Transforming Grading Practices in the Computing Education Community

Adrienne Decker
Engineering Education
University at Buffalo
Buffalo, NY, USA
adrienne@buffalo.edu

Stephen H. Edwards
Computer Science
Virginia Tech
Blacksburg, VA, USA
edwards@cs.vt.edu

Brian M. McSkimming
Engineering Pathways
University of Oklahoma
Norman, OK, USA
mcskimming@ou.edu

Bob Edmison
Computer Science
Virginia Tech
Blacksburg, VA, USA
kedmison@vt.edu

Audrey Rorrer
Computer Science
UNC Charlotte
Charlotte, NC, USA
audrey.rorrer@charlotte.edu

Manuel A. Pérez-Quiñones
Software and Information Systems
UNC Charlotte
Charlotte, NC, USA
perez.quinones@uncc.edu

ABSTRACT

It is often the case that computer science classrooms use traditional grading practices where points are allocated to assignments, mistakes result in point deductions, and assignment scores are combined using some form of weighted averaging to determine grades. Unfortunately, traditional grading practices have been shown to reduce achievement, discourage students, and suppress effort to such an extent that some common elements of traditional grading practices have been termed toxic. Using grades to reward or punish student behavior does not encourage learning and instead increases anxiety and stress. These toxic elements are present throughout computing education and computer science classrooms in the form of late penalties, lack of credit for code that doesn't compile or pass certain unit tests, among others. These types of metrics, that evaluate behavior are often influenced by implicit bias, factors outside of the classrooms (e.g., part-time employment), and family life situations (e.g., students who are caregivers). Often, students in these situations are disproportionately from low-socioeconomic backgrounds and predominantly students of color. Through this paper, we will present a case for adoption of equitable grading practices and a call for additional support in classroom and teaching technologies as well as support from administrations both at the department and university level. By adopting a community of practice approach, we argue that we can support new faculty making these changes, which would be more equitable and inclusive. Further, these practices have been shown to better support student learning and can help increase student learning gains and retention.

SIGCSE 2024, March 20–23, 2024, Portland, OR, USA.

ACM ISBN 979-8-4007-0423-9/24/03.
https://doi.org/10.1145/3626252.3630953

CCS CONCEPTS

• Social and professional topics~Professional topics~Computing education•Social and professional topics~Professional topics~Computing education~Computing education programs~Computer science education

KEYWORDS

Equitable grading, grading for equity, grading practices

ACM Reference format:

Adrienne Decker, Stephen H. Edwards, Brian M. McSkimming, Bob Edmison, Audrey Rorrer, Manuel A. Pérez-Quiñones. 2024. Transforming Grading Practices in the Computing Education Community. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V.1 (SIGCSE 2024). ACM, New York, NY, USA, 7 pages.* https://doi.org/10.1145/3626252.3630953

1 INTRODUCTION

Douglas Reeves wrote: "If you wanted to make just one change that would immediately reduce student failure rates, then the most effective place to start would be challenging prevailing grading practices" [39]. We believe that we must challenge prevailing grading practices to transform the way the community of CS educators and learners perform, understand, and use assignment grading.

Most CS classrooms use traditional grading practices where points are allocated to assignments, mistakes result in point deductions, and assignment scores are combined using some form of weighted averaging to determine grades. "Traditional grading practices have been used for over one hundred years, and to date, there have been no meaningful research reports to support it" [43]. Unfortunately, traditional practices *reduce achievement, discourage students, and suppress effort* [20, 43]. Some common elements of traditional grading practices are so negative that they have even been termed **toxic** [39] including: using zeros for missing work; averaging scores throughout the academic term; and the use of heavily weighted high-stakes tests or project assignments where one assignment's score can make the difference between achieving an average grade or needing to repeat a course. Using grades punitively is an ineffective teaching tool because it increases anxiety and stress, yet these toxic grading practices are virtually ubiquitous throughout higher education.

Unfortunately, "most teachers have not received adequate training in reliable and valid assessment methods in their teacher preparation and often default to the way they saw their teachers grade when they were in school" [12]. According to Feldman [18], traditional grading includes a component that evaluates student's behaviors, often including timeliness, effort, and other behavioral measures. These metrics, which are not based on content knowledge or learning outcomes, are often influenced by implicit biases, personal factors outside of the classroom (e.g., part-time employment), and by family life situations (e.g., students who are caregivers). Students in these situations are disproportionately students from marginalized communities (e.g., low-socioeconomic background, first generation, and by proxy, students of color). "Unfortunately, many educators have fallen into the trap of believing that punitive grading should be the chief consequence for poor decisions and negative behaviors. These teachers continue to argue that grading as punishment works, despite over 100 years of overwhelming research that suggests it does not" [12].

In order to successfully transform these practices, we need to engage the community of educators in this space. Computing education faculty need to be guided towards these changes with support of a community and with tools to help facilitate this change. This transformation will need to incorporate a variety of equitable grading practices, some for grading by hand, others supported within various automated grading systems. This transformation has the potential to provide the following benefits:

- To guide and engage students as they develop and improve their development of self-regulatory learning practices as they set goals for final grades, plan how they will achieve those goals, monitor how they are progressing, and evaluate how to update or modify their plans as they proceed.
- To shift students' focus on learning outcomes instead of points by framing grading in terms of demonstrations of specific learning objectives in each assignment. When students do not complete an assignment successfully, they are explicitly informed of which concepts were not demonstrated, and therefore exactly which learning objectives they need to improve upon.
- To increase academic achievement in class [20].
- To create a learning atmosphere that facilitates students' individual ownership of achievement of the goals they set for themselves, thereby diminishing the contentious relationship with professors about grading ("You took points away from me").
- To reduce student grading stress and frustration.
- To provide students with clear expectations while maintaining high standards of performance.
- To increase support for students that are traditionally disadvantaged by the higher educational system by grading based on content and not behavior (e.g., late penalties, missed deadlines).

2 BACKGROUND

In this section, we will present additional information about alternative grading practices, self-regulation, communities of practice, and tool support, each of which is a key aspect of realizing a vision of transforming grading practices in our discipline.

2.1 Grading

Feldman [18], Rapaport [37], and Nilson [33] propose alternative grading practices that have influenced our thinking of how to transform grading in computing education. All three authors advocate for a reduced grading scale. Feldman [18] suggests that a 0-100 scale, with 0-59 being failure, tilts the scale towards failure. Furthermore, if multiple grades are averaged to gain a final score, a 0 in an assignment has an over-weighted punitive effect on the student's score. Feldman advocates we use "minimum grading", meaning that the minimum score that a student could obtain for a 0 (e.g., no submission) is a 50. For a deeper discussion, see Feldman's Chapter 7, "Practices that are Mathematically Accurate" [18]. Nilson [33] suggests a pass/fail (Satisfactory/Unsatisfactory) grading scale with opportunities for students to resubmit work with a fail grade, thereby providing additional opportunities for practice. Rapaport's triage grading [37] provides a third option, such that any item to be graded (whether it be an assignment, a question/multi-part question on an assignment, specific learning objectives within an assignment, or specific items in a rubric) gets one of three grades:

- **Full** credit if it is clearly/substantially correct. (i.e., 3 points)
- **Minimal** credit if it is clearly/substantially incorrect. (i.e., 1 point)
- **Emerging** credit if the item is neither of the above. (i.e., 2 points)
- **Zero** credit if the item does not exist. (i.e., 0 points)

Nilson [33] recognizes the importance of tying the grading scale to the learning outcomes of the assignment. Typically, one does not need perfection to demonstrate evidence of meeting the intended learning outcomes of an assignment, however this detail tends to be missing in most grading practices in higher education.

Partial credit grading within a 0-100 scale has several disadvantages as recognized by Feldman, Nilson, Rapaport and others [39, 22]. First, students feel they can argue for a few more points after the fact without doing additional work [33, 37]. Rapaport notes that within triage grading, if the item to be graded is not completely correct yet not substantially incorrect, then there is only one other option.

Second, averaging scores across assignments using arbitrary weighting schemes for different types of assignments is also problematic [33]. For example, assignments done outside of the classroom might not be done independently and thus be weighted lower than in-class work. Assignments done in the classroom (e.g., exams) may be viewed as having more authenticity. Thus, the final weight given to different categories of assignments (e.g., lab, homework, programming, exams) is determined not because of any learning objective measure but because of faculty trust/mistrust of our own instruments reflecting our learning outcomes.

Further, partial credit is not ideal in that it is often seen as a "single deadline," thus increasing the stakes of the assignment (e.g., must submit it by the deadline or lose points as a late penalty). Students may see partial credit as a game whose goal is to maximize points without changing effort (e.g., bargaining to get more points in an assignment) [33]. Ultimately, students end up seeing the score as an externally influenced metric (e.g., "you gave me a C") rather than a reflection of the learning demonstrated in an assignment (e.g., external vs. internal locus of control [13]). Feldman suggests that when grades are seen as extrinsic motivators, they might not encourage a growth mindset [13].

The opportunity for students to resubmit work that sufficiently demonstrates comprehension or understanding of a concept within a grading system is key to the desired transformation. Resubmission of an assignment lowers the stakes of the assignment, provides students with additional opportunities to practice and complete the assignment, and more closely matches the learning outcomes of the assignment or course [33]. This approach of pass/fail and multiple submissions encourages high standards and low stakes, an approach that Bowen [5] argues is best for student's learning.

Perhaps most important, all three authors recognize the necessity of useful and meaningful feedback to students. The triage grading methodology proposed by Rapaport suggests that submissions awarded emerging credit can then be further assessed and evaluated as the source of feedback both to the instructor (areas needing additional attention, perhaps in a classroom setting) and the student (potentially promoting more useful self-reflection and better future self-regulation). Triage grading provides the student implicit feedback which is directly contained within the scoring. For instance, a 3 says 'you got it right (for all practical purposes)', a 2 says 'almost, but not quite', a 1 says 'nope', and a 0 says 'you did not even try'; weighting of the various items can indicate relative importance [37, 38].

2.2 Self-Regulated Learning and Motivation

The ability to self-regulate provides an advantage throughout our development as a society, a species, and an individual. Specifically, "Self-regulation refers to self-generated thoughts, feelings, and actions that are planned and cyclically adapted to the attainment of personal goals." [48] As such, some [47] have argued that there are no un-self-regulated persons nor could there be a complete absence of self-regulation; rather, there exists varying quality and quantity of one's self-regulatory processes [48].

According to Zimmerman [41, 48], self-regulatory processes are contained within three cyclic phases: forethought, performance/volitional control, and self-reflection as shown in Figure 1. The forethought phase of self-regulation refers to processes that influence the marshalling of efforts towards action (or in-action) and work toward activating the necessary personal resources for action. Performance control suggests processes

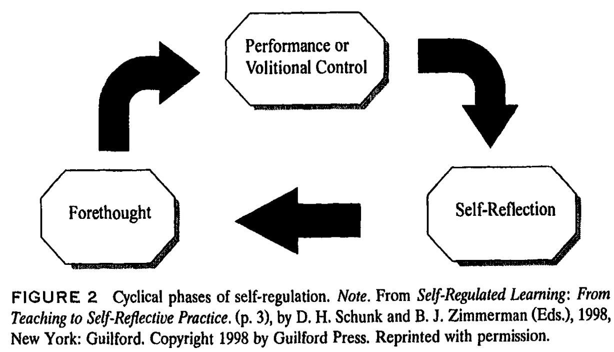

Figure 1: Phases of Self-Regulation

occurring during effort/action, and self-reflection involves processes occurring after action/effort and serve to influence one's response to the experience. These self-reflections are important for processes within the forethought phase, thus resulting in a full self-regulatory cycle.

While self-regulation of cognition and behavior are important components within learning, they are often insufficient on their own to describe and promote overall achievement [34]. Students need to be motivated to use the associated cognitive strategies present within self-regulation.

2.3 Communities of Practice

The identification of conceptual communities of practice (CoP) and their associated framework was the result of work by Lave & Wenger [26] considering the learning resultant from belonging to a group as one transitions from peripheral participation within specialized communities (newcomers) into acceptance and active participation within those communities (old-timers). The 30 years since their seminal work has witnessed a metamorphosis in both the communities of practice framework [44, 45, 46] but also in the educational research performed embracing these ideals [1, 4, 7, 29] and the associated situated learning frameworks [5, 21, 28, 46].

A current definition of CoP suggests that these communities are formed by individuals pursuing collective learning and experiences within a specific domain/endeavor, whether it be a tribe learning to survive, a group of engineers working on similar problems, or a gathering of first-time managers helping each other cope. Specifically, "*Communities of practice are groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly.*" [44]

The three necessary elements of a CoP are:

1. **Domain**—Membership within a CoP implies a commitment to a shared domain of interest. While this may not necessitate an expertise recognized outside of the community itself, there is a valued collective competence and learning capability within the CoP with regard to the domain regardless of the outside perceived value.
2. **Community**—In pursuing interest in domain, members engage in joint activities and discussions, help each other, and share information. They build relationships that enable them to learn from each other and they care about their standing with each other. While the members may not necessarily work together on a daily basis, the interactions they have when coming together are essential to their personal identity and membership within the CoP.
3. **Practice**—A CoP is not merely a community of shared interests, e.g., people who like similar music, games, etc. Members of the community are practitioners developing a shared repertoire of resources including experiences, stories, ways of addressing recurring problems: a shared practice.

