.867
Q3 (Peer Review Helped Learning)	159	3.08	1.180
Q4 (Assigning Grades Easy)	159	3.60	.942
Q5 (Same Effort if Not Required)	159	2.52	1.321
Q6 (Spent Enough Time on HW)	159	3.95	1.101
Q7 (Spent Enough Time on Peer Review)	159	3.74	1.052
Q8 (Grading Approach is Fair)	159	3.97	1.058

Table 1: Summary of Student Survey Data

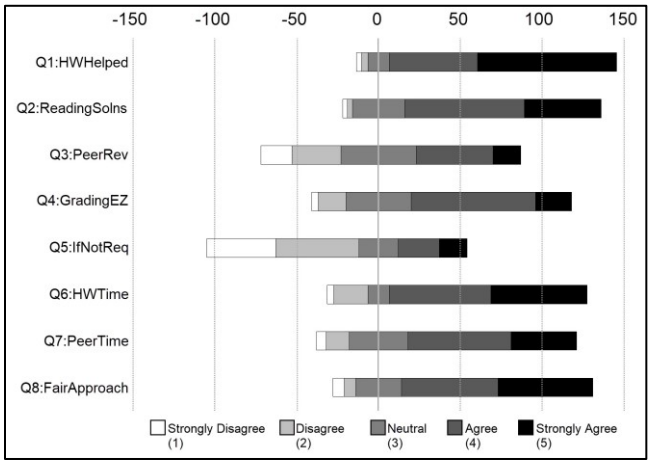

Figure 5: Visualization of Student Survey Data. Each segment of a bar represents the number of students who selected that choice. The neutral choice is always centered on the zero axis.

grading; many FOCS questions have multiple (and sometimes, infinite) solutions, so students may indeed benefit from seeing peer solutions that differ from the model solution(s) provided.

5.2.3 Students Think SPARK Grading Is Straightforward. The majority (61.64%) of students agree or strongly agree that it was easy to decide what grade to assign their peers (Q4). While we are pleased with this result, in retrospect we wonder if the question's wording was a bit ambiguous. For example, students might equate "easy" with "quick," and with 11 problems to grade for each of three peers, it might not have felt "easy". In addition, if a student is grading a solution that is quite different from the model solution(s), they may find it more difficult to grade.

5.2.4 Students Acknowledge That They Would Not Have Put In the Same Amount of Time and Effort Without SPARK. The majority of students (58.49%) reported that they would not have put in the same time and effort had the homework problem sets not been required (disagree or strongly disagree with "I would still have put the same amount of time and effort") (Q5). The 26.42% on the other side (agree or strongly agree) is larger than we expected, perhaps because of Q5's negative wording, which flips the direction of the rating scale.

5.2.5 Students Spend "Enough" Time On Their Homework And Peer Grading. The majority of students agree or strongly agree that they spent enough time on the homework (76.10%) (Q6) and peer grading (64.78%) (Q7).

5.2.6 Students Think That SPARK Grading Is Fair Almost three-quarters of students (73.58%) agree or strongly agree that our approach to computing their final grade is fair (Q8).

The meaning of "fair" is, of course, subject to interpretation. Some students may be happy with their grades and conclude that SPARK is fair. Others may come to the opposite conclusion because students who do "A" and "B" work get the same grade.

Regardless of interpretation, a mere 8.81% of students disagree or strongly disagree. In other words, 91.19% of students were neutral or positive about SPARK's fairness.

We believe that this result strongly supports SPARK's value as a tool to incentivize student completion of low-stakes homework assignments while minimally impacting the instructor's grading load.

6 SPARK in Robotics Classes

We have also used SPARK for a small number of assignments in both undergraduate and graduate robotics classes. Since SPARK incentivizes timely completion of projects without adding significantly to the instructor's grading load, we can use it for small programming projects earlier in the semester as well as for intermediate checkpoints on larger projects that otherwise might not have been graded. This can help students and faculty become aware of problems earlier. A sample robotics task might ask students to program their robot to follow a dark line on a light background. Students are asked to create a video of their work, post it on YouTube or a similar site (unlisted if they choose), and then to submit a URL of their video on the LMS. Reviewers watch three videos and use a 0-1-2 scale to confirm that the robot correctly performs the maneuver as well as give a 0-1-2 rating for "viewability" to encourage video clarity.

7 Concluding Thoughts

SPARK has enabled us to *require* rather than *suggest* homework problem sets in FOCS as well as programming assignments in robotics classes. SPARK incentivizes our students to do the work that we believe is essential for their success without significant increases in faculty grading load or the need for specialized auto-graders. Furthermore, in stark contrast to traditional forms of summative peer assessment, students are overwhelmingly positive about the SPARK approach to peer grading.

From a faculty perspective, we have already found SPARK to be an invaluable tool in two very different CS courses and we plan to include SPARK in other CS courses in future semesters. In addition, we are intrigued by the fact that many FOCS students reported that the act of grading one's peers using a 0-1-2 scale *increased* their understanding of the material. We look forward to digging deeper into the impact that SPARK might have, if any, on student learning.

REFERENCES

[1] Michael Mogessie Ashenafi. 2017. Peer-Assessment in Higher Education - Twenty-First Century Practices, Challenges and the Way Forward. *Assessment and Evaluation In Higher Education* 42, 2 (2017), 226–251. DOI: https://doi.org/10.1080/02602938.2015.1100711

[2] Gökhan Baş, Cihad Şentürk, and Fatih Mehmet Ciğerci. 2021. Homework and Academic Achievement: A Meta-Analytic Review of Research. *Issues in Educational Research* 27, 1 (January 2017), 31-50.

[3] Veronika Bejdová, Zuzana Kubincová, and Martin Homola. 2014. Are Students Reliable Peer-Reviewers? *2014 IEEE 14th International Conference on Advanced Learning Technologies*, 2014, 270-272. DOI: https://doi.org/10.1109/ICALT.2014.84

[4] Andrew Berns. 2020. Scored out of 10: Experiences with Binary Grading across the Curriculum. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education (SIGCSE '20)*, March 11-14, Portland, Oregon, ACM Inc., New York, NY, USA, 1152–1157. DOI: https://doi.org/10.1145/3328778.3366956

[5] Blackboard. 2021. Blackboard, Learning Management Systems (LMS) and Software. Retrieved July 2021 from https://www.blackboard.com/teaching-learning/learning-management.

[6] Burnett, W., and G. Cavaye. 1980. Peer Assessment By Fifth Year Students Of Surgery. *Assessment in Higher Education* 5, 3 (1980), 273-278.

[7] Harris Cooper. 1989. Synthesis of research on homework. *Educational Leadership* 47, 3 (Nov. 1989), 85-91.

[8] Harris Cooper, Jorgianne Civey Robinson, and Erika A. Patall. 2006. Does Homework Improve Academic Achievement? A Synthesis of Research, 1987–2003. *Review of Educational Research* 76, 1 (March 2006), 1-62. DOI: https://doi.org/10.3102/00346543076001001

[9] Francis T. Cullen Jr, John B. Cullen, Van L. Hayhow Jr, and John T. Plouffe. 1975. The Effects of the Use of Grades as an Incentive. *The Journal of Educational Research* 68, 7. (March 1, 1975), 277-279.

[10] Nancy Falchikov and Judy Goldfinch. 2000. Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks. *Review of Educational Research* 70, 3 (September 2000), 287–322, DOI: https://doi.org/10.3102/00346543070003287

[11] Instructure. 2021. Canvas LMS | Learning Management System. Retrieved July 2021 from https://www.instructure.com/canvas

[12] Instructure Community. 2017. Download all peer reviewers/scores using Rubrics API. Retrieved July 2021 from https://community.canvaslms.com/t5/CanvasLive/Download-all-peer-reviewers-scores-using-Rubrics-API/ba-p/382946

[13] D. Koban, M. Fukuzawa, R. Slocum, M. Fletcher & J. Pleuss (2020) Differential Effects of Incentivized Homework on Student Achievement in Undergraduate Mathematics, *PRIMUS: Problems, Resources, and Issues in Mathematics Undergraduate Studies* 30, 5 (May 2020), 501-519, DOI: https://doi.org/10.1080/10511970.2018.1530703

