Advanced Education @ Scale: Feedback Isn't Cheap, But It's Effective

Spencer Wilson

University of Colorado Boulder

Abstract. With the growing promise of high income for advanced education in IT fields, many higher academic institutions continue to observe higher enrollment in their computer science programs and related fields [1] [2] [3]. Over time, this has lead to higher enrollment in courses, limiting the units of one on one support available to students, forcing a change in methods to teach effectively at a larger scale. Accordingly, many courses have gone to focus on teaching at scale, or continuing their previous teaching methods with limited exploration in adaptation to scale. In this paper, we propose the use of peer to peer reflective un-grading interviews to scale effective teaching in the Principles of Programming Languages course based on lessons learned from review of literature on effective education at scale.

Keywords— Education at Scale, Computer Science, Principles of Programming Languages, Peer Grading, Ungrading, Interview Grading

1 Introduction

As is standard in academia, there is a lag between increased enrollment and funds for the hiring of additional staff. So, over time, we have seen the number of staff resources available per student decreases. With this unit resource of student to staff interaction decreasing a key value add proposition disappears and an important question arises: How do we provide an effective learning experience to our students at scale? In this paper we explore the use of peer to peer interview grading in an ungraded and self reflective model for pair assessment on complex lab assignments for the Principles of Programming Languages.

We will propose a matrix system of measuring methods of education against effectiveness as the number of students enrolled in the course increases. We then go on to detail and chart the current dimensions of two sample course structures for the Principles of Programming Languages with respect to this matrix system. Finally, we will demonstrate the value added by changing the current method of interview grading in one course from un-grading non-reflective model managed by course staff to a reflective un-grading model manged by peer to peer assessment.

2 Background

2.1 Scale-ability

For this paper, scale-ability in education refers to providing consistent learning opportunities to as many students as possible. Some obvious places to look for scale-able

education tools are the use of artificial intelligence in the classroom, and the world of online learning [4] [5] [6]. While Ai in the classroom sounds awesome, it is currently burdensome to implement, so we'll move on. What is found to be most important in scaling education online is the encouragement of collaboration between students. After all, more students in the classroom means more students that can interact with other students. Beyond technology integration's, this is the most scale-able resource for the course as enrollment increases.

Peer Grading Having students grade each other is considered a must have for effective online education at scale CITE. While many students are resistant to peer grading and do not believe it to be as helpful as feedback from their course staff, it has been shown to be effective. This scales infinitely as more students yields, more people to perform reviews [4]. Perhaps the most important aspect of doing this effectively at scale is to have a way of assessing the students review capabilities. The literature suggests an effective method to ensure effective peer grading is to have some kind of training assignment. Here, students complete an assignment to demonstrate acceptable knowledge of the peer review process early in the semester [7]. This method has been employed extensively in the online learning environment where scale is potentially limitless.

Discussion Forums Additionally, to increase a sense of belonging and community in a large class - be it online or in person - we see a recommendation for online discussion forums [8] [4]. Here many students are able to engage with the material and start discussions with their peers. It is best practice to have course staff monitor and collaborate on this forum as well. While this which requires some time, this is often worth the effort for larger sized classes as it engages students on some semi-synchronous forum where they can ask questions and discuss topics beyond the confines of class time.

2.2 Effectiveness

For this paper, effectiveness in learning refers to providing learning that engages the student and enables future creations. We will focus on the bloom taxonomy of learning which suggests a linear progression of ability exists in student minds from remember, understand, apply, analyze, evaluