From Pointless to Point-Less Grading

A Call for a Better Grading Process

Andrew Berns
University of Northern Iowa
andrew.berns@uni.edu

J. Philip East
University of Northern Iowa
east@cs.uni.edu

ABSTRACT

For as long as we can remember, grading has felt like one of the necessary evils of instruction. We graded in much the same way as has been done for decades, totaling points which were earned for a variety of reasons. We had nearly resigned ourselves to accepting the frustrations such a traditional grading system brought with it for students and instructors alike. Recently, however, we have come to believe grading does not need to be a frustrating and fruitless task and can augment our instruction instead of undermine it. In this position paper, we discuss a new grading paradigm which has revolutionized the way in which we teach, and which we believe can do the same for the entire computer science education community.

CCS CONCEPTS

• **Social and professional topics** → *Computer science education*; *Student assessment*;

KEYWORDS

grading systems, equity, assessment, diversity

ACM Reference format:
Andrew Berns and J. Philip East. 2020. From Pointless to Point-Less Grading. In *Proceedings of 51st ACM Technical Symposium on Computer Science Education, Portland, OR, USA, March 11 - March 14, 2020 (SIGCSE '20),* 7 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 WHAT IF . . .

We love teaching and hate grading. Students clearly benefit from good feedback, and as experts in our field we can often generate a lot of it. However, grading generally involves more than just providing feedback – it requires (or so it seems) that we assign points to student work so we can record a fair and accurate indication of student capability at the end of the course. We have spent hours agonizing over issues such as the deduction for missing semi-colons in handwritten code, or the penalty doled out for an off-by-one error in a loop. Compounding the frustration, students often look at the assigned score and stop there, ignoring the "wisdom" of our years of expertise poured out in our feedback and instead taking away only the number written on the first page. This number can also be deflating to students and may quickly damage motivation and morale in a class. Unfortunately, it seemed that the frustrations of grading were simply an unavoidable unpleasantness one had to accept as an instructor. For a while, we had resigned ourselves to our fate and continued on with grading as a pointless exercise.

But, what if the unpleasant aspects of grading were *not* inevitable? What if we had a way of grading that turned our grading dystopia into a utopia? It would be a relief if we had a grading practice where …

- …grades were an accurate reflection of each individual's capabilities with respect to desired course outcomes.
- …grades were more equitably assigned, free from influences and experiences outside of our students' control.
- …we did not need to spend extensive energy making sure students do not plagiarize someone else's work.
- …we no longer had conflicts with students over the number of points awarded.
- …students were motivated and took ownership of their own learning, seeing the connection between their effort and their grades.
- …the course design and assessments used were synchronized, leading to a clear understanding of how what we ask students to do aligns with the course objectives.
- …our time grading is spent applying our expertise efficiently and not acting as an accountant of points.
- …we spent less time grading and more time thinking about how we can improve our students' learning.

Recently we have come to believe our ideal grading practice does in fact exist. The axiom guiding our examination of grading practice is that the goal of teaching is student learning or at least that at the end of instruction students have the desired capability. Students start at different places, learn at different paces, have individual circumstances that affect the time available for our classes, etc. Our primary concern is that *at the end of our course*, the student knows and is able to do those things the course is meant to teach. The grades we assign should be designed to reflect student knowledge and capability at that time.

In this paper we share how we think a significantly bettter grading practice can be achieved. We describe our past grading practice to provide a basis for understanding what we propose, present the key ideas of equitable grading and provide possible exemplars for the practice of equitable grading in an effort demonstrate it as being both possible and practical, and argue that our (the computer science education community's) grading practices should change substantially and with all due haste.

SIGCSE '20, March 11 - March 14, 2020, Portland, OR, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 OUR PRIOR PRACTICE

Generally, teachers teach how they were taught and try out new ideas as they are encountered. The way we were taught consisted mostly of presentations on content via lecture followed by the assignment of some homework, with the occasional inclusion of relatively small quizzes and/or larger examinations. Typically the products we produced were examined and points for each was awarded. An overall course grade was determined by some weighted combination of those points normalized to a percentage score and attendant grades were assigned, e.g., F (0-60%), D (60-70%), C (70-80%), B (80-90%), and A (90-100%). As students, we did well in this system which we felt was fair and accurate, and this traditional points-based approach was the basis for our own teaching and grading.

Some additional elements might work their way into our grading practice depending on the course and on our state of mind when producing syllabi. Often attendance and class participation were included when determining grades. Occasionally, extra credit activity was included as a mechanism to allow students to make up for some missed or poor work from earlier in the course. Non-trivial systems for managing late work were introduced, such as deducting a certain percentage for every time unit late, or allowing students a fixed number of late days to use. In some upper division courses (and occasionally in lower division courses) we would assign group projects. On those occasions we would typically assign a similar grade to all group members. Sometimes those group scores would be adjusted based on group members' assessments of contributions by the various group members.

"Grading" student submissions started with the identification of flaws. In the homework, perceived flaws were noted and suggestions for improvement provided as feedback. Typically, a submission's grade was determined by deducting points from a total possible amount. Sometimes homework grading and feedback were provided relatively quickly but more often a week or even more time might elapse between submission and feedback. Quizzes and exams were graded more quickly as feedback was typically provided by reviewing the examination items in class. Knowing we might make mistakes while grading caused us to allow students to question the grading process. While a few students would focus on their understanding, typical grading interactions with students focused on the number of points awarded and whether/how the student might get more points.

We worked within this traditional grading system for years, making occasional tweaks attempting to improve our grading practice. We would hear or think of something that might be better and give it a try. For example, we examined (and sought from others) alternatives in grading [1]. We examined the use of in-person grading [1].Workshops were attended to consider incorporating substantially different instructional approaches such as *Peer-Led-Team-Learning* (see for example [10]) and *POGIL* (Process Oriented Guided Inquiry Learning) [3]. We tried implementing *exam wrappers* [4], where students were provided an opportunity to complete metacognitive work (reflection) before or after exams. For years we continued with these piecemeal improvements in the hopes of improving our instruction.

2.1 Persistent Difficulties

Most of the instructional improvements above, however, seemed to provide little or no impact on our overall approach to grading and we continued to encounter problems. Students that demonstrated competence later than our arbitrarily selected deadlines were penalized even if they learned the concept eventually. A poor performance early on in the course could sink a student's grade for the remainder, regardless of their future actions, which could crush a student's morale and motivation for the course and the discipline in general. Students who were already competent in what we were trying to teach also lost motivation as they were forced to complete assignments to help them "learn" what they already knew, a task they justifiably viewed as "busy work". Scores were also often based on things unrelated to course outcomes. For instance, a student might get lucky with their group assignment and receive a passing grade even if they individually learned little or nothing. All too often, the hours we spent providing feedback were wasted, with students simply looking at the numeric score and repeating the identified mistakes. The majority of the time, discussions with students about their prior performance was as part of a grading dispute where a student felt their work warranted a few more points than we had assigned.

While these frustrating grading problems would cause us to frequently lament the state of "students these days", in hindsight these problems were all predictable outcomes with our traditional way of grading. If a student will not be asked to perform a nearly-identical task again, why should they care about our suggestions for improvement? If students are penalized for missing deadlines even for reasons that might be outside of their control (like illness, a family emergency, or limited time outside of school due to other responsibilities), and these penalties can damage their entire grade, why wouldn't they resort to any means necessary to meet our deadlines, even if it means plagiarizing someone else's work? If poor performance early on can damage a student's grade, why should they bother working hard if they get off to a bad start? Why should a bright student develop a passionate interest in more advanced topics when they are being forced to repeatedly perform tasks to learn what they already know? If the difference between a letter grade comes down to the total of a few points here and there, why shouldn't a logical student expend significant effort in having us allocate more points to them? If we need to assign a specific point value from a large range to everyone's work, why wouldn't we expect it to take us hours to do so fairly and consistently? The fact of the matter was our grading practice was pointless: it did not help others know how much a student knew and could do, it did not help students stay motivated and work to improve, and it did not help us teach more effectively.