The combination of these three elements constitutes a CoP. Additionally, Wenger [45] describes three dimensions of practice which facilitate an effective CoP:

1. **Mutual Engagement**—Through participation in the CoP, members establish norms and build collaborative relationships which tie the community together. This mutual engagement results in a diverse group of people working together to create and negotiate meaning, creating the social fabric of the CoP.
2. **A Joint Enterprise**—Interactions within members of the CoP result in the creation of a shared understanding of the interconnectedness of each member thus resulting in the development of mutual accountability, shared goals, and a cooperative identity.
3. **A Shared Repertoire**—As part of its practice, the community produces a set of communal resources, which is termed their shared repertoire to be used in the pursuit of their joint enterprise which can include both literal and symbolic meanings.

As with any community, cultivation of a successful CoP can be a challenging endeavor. CoPs are dynamic learning communities which shape not only itself but also the individual identities of the COP's members throughout their engagement. Some of the suggestions made by Wenger et. al [46] to build a successful CoP include:

- **Focus on the value of the community**—CoPs should create opportunities for participants to explicitly discuss the value and productivity of their participation in the group.
- **Combine familiarity and excitement**—CoPs should offer the expected learning opportunities as part of their structure, and opportunities for members to shape their learning experience together by brainstorming and examining the conventional and radical wisdom related to their topic.
- **Find and nurture a regular rhythm for the community**—CoP's should coordinate a thriving cycle of activities and events that allow for the members to regularly meet, reflect, and evolve. The rhythm, or pace, should maintain an anticipated level of engagement to sustain the vibrancy of the community, yet not be so fast paced that it becomes unwieldy and overwhelming in its intensity.

2.4 Automated Tools

The three most important groups of classroom tools used by CS educators are automated grading systems, small exercise practice systems, and learning management systems. As part of adopting equitable grading practices, we need to address all three groups of tools to ensure that educators can use these tools as part of regular classroom activities.

Automated grading systems are well-known among CS educators and are seeing increasing use as educators turn to them to cope with increasing numbers of students. Both open-source solutions, such as Web-CAT [14, 15] and INGInious [23], and commercial solutions, including Gradescope [19] and CodeGrade [8], are commonly used in many classrooms. Gradescope, for example, is used by over a thousand universities, while Web-CAT's primary server at Virginia Tech has processed nearly 5 million submissions from over 50 thousand users and is installed at many other institutions.

Similarly, classrooms are starting to see increased usage of homework practice systems for small coding exercises. CodeWorkout [9, 17] is one example, providing support for drill-and-practice questions in Java, Python, and C++. It allows instructors to give students questions that require writing individual methods or blocks of code, as well as questions involving partially complete code where a student fills in the blanks, questions providing complete but buggy code for the student to fix, or multiple-choice questions of any form. Such systems can be used to provide syntax practice exercises [16, 27] as well as homework on basic programming skills. Most such systems provide results using numeric scores.

Finally, many modern classrooms use some form of learning management system (LMS) to manage student access to course content and to grades. Canvas [24], Moodle [32], and Blackboard [3] are common examples.

3 CREATING A COMMUNITY OF PRACTICE

We believe that embracing equitable and inclusive practices will promote self-regulation of student learning and can easily be incorporated by instructors. Adoption of such practices will result in improved learning and retention throughout computer science. To implement the alternative grading and feedback protocols, we propose two parallel paths to success: (1) development and nurturing of a CoP focused on the alternative grading/feedback protocols and (2) significant expansion of common grading and practice software used throughout the computing education landscape.

Embracing the CoP framework through a series of professional development activities provides the necessary precursors and catalysts to form a strong and effective CoP. Effective CoPs need to be involved in mutual engagement, a joint enterprise, and development of a shared repertoire [45]. There are several examples of faculty who have adopted and promoted these types of practices. We believe that they can be effective leaders and mentors for other faculty who are interested in adopting these practices (a first cohort in the CoP). In turn, as new faculty are brought into the CoP, they can become mentors to additional cohorts of faculty.