[14] Vicky S. Kritikos, Jim Woulfe, Maria B Sukkar, and Bandana Saini. 2011. Intergroup Peer Assessment in Problem-Based Learning Tutorials for Undergraduate Pharmacy Students. *American Journal of Pharmaceutical Education* 75, 4 (2011), 73–73. DOI: https://doi.org/10.5688/ajpe75473

[15] Hongli Li, Yao Xiong, Charles Vincent Hunter, Xiuyan Guo, and Rurik Tywoniw. 2020. Does Peer Assessment Promote Student Learning? A Meta-Analysis. *Assessment & Evaluation in Higher Education* 45, 2, 193–211. DOI: https://doi.org/10.1080/02602938.2019.1620679

[16] Angela L. Minichiello, Laurie S. McNeill, and Christine E. Hailey. 2012. Comparing Engineering Student Use of Solution Manuals and Student/Faculty Perceptions of Academic Dishonesty. In *Proceedings of 2012 ASEE Annual Conference & Exposition*, June 10-13 2012, San Antonio, Texas. ASEE Conferences, 25.330.1 - 25.330.19. DOI: https://doi.org/10.18260/1-2--21088

[17] Chloe Patton. 2012. 'Some Kind of Weird, Evil Experiment': *Student Perceptions of Peer Assessment. Assessment and Evaluation in Higher Education* 37, 6 (September 2012), 719–731. DOI: https://doi.org/10.1080/02602938.2011.563281

[18] Phanikiran Radhakrishnan, Dianne Lam, and Geoffrey Ho. 2009. Giving University Students Incentives to do Homework Improves their Performance. *Journal of Instructional Psychology* 36, 3 (Sept. 2009).

[19] Thomas Shepard and Deify Law. 2015. A comparison of student learning between graded homework and suggested problems. In *Proceedings of 2015 ASEE Annual Conference & Exposition*, June 14-17, 2015, Seattle, WA, ASEE Conferences. 26.24.1 - 26.24.12. DOI: https://doi.org/10.18260/p.23365

[20] Keith Topping. 1998. Peer Assessment between Students in Colleges and Universities." Review of Educational Research, 68, 3 (Autumn 1998), 249–276. DOI: https://doi.org/10.2307/1170598

[21] Keith J. Topping. 2009. Peer assessment. *Theory Into Practice* 48, 1 (Oct 2009), 20-27. DOI: https://doi.org/10.1080/00405840802577569

[22] Ulrich Trautwein & Olaf Köller. (2003). The Relationship between Homework and Achievement - Still Much of a Mystery. *Educational Psychology Review* 15, 2, 115-145. DOI: https://doi.org/10.1023/A:1023460414243

[23] Cathy Vatterott. 2018. Rethinking Homework: Best Practices That Support Diverse Needs (second edition) Association for Supervision & Curriculum Development, Alexandria VA.

[24] James Widmann and Kim Shollenberger. 2006. Student Use of Textbook Solution Manuals: Student and Faculty Perspectives In A Large Mechanical Engineering Department. In *Proceedings of 2006 ASEE Annual Conference & Exposition*, June 2006, Chicago, Illinois. ASEE Conferences. 11.1168.1 - 11.1168.9. DOI: https://doi.org/10.18260/1-2--432

[25] Wikipedia. 2021. List of colors (compact). Retrieved July 2021 from https://en.wikipedia.org/wiki/List_of_colors_(compact)

[26] Ryan Zerr. 2007. A Quantitative and Qualitative Analysis of the Effectiveness of Online Homework in First-Semester Calculus, *The Journal of Computers in Mathematics and Science Teaching* 26, 1, Association for the Advancement of Computing in Education (AACE), 2007, p. 55–73.

[27] L. Zrnić, L. Korov, J. Petrović and P. Pale. 2020. Students' Perception Of Summative Peer Review Grading, In *Proceedings of The 43rd International Convention on Information, Communication and Electronic Technology (MIPRO 2020)*, 1526-1530, DOI: https://doi.org/10.23919/MIPRO48935.2020.9245112