2.2 Toward a New Paradigm

Recently we have been working on focusing on what we want students to be able to *do* instead of just *know*. This practice of looking more specifically at course outcomes has seemed a step in the right direction. An outcomes orientation allowed us to approach course design with a new mindset, focusing specifically on capabilities rather than rote knowledge alone. With programming courses, this seems rather obvious, but it is less so with other types of courses.

For example, in a programming course the the goal is not for students to know language features, but rather that they be able to apply language features in a variety of contexts with a variety of data to solve a variety of problems. In our algorithms course, it is more important that students apply complexity analysis when they design algorithms, for instance, rather than regurgitate the formal definition and complete lengthy proofs. In our software engineering course, it is important that students demonstrate an understanding of how various activities can work together to create software in a systematic and disciplined way, and not simply be able to list the twelve practices of extreme programming.

Still, as we were working on our instructional practice, and to a lesser extent grading practice, we were not very satisfied. The explorations had some effect on the instructional process but little or no effect on grading activity. Our attitude and understanding of grading mostly remained as it was. What we lacked was a grading ethos: a set of tenants to guide our entire evaluation practice. We found this when we encountered the idea of *equitable grading*.

In his book [8], Feldman makes the argument that the traditional points-based grading system fails to accurately convey a student's achievement in a course for a number of reasons, including grading items unrelated to student performance such as attendance, disadvantaging students with less prior experience than others, and variably (and somewhat arbitrarily) selecting the weighting for various items. Feldman urges educators to instead adopt a more equitable approach to grading, which was exactly the overarching set of principles we were seeking in improving our grading practice.

3 IMPLEMENTING EQUITABLE GRADING

Feldman [8] identifies the three pillars of equitable grading (p.71) as:

- Grading must/should be **mathematically accurate**, validly reflecting a student's academic performance.
- Grading must/should be **bias-resistant**, preventing non-academic or biased subjectivity from infecting grades.
- Grading must/should **motivate students** to strive for academic success, persevere, accept struggles and setbacks, and to gain critical lifelong skills.

These principles and a common sense understanding of learning lead to a number of conclusions about how to apply equitable grading. The conclusions typically stem from several principles or elements of a principle. Feldman [8] provides examples of applying the principles to particular situations, each somewhat different, and obviously not tailored for computer science in particular.

Below, we identify a number of suggestions that we have begun using in our courses, and which we believe embody equitable grading. We believe these practices can be combined and used effectively to implement equitable grading in other computer science courses. As noted above, the goal of instruction in computer science is that students develop or possess desired capability. Grading activity should focus on this goal.

3.1 Separate Learning and Assessment Activities

The purpose of homework is to allow students to learn. We tell students that learning is a process and they need to work hard and learn from their mistakes. However, our advice seems hollow when homework is included in grades and we are penalizing students for making mistakes we tell them are a natural part of the learning process. Additionally, typical practices associated with grading homework, such as late penalties, zeroes for non-submission, zeroes for using the work of others, and some aspects of grading group work do *not* measure academic capability of individual students, who ultimately receive the final grade.

Therefore, we suggest separating the *learning* activities from the *assessment* activities. Homework assignments provide students a great opportunity to practice what we've shown them, but a really poor opportunity for us to score what they've learned. We suggest providing students with a large set of ungraded homework problems that they can do to learn and then administering an assessment used to determine the extent of student learning, knowledge, or capability. The learning activity is essentially the homework but without grading. For assessments, we have had positive experiences using short, typically in-class, competency demonstrations. Competency demos could be proctored quizzes or short tests, either pencil-and-paper or on-line. A relatively brief oral examination could also be used. The main point is to make it clear to students that these activities in particular are being used to demonstrate competency in one or more of the learning outcomes for the course. That way, student grades reflect capability *after* the learning. Grading the competency demonstrations is much less onerous than grading homework. This does, of course, require additional effort on the part of the instructor to design good assessments based on their course outcomes, a task that we continue to refine every semester. We are energized, however, by the belief that these assessments will help create a more accurate and equitable grading system.