Through extensive expansion and alterations to common grading and practice software, potential *external* roadblocks to and assumptions about adoption barriers are minimized. Through the supportive and learning activities of the CoP, potential *internal* roadblocks to and assumptions about adoption barriers are minimized. As such, the combined efforts from these two parallel paths minimizes resistance to adoption of the alternative grading/feedback protocols.

Finally, research on the performance gains from these alternate grading philosophies [10, 30, 20, 35, 36, 40, 42] suggest that instructor and student adoption of equitable grading practices may lead to improvements in student performance and self-regulation capabilities. This is a direct result of engaging students in developing self-regulated learning practices by focusing on learning outcomes and providing feedback to students based on outcomes, while simultaneously removing the toxic features of traditional points-based grading.

3.1 Development of Training Materials

To support such a community of practice, there is a need to develop training materials for engaging CS educators in the transformation. The goal of the training will be to model the equitable grading theories using explanations, examples, and samples for the faculty to ease adoption. In addition, it is important to provide information to the members of the CoP about how to incrementally adopt these approaches, perhaps over multiple semesters (i.e. series of small steps to take, in what order, what's involved, expected impact, expected time to implement, etc.).

Another key component of the training materials will be information about how to communicate these practices to the students. The students in the courses will need to be educated about the system and the affordances as much as the faculty who are adopting it. They need to not only understand what the changes are, but also how they need to possibly change their behaviors to take advantage of it.

Above and beyond their classroom, instructors need to feel confident to talk to their administration about this change of grading. They should be equipped with information, data, and talking points to help doubting administrators understand why these changes are good for students and good for learning.

3.2 Tools to Support Educators and Students

In terms of transforming grading practices, one of the most significant obstacles is the perceived effort or amount of work needed to use the new approach. For educators, minimizing the impact of transitioning to specification grading is critical for project success. To this end, it is necessary to ensure that the classroom tools used regularly by CS educators will support specification grading (or similar) with minimal effort, and that these tools will give learners appropriate feedback. We posit that augmenting current classroom systems that faculty use would enable this transition to be smoother for faculty.

The automated grading systems previously discussed provide extensible capabilities for processing work in virtually any programming language, and to customize the nature of the run-time testing and analysis performed. Web-CAT uses a plugin system to allow for extensible support for different programming languages, different approaches to software testing, different assessment strategies, and different feedback generation approaches. INGInious, Gradescope, and CodeGrade all use Docker [11] containers to provide extensible support for different programming languages and testing strategies. Of these systems, Web-CAT and CodeGrade provide greater flexibility in presenting custom or extensible feedback to students. For Web-CAT, this would involve creating a specification grading plugin that will generate feedback in the appropriate form. For CodeGrade, a custom docker image could be created that can be used to generate specification grading style feedback in HTML format for presentation to students.

Gradescope and INGINious both use more constrained APIs for instructor-provided Docker images where results presented back to students are communicated in the form of passed or failed test cases as well as numeric scores. To adapt such tools for specification grading results can be encoded using a choice of ordinal scales (such as 2-valued pass/fail, or 4-valued triage system [37]). This approach allows assignment feedback to be encoded using simple 0-3 scores that will fit within models constrained to be numeric, while still preserving the meaning of specification grading for students. These Docker images can provide a ready-made base for instructors using this approach to build their own classroom assignments with no more effort than they currently use for those tools, with specification grading features already handled in a clean and well-designed way for them to reuse.

Finally, LMS systems play a key role in how grades are presented in the class gradebook. Educators who use these systems are familiar with how scores are presented in a traditional grading model but will need specific LMS support when applying equitable grading practices. Canvas provides a mastery gradebook view [6, 25] to support mastery-based grading approaches, and Moodle supports this approach through competencies [31]. In addition, assignments can be configured to display assignment scores using custom-defined "grading scales" with custom-labeled levels, naturally supporting 2-valued or 4-valued grading levels to be displayed. However, while these features exist, many in our community are unaware of them, how they should be used, or where they may be limited.

To support educators and learners, we need the development of clear course gradebook templates and guidelines that can be closely integrated with the teacher materials. Educators also need clear guidance on how to use LMS-based test or quizzing support for equitable grading practices. Educators also need clear strategies for how to integrate any outside resources that only support traditional numeric grading into their LMS without creating conflicts. These are problems that educators would traditionally have to solve on their own, but if we lean into the CoP model, the CoP should work to provide our community with ready-made, working solutions as well as information on how to use them.

4 EQUITABLE GRADING FOR COMPUTING EDUCATION

Throughout this paper, we have argued that we should work to create a community of practice to enable faculty in computer science to adopt equitable grading practices in their courses.