One might ask, "If homework is not graded, how do students receive feedback?" Our response is that alternatives must be developed. Computer science in particular has a plethora of options available to allow feedback without over-burdening the instructor with excessive work. For feedback on correctness of programs, there are a number of both free and for-pay "autograders" available which can evaluate student programs on things like functional correctness, style, and other metrics like complexity. Alternatively, the instructor could provide answer keys and students could check their own work (since the work is not being graded, providing answer keys should not be an issue).

For quality-related elements of homework that go beyond functional correctness, there are also a number of free and paid tools to facilitate code walkthroughs. Students can work in small groups or as a whole class to complete this activity. Not only does this provide feedback, it also helps familiarize students with a practice that they will almost surely need to participate in if seeking a job developing software. An additional mechanism for providing feedback for any type of student work is peer-review of student developed material. In his ICER keynote, Nicol [13] indicated that peer-review is not only good for those whose material is being reviewed but also for those doing the reviewing. His research [12] suggests:

- students should receive guidance that they understand and can buy into
- receiving reviews/feedback is useful for the improvement of the products they produced

- providing reviews/feedback engages reflection and critical thinking about both the review criteria and the production process and artifact

Additionally, the peer-review activity was generally well received by students under the given circumstances: review criteria provided, two peer reviews were assigned, a review of their own work was expected, and an opportunity to revise their work was provided.

It is worth noting here that we are not saying that we should not monitor a student's progress as they complete the learning activities (homework) – in fact, we feel it can be quite helpful for an instructor to track progress to identify struggling students and topics that might need further clarification. The techniques listed above make tracking student progress easier while still eliminating the need for painstakingly scoring every learning activity a student completes. The result is that we can spend more time helping students learn, not totalling points on a page.

3.2 Allow Assessment Re-Takes

We know that learning is highly individual: each student learns at their own pace in their own way. Furthermore, we know that mastery comes from repeated practices where we perform some action, collect feedback on our performance, use this feedback to improve, and repeat the action. Therefore, students should be given a place to perform these learning cycles to reach mastery without being penalized for the natural and expected mistakes they will make, i.e., provided opportunities for do-overs. Feldman [8], suggests that later assessments addressing some earlier learning goal be given more weight than an earlier assessment of the same learning as an alternative to re-takes. As noted above, however, we suggest using competency demos *plus* the re-take capability.

It is worth noting that this practice need not mean the student retakes the exact same competency demo they attempted before. In some cases, it is appropriate to provide feedback and allow the student to revise and resubmit (for instance, if they are working on a unique project of their own choosing). We have also had success in letting students make another attempt at demonstrating a particular outcome by completing a slightly different activity. For instance, a student may be asked to complete another quiz whose questions are of the same type but with different specifics. We have found it helpful to require a short period of time between repeats (for instance, a minimum of two or three days) so that students can have time to think about their understanding and learn more, and also to keep our workload manageable. For some courses, this does mean the instructor has to create multiple versions of some competency demos. However, we have found it fairly straightforward (although not necessarily "easy") to do so as the competency demos are now tied directly to student outcomes. We feel it is worth the extra instructor time up-front in exchange for improved student learning in the future.

It may seem that by allowing students unlimited re-takes, the instructor will be overwhelmed by them later in the semester and their performance and mental health will suffer. However, we have found competency demonstrations are much easier to grade than traditional homework, meaning re-takes can still be assessed quickly. Furthermore, not every student will necessarily re-take every competency demo until they earn an A. With our simpler grading system, students are aware of what their current grade is and what their final grade will likely be, and they are better equipped to make decisions like skipping particular activities based on their individual priorities. For these reasons, we have not found it problematic to provide re-takes on assessments.

Competency demos and re-takes are not the only way to accomplish favoring later assessments when grading. Comprehensive in-term and final exams could accomplish this goal also (though it has not been our approach).

3.3 Eliminate Late Penalties

We believe it is more equitable to give students the flexibility to master a particular objective at their own pace, though it may be later than we expect or later than some other students. There are many factors outside of a student's control that might affect their ability to turn in something at a particular deadline. For instance, students might have a very limited time to work outside of school hours, or a significant life event might cause students to fall behind. Since grades are meant to reflect student capabilities *at the end of the semester*, it is not necessary to penalize students for missing earlier deadlines. Therefore, we suggest eliminating penalties for late work.

In our discussions with other faculty, we have had several understandable concerns raised regarding our "no late penalty" policy. For one, there is concern that students might put off all assignments until the very end and then fail to complete them all due to a shortage of time. Note, however, that we are not suggesting one eliminate deadlines, but rather just eliminate the penalty for missing them. Our experience has been that most students do quite well in keeping up with the deadlines we provide, even if they know there is no penalty for late submissions. And, having assessments shortly after the due date will certainly encourage students to get the work done on time. As mentioned previously, we can also monitor students' progress and reach out to those who have fallen behind in competency demonstrations. Unlike the traditional system we used in the past, however, students who fall behind can still succeed in the course.

3.4 Focus on Outcomes for the Individual

When designing learning and assessment activities, we should focus only on the learning outcomes for the course. By enforcing this "outcome traceability", we can make sure our final grades are an accurate representation of each student's abilities. This means several items in our homework and exams that were unrelated to the overall course outcomes were removed and we adjusted our assessments appropriately. For every activity we did in class or had students do outside of class, we asked ourselves how this was connected to the individual outcomes for the course, and revised the activity until we could answer the question satisfactorily. Not only is this a great course design practice, but it also ensures the accuracy of student grades.

With the focus exclusively on individual outcomes, several other grading practices that we had commonly used were eliminated. For instance, we eliminated extra credit activities. We had originally introduced extra credit activities to allow struggling students a chance to complete additional work and raise their scores. To

ensure accurate and fair grading for struggling students, it makes more sense to have them continue working towards demonstrating capability in a course outcome. Another practice we eliminated was assigning points to students based upon participation. While it may be true that students learn better through engagement, participation is not a measurement of student capability, is subject to inaccurate recording, and penalizes students who are more reticent or whose culture discourages such behavior. Similarly, we stopped assigning a single score for group projects: again, while working in a group may be helpful for students, the final group product by itself is not an accurate reflection of what each *individual* is capable of.

It is important to note that focusing on the outcomes for an individual does not mean we cannot assess a student's ability to, say, work as part of a team. For classes where the ability to work as part of a team is included as a learning outcome, then it is wholly appropriate (and in fact necessary) to assess this. However, we should make sure that our assessment is in fact a measurement of what the *individual* can do in a group, not the final product of the group as a whole. We have had success in having the students complete an essay on what makes a good team and how their experiences have helped them see this, and combining this with self and team evaluations for determining competency in these kinds of outcomes.

3.5 Shrink the Scale

As mentioned earlier, we've noticed a few problems with the traditional grading scale we had used in the past. First, the range for acceptable or excellent work was much smaller than the range for failing work. For example, we had 11 ways to denote top-level work (assigning an A for scores between 90 and 100) and 60 ways to denote failing work (assigning an F for scores between 0 and 59). This meant that the penalty for a bad score could be extremely damaging to a student's overall score. For instance, consider a student who submits three assessments and earns an A on each, but fails to submit a fourth and is given a zero. With the traditional grading scale (0-100), this student's average would be a C: the same score a student might receive if they had *no* exceptional work for the entire semester. Using larger point scales also increased the time it took for us to grade, and opened up more student disputes: was this student error worth 5 points out of 100, or 3? If we're assigning grades in the same way we compute our taxes, then keeping track of these minor details makes sense. However, if we want our grades to accurately reflect what a student knows at the end of the course, there is no need for such granularity.