The arguments for these transformative grading practices presented include:

1. A reduced grading scale eliminating traditional partial credit makes it easier to grade and eliminates negotiations with students about "I need one more point."
2. A more equitable grading scale by eliminating the 0-100 orientation to failure.
3. Direct tie the grading scale to the learning outcomes of the assignment.
4. Allow and encourage resubmissions of assignments thus lowering the stakes of the assignment and providing students with additional practice opportunities. Resubmissions also support students with behavioral challenges (e.g., part-time employment) outside of class.
5. Rapid, useful, and meaningful feedback promotes development of self-regulated learning practices.

It is also important to note that the CC2020 curriculum overview report [2] extols the virtues of competency-based grading indicating that it has many of the same qualities as described for specifications grading [29] and triage grading [37].

For this vision to be successful, we need to utilize practices from all the grading systems mentioned and others that are currently being developed in the community by faculty interested in better supporting their students in assessment and feedback. When thinking about faculty adoption, we must address upholding high academic standards, and the time commitment from the faculty (for both learning and ultimately using the approach) [29].

Overall, we that the vision of equitable grading for computer science classes is achievable and that we can create a community of practice around Feldman's three pillars: accurate, bias-resistant, and motivational grading [18], that have the following elements at its core:

1. *minimum grading* to support mathematically accurate scores.
2. *learning outcomes-based grading scales* so that quick feedback readily shows where more work is needed.
3. *smaller grading scales* to make grading faster, reduce negotiation about points, and possibly increase internal locus of control (reducing the "*you* gave me this grade").
4. *multiple submissions* to lower stakes of assignments, provide additional opportunities to learn, and allow students to match the outcomes of the assignment more closely by having more attempts at getting the right answer.
5. *removal of late penalties* to allow students to progress through the material at their own pace.
6. *automatic grading and testing support* to enable faculty to more easily grade assignments and provide feedback using these techniques.
7. *integration into existing LMS* to enable faculty to easily adopt these grading schemes where students normally see their course grades.
8. *resources for adoption* that enable faculty to understand the techniques, to educate themselves, their students, and their administration about the techniques.

We believe that the combination of creating a diverse community of people all focused on employing transformative and equitable grading practices will result in a more effective, more inclusive learning experience for students. At the same time, we believe instructors will benefit by learning through the experiences of colleagues and improved tooling support, allowing them to focus more on their students, and less on the mechanics of providing equitable grading.

If you are interested in joining us on this journey, please go to: https://cs-equitable-grading-practices.github.io/.

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. National Science Foundation under Grant Nos. 2235337, 2235643, and 2235644. This material is based in part based upon work done by the last author while serving at the National Science Foundation.

REFERENCES

[1] Akerson, V. L., Cullen, T. A., and Hanson, D. L. 2009. Fostering a community of practice through a professional development program to improve elementary teachers' views of nature of science and teaching practice. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching* 46, 10, 1090-1113. http://dx.doi.org/10.1002/tea.20343.

[2] Association for Computing Machinery CC2020 Task Force. 2020. Computing Curricula 2020: Paradigms for Global Computing Education. Retrieved 1 August 2023 from https://www.acm.org/binaries/content/assets/education/curricula-recommendations/cc2020.pdf

[3] Blackboard. 2021. Educational Technology Services | Blackboard | North America. Retrieved 1 August 2023 from https://www.blackboard.com/

[4] Blanton, M. L. and Stylianou, D. A. 2009. Interpreting a Community of Practice Perspective in Discipline-Specific Professional Development in Higher Education. *Innovative Higher Education* 34, 2 (2009/06/01), 79-92. http://dx.doi.org/10.1007/s10755-008-9094-8.

[5] Bowen, J. A. 2012. Teaching naked: How moving technology out of your college classroom will improve student learning. John Wiley & Sons. San Francisco, CA, USA.

[6] Canvas, P. 2020. Introduction to the Learning Mastery Gradebook. Retrieved 1 August 2023 from https://psu.instructure.com/courses/1881551/pages/introduction-to-the-learning-mastery-gradebook?module_item_id=22916901

[7] Chalmers, L. and Keown, P. 2006. Communities of practice and professional development. *International Journal of Lifelong Education* 25, 2, 139-156. http://dx.doi.org/10.1080/02601370500510793.