Our recommendation, then, is to "shrink the scale" and assign only a few discrete steps for scores. Feldman [8] suggests several practices such as using a scale from 0 to 4 like the typical GPA calculation. We have had success going even further and using a simple pass/fail grading scale and requiring all passes for an A, perhaps 1 fail for a B, etc. Since students will have a chance to improve and attempt the assessment again, the need for various measures of "almost there" is removed. There are of course similar alternatives if one is not comfortable with the binary grading scale. A colleague has implemented a three point scale (0, 1, and 2) to distinguish between unsatisfactory work, acceptable work, and exceptional work. All of these approaches seem equitable to us, as they all move the focus away from the totalling of many points and more towards simpler metrics of competency.

4 HOW EQUITABLE GRADING REVOLUTIONIZES OUR WORK

We began our paper with a discussion about what our ideal grading practice would do. In this section, we discuss how implementing the equitable grading practices above has (and continues to) revolutionize our instruction and realize our ideal grading practice.

4.1 Accurate Assessment

Focusing on what each student is able to do at the *end* of the semester and not the path the student took to get there results in their final grade being a much more accurate representation of their individual capabilities. Students' grades are determined strictly by assessments which we've intended to measure a particular set of outcomes for the course, and nothing else. A passing grade cannot be achieved by having exceptional group members or by collecting small numbers of points here and there, but rather must be earned by demonstrating capability in a minimal set of outcomes. Similarly, a poor grade cannot be received by taking a little longer than others to learn something, or by not "participating" enough in class. With the new paradigm, if we know a student's grade for a course and we have the outcomes for the course, we can build a much more accurate picture about what they are able to do than in our previous points-based system.

4.2 Equitable Assessment

Of course, an accurate assessment is also an equitable assessment. Students who already are competent in a particular outcome need not spend time with learning activities. Students who are unfamiliar with a particular outcome can have multiple opportunities to learn without penalty, even if doing so requires additional time. Students are also free to learn in different ways: a student can choose to work closely with a classmate to complete some learning activities if they wish, or work alone and review the solutions to learning activities. Assessments can be repeated if done unsatisfactorily, giving students a low-pressure environment to demonstrate what they can do. The new grading paradigm tries to ensure every student has a path to success, regardless of their prior knowledge and experiences, and regardless of factors outside of their control. Life events no longer guarantee a student cannot do well in a class as they have the flexibility to work hard and learn the concepts without penalty later in the semester.

4.3 Plagiarism Concerns Eliminated

Our concerns regarding plagiarism have been virtually eliminated by adopting this new grading paradigm. As learning activities are ungraded and simply meant to prepare a student for the competency demonstration, plagiarizing to solve a homework activity now literally at worst only hurts the offending student. As mentioned above, some students might actually find reviewing known answers as an effective technique for learning, and they are not punished for doing so. We have found it much easier to ensure students complete their own work on competency demonstrations,

as these are typically done in-class such as through a quiz or exam. Even for competency demonstrations that are done outside of class, with no deadlines for late work, students are much less likely to encounter the situation where the only way they can complete an activity is to cheat. Our experience has been students do not begin a course intending to cheat their way through it, and allowing them flexibility in deadlines and retakes virtually eliminates the factors we felt push them towards academic dishonesty in the first place.

4.4 Elimination of Grading Disputes

In the time we have been using a "points-less" approach, we have not had a single student contest our final decision on the score for a competency demo. Most grading disputes used to be over a few percentage points here and there, a distinction that is not able to be made when scoring with a much coarser scale like pass/fail or 0 through 4. Furthermore, students have the opportunity to repeat the assessment and correct any mistakes we feel they made regardless of why they made them, so grading disputes arising from misunderstandings of the assignment are minimized as well.

4.5 Increased Student Motivation

We have found students seem to adopt more of a growth mindset and are more motivated when we have implemented the practices of this new grading paradigm. We believe that students are much more receptive to the message that learning is a process that requires effort and hard work more than innate ability when our grading practices allow them to participate in this process. Students are able to see more of a connection between the work they do and the outcomes they achieve. It is worth noting as well that we have found advanced students are also happier than before, particularly because they no longer feel they are wasting time completing "busy work" and can now bypass "easy" activities.

Our personal observations are in agreement with a large body on grading and motivation. There has been increased attention given to promoting a growth mindset with students (for instance, consider the work of Murphy and Thomas [11]), which has numerous benefits, including helping retain women in computer science [7] and improving performance for introductory programming courses [5]. Similarly, intrinsic motivation has been found to be strongly correlated with student performance in programming courses [2]. However, multiple studies suggest that extrinsic rewards like grades actually *decrease* intrinsic student motivation, as summarized by Deci, Koestner, and Ryan's survey of 128 studies on motivation [6]. We believe the equitable grading practices we enumerate here will help improve motivation and develop a growth mindset.

4.6 Improved Course Design

Re-thinking how we grade has also led us to re-think our course design for the better. We have found the focus on learning outcomes instead of content is a natural lead-in to the *backwards design* approach described by Wiggins and McTighe [15] (a newer edition exists). There is now transparency in how assessments align with course outcomes, which is becoming increasingly important for things like accreditation in higher education.

Designing accurate assessments to measure student outcomes has also helped us design better activities to use during class time. Dry lectures have been replaced with activities that better correlate with the student outcomes.Anecdotally, students have seemed more energized in class, and have been more willing to participate in our in-class activities.

4.7 Improved Time Utilization

We have found our grading system to be much more efficient than the traditional system in two ways. First, the proportion of time we spend giving useful feedback now far outweighs the time we spend determining the precise "value" of work products. As experts in the field, we can often quickly determine if a particular product meets our criteria. If it does, there is little corrective commentary needed, and our time is instead spent providing feedback and suggestions for improvement to students who actually need (and welcome) it. Not only was this a better use of our time, but we also found this task to be much more pleasant than the more punitive tone we took with traditional grading systems. Second, there is now no need for complex formulas to determine final grades. While students did have a few questions initially, the simplicity of the grading system seemed to reduce the end-of-semester grade inquiries. We also found it refreshing to free up some of the accounting work that was required from traditional grading systems.

4.8 Focus on Learning

Last but not least, our new way of thinking about grading has left us both energized and able to spend more time thinking about how we can continue to improve student learning. We believe we spend less time on activities that don't improve learning (like pointlessly assigning points) and more time working on activities that help students achieve the objectives of the course. The feelings of grading burnout we used to have as we played "points police" are no longer present, and we find ourselves more excited about assessment than we were before. The value we have found in improved instructor morale cannot be overstated.

5 CONCLUDING THOUGHTS

It is important for us to note that we are not the first ones to notice the problems with traditional grading in computer science and push towards a better form of assessment. Some folks even call for the elimination of grading, including one computer science educator [14]. We do not recommend doing away with grading. However, the recurring discussions of auto-grading, plagiarism, etc., on the SIGCSE discussion list strengthens our call for something new, perhaps equitable grading. We feel the grading for equity approach advocated by Feldman [8] enhances both the teaching and the learning experience. We encourage you to read the book.

While our goal here is to inform and persuade, we are at least hopeful that this can lead to a new discussion of learning and grading. As [9] suggests, we would like to see

> "...further reflection and discussion of approaches to grading programs. As a community we should be able to articulate the reasons for our choices for the benefit of our students and to clarify our own thinking. Furthermore, we should discuss those reasons with our colleagues, not competitively, but to learn from each other."

The move towards more equitable grading has revolutionized our teaching for the better, and we believe it will also do so for the broader computer science education community.

REFERENCES

[1] Anonymous Authors. [n. d.]. Anonymous Title. ([n. d.]).