[8] CodeGrade. 2020. CodeGrade. Retrieved 1 August 2023 from https://www.codegrade.com/

[9] CodeWorkout. (2021). CodeWorkout. Retrieved 1 August 2023 from https://codeworkout.cs.vt.edu/

[10] Craig, T. A. 2011. Effects of standards-based report cards on student learning Ph.D. Dissertation. Northeastern University, Boston, MA, USA. https://www.proquest.com/docview/926432787

[11] Docker. 2021. Empowering App Development for Developers. Retrieved 1 August 2023 from https://www.docker.com/

[12] Dueck, M. 2014. *Grading smarter, not harder: Assessment strategies that motivate kids and help them learn.* ASCD, Alexandria, VA, USA.

[13] Dweck, C. S. 2007. *Mindset: The new psychology of success.* Random House, New York, NY, USA.

[14] Edwards, S. H. 2004. Using software testing to move students from trial-and-error to reflection-in-action. In *Proceedings of the 35th SIGCSE Technical Symposium on Computer Science Education (SIGCSE '04)* ACM, Norfolk, Virginia, USA, 26-30. http://dx.doi.org/10.1145/971300.971312

[15] Edwards, S. H. 2019. Web-CAT: the Web-based Center for Automated Testing - Web-CAT. Retrieved 1 August 2023 from http://web-cat.org/

[16] Edwards, J., Ditton, J., Trninic, D., Swanson, H., Sullivan, S., & Mano, C. 2020. Syntax exercises in CS1. *Proceedings of the 2020 ACM Conference on International Computing Education Research (ICER '20)*, Virtual Event, New Zealand. http://dx.doi.org/10.1145/3372782.3406259

[17] Edwards, S. H. and Murali, K. P. 2017. CodeWorkout: Short Programming Exercises with Built-in Data Collection. In *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE '17)* ACM, Bologna, Italy, 188-193. http://dx.doi.org/10.1145/3059009.3059055.

[18] Feldman, J., 2019. *Grading for equity: What it is, why it matters, and how it can transform schools and classrooms.* Corwin Press, Thousand Oaks, California, USA.

[19] Gradescope. 2020. Gradescope. Retrieved 1 August 2023 from https://www.gradescope.com/

[20] Great School Partnership. 2020. Research Supporting Proficiency-Based Learning: Grading + Reporting. Retrieved 1 August 2023 from https://www.greatschoolspartnership.org/proficiency-based-learning/research-evidence/research-supporting-ten-principles-grading-reporting/

[21] Gunawardena, C. N., Hermans, M. B., Sanchez, D., Richmond, C., Bohley, M., and Tuttle, R. 2009. A theoretical framework for building online communities of practice with social networking tools. *Educational Media International* 46, 1, 3-16.

[22] Guskey, T. R. 2000. Grading policies that work against standards … and how to fix them. *NASSP Bulletin*, 84, 620, 20–29. https://doi.org/10.1177/019263650008462003

[23] INGInious. 2021. UCL-INGI/INGInious. Retrieved 1 August 2023 from https://github.com/UCL-INGI/INGInious

[24] Instructure. 2021a. Higher Education LMS | Canvas for Colleges & Universities. Retrieved 1 August 2023 from https://www.instructure.com/product/higher-education/canvas-lms

[25] Instructure. 2021b. How do I use the Learning Mastery Gradebook to view outcome results in a course? Retrieved 1 August 2023 from https://community.canvaslms.com/t5/Instructor-Guide/How-do-I-use-the-Learning-Mastery-Gradebook-to-view-outcome/ta-p/775

[26] Lave, J. and Wenger, E. 1991. *Situated learning: Legitimate peripheral participation.* Cambridge university press, Cambridge, UK.

[27] Leinonen, A., Nygren, H., Pirttinen, N., Hellas, A., and Leinonen, J. 2019. Exploring the Applicability of Simple Syntax Writing Practice for Learning Programming. In *Proceedings of the Proceedings of the 50th ACM Technical Symposium on Computer Science Education (SIGCSE '19).* Minneapolis, MN, USA, ACM, 84–90. http://dx.doi.org/10.1145/3287324.3287378

[28] Lotter, C., Yow, J. A., & Peters, T. T. 2014. BUILDING A COMMUNITY OF PRACTICE AROUND INQUIRY INSTRUCTION THROUGH A PROFESSIONAL DEVELOPMENT PROGRAM. *International Journal of Science and Mathematics Education* 12, 1 (2014/02/01), 1-23. http://dx.doi.org/10.1007/s10763-012-9391-7.