[2] Susan Bergin and Ronan Reilly. 2005. The influence of motivation and comfort-level on learning to program. In *Proceedings of the 17th Workshop of the Psychology of Programming Interest Group, PPIG 05. University of Sussex, Brighton, UK, June 2005*. Psychology of Programming Interest Group, 293–304. http://mural.maynoothuniversity.ie/8685/

[3] Chris Mayfield Aman Yadav Tammy Pirmann Clif Kussmaul, Helen Hu. 2018. CSPOGIL (Process Oriented Guided Inquiry Learning in Computer Science). (2018). https://cspogil.org/Home

[4] Michelle Craig, Diane Horton, Daniel Zingaro, and Danny Heap. 2016. Introducing and Evaluating Exam Wrappers in CS2. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education (SIGCSE '16)*. ACM, New York, NY, USA, 285–290. https://doi.org/10.1145/2839509.2844561

[5] Quintin Cutts, Emily Cutts, Stephen Draper, Patrick O'Donnell, and Peter Saffrey. 2010. Manipulating Mindset to Positively Influence Introductory Programming Performance. In *Proceedings of the 41st ACM Technical Symposium on Computer Science Education (SIGCSE '10)*. ACM, New York, NY, USA, 431–435. https://doi.org/10.1145/1734263.1734409

[6] Edward L. Deci, Richard Koestner, and Richard M. Ryan. 1999. A Meta-Analytic Review of Experiments Examining the Effects Of Extrinsic Rewards on Intrinsic Motivation. *Psychological Bulletin* 125 (1999), 627–668.

[7] Wendy M. DuBow, Beth A. Quinn, Gloria Childress Townsend, Rosario Robinson, and Valerie Barr. 2016. Efforts to Make Computer Science More Inclusive of Women. *ACM Inroads* 7, 4 (Nov. 2016), 74–80. https://doi.org/10.1145/2998500

[8] J. Feldman. 2018. *Grading for Equity: What It Is, Why It Matters, and How It Can Transform Schools and Classrooms*. SAGE Publications. https://books.google.com/books?id=ORhoDwAAQBAJ

[9] Sue Fitzgerald, Brian Hanks, Raymond Lister, Renee McCauley, and Laurie Murphy. 2013. What Are We Thinking when We Grade Programs?. In *Proceeding of the 44th ACM Technical Symposium on Computer Science Education (SIGCSE '13)*. ACM, New York, NY, USA, 471–476. https://doi.org/10.1145/2445196.2445339

[10] Susan Horwitz, Susan H. Rodger, Maureen Biggers, David Binkley, C. Kolin Frantz, Dawn Gundermann, Susanne Hambrusch, Steven Huss-Lederman, Ethan Munson, Barbara Ryder, and Monica Sweat. 2009. Using Peer-led Team Learning to Increase Participation and Success of Under-represented Groups in Introductory Computer Science. In *Proceedings of the 40th ACM Technical Symposium on Computer Science Education (SIGCSE '09)*. ACM, New York, NY, USA, 163–167. https://doi.org/10.1145/1508865.1508925

[11] Laurie Murphy and Lynda Thomas. 2008. Dangers of a Fixed Mindset: Implications of Self-theories Research for Computer Science Education. In *Proceedings of the 13th Annual Conference on Innovation and Technology in Computer Science Education (ITiCSE '08)*. ACM, New York, NY, USA, 271–275. https://doi.org/10.1145/1384271.1384344

[12] David Nichol, Avril Thompson, and Caroline Breslin. 2014. Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education* 39, 1 (2014), 102–122. https://doi.org/10.1080/02602938.2013.795518

[13] David J. Nicol. 2014. Unlocking Learners' Evaluative Skills: A Peer Review Perspective. In *Proceedings of the Tenth Annual Conference on International Computing Education Research (ICER '14)*. ACM, New York, NY, USA, 1–2. https://doi.org/10.1145/2632320.2632359

[14] Beckie Supiano. 2019. Grades Can Undermine Learning. What Should Professors Use Instead? *Chronicle of Higher Education* (2019), A12–A15.

[15] Grant Wiggins and Jay McTighe. 1998. *Understanding by Design*. Association for Supervision and Curriculum Development, Alexandria, Virginia, USA.