[29] Mak, B. and Pun, S.-H. 2015. Cultivating a teacher community of practice for sustainable professional development: beyond planned efforts. *Teachers and Teaching* 21, 1 (2015/01/02), 4-21. http://dx.doi.org/10.1080/13540602.2014.928120.

[30] Mirsky, G. M. 2018. Effectiveness of specifications grading in teaching technical writing to computer science students. *J. Comput. Sci. Coll.* 34, 1 (October 2018), 104–110. https://dl.acm.org/doi/abs/10.5555/3280489.3280505

[31] Moodle. 2021a. Competencies - MoodleDocs. Retrieved 1 August 2023 from https://docs.moodle.org/311/en/Competencies

[32] Moodle. (2021b). Moodle - Open-source learning platform | Moodle.org. Retrieved 1 August 2023 from https://moodle.org/

[33] Nilson, L. B. 2015. *Specifications grading: Restoring rigor, motivating students, and saving faculty time.* Stylus Publishing, LLC, Sterling, VA, USA.

[34] Pintrich, P. R. and De Groot, E. V. 1990. Motivational and self-regulated learning components of classroom academic performance. *Journal of educational Psychology* 82, 1, 33-40. http://dx.doi.org/10.1037/0022-0663.82.1.33.

[35] Pope, L., Parker, H. B., and Ultsch, S. 2020. Assessment of Specifications Grading in an Undergraduate Dietetics Course. *Journal of Nutrition Education and Behavior.* 52, 4, 439-446. https://www.jneb.org/article/S1499-4046(19)30959-5/fulltext

[36] Prasad, P. V. 2020. Using Revision and Specifications Grading to Develop Students' Mathematical Habits of Mind. *PRIMUS*, 30:8-10, 908-925, DOI: 10.1080/10511970.2019.1709589

[37] Rapaport, W. J. 2011. A triage theory of grading: The good, the bad, and the middling. *Teaching Philosophy* 34, 4, 347-372. http://dx.doi.org/10.5840/teachphil201134447.

[38] Rapaport, W. J. 2012. HOW I GRADE. Retrieved 1 August 2023 from https://cse.buffalo.edu/~rapaport/howigrade.html

[39] Reeves, D. B. 2008. Leading to Change / Effective Grading Practices. ACSD. Retrieved 1 August 2023 from https://www.ascd.org/el/articles/effective-grading-practices

[40] Sanft, K. R. , Drawert, B., and Whitley, A. 2021. Modified specifications grading in computer science: preliminary assessment and experience across five undergraduate courses. *J. Comput. Sci. Coll.* 36, 5, 34–46. https://dl.acm.org/doi/abs/10.5555/3447307.3447310 [41] Schunk, D. H. and Zimmerman, B. J. 1998. Self-regulated learning: From teaching to self-reflective practice. Guilford Press, New York, NY, USA.

[42] Shoen, H. L., Cebulla, K. J., Finn, K. F., and Fi, C. 2003. Teacher Variables That Relate to Student Achievement When Using a Standards-Based Curriculum. *Journal for Research in Mathematics Education JRME* 34, 3 (01 May. 2003), 228-261. http://dx.doi.org/10.2307/3003477

[43] Townsley, M., & Buckmiller, T. 2016. What Does the Reesearch Say about Standards-Based Grading? A Research Primer. https://eric.ed.gov/?id=ED590391

[44] Wenger-Trayner, E. and Wenger-Trayner, B. 2015. Communities of practice: a brief overview of the concept and its issues. https://www.wenger-trayner.com/introduction-to-communities-of-practice/

[45] Wenger, E., 1998. *Communities of practice: Learning, Meaning, and Identity.* Cambridge University Press, Cambridge, UK.

[46] Wenger, E. McDermott, R. A., and Snyder, W. 2002. *Cultivating communities of practice: A guide to managing knowledge.* Harvard Business Press, Boston, Massachusetts, USA.

[47] Winne, P. H. 1997. Experimenting to bootstrap self-regulated learning. *Journal of educational Psychology* 89, 3, 397-410. http://dx.doi.org/10.1037/0022-0663.89.3.397.

[48] Zimmerman, B. J. 2000. Attaining self-regulation: A social cognitive perspective. In *Handbook of self-regulation*, M. Boekaerts, P. R. Pintrinch, and M. Zeidner Eds. Academic Press, San Diego, CA, USA, 13-39. http://dx.doi.org/10.1016/B978-012109890-2/50031-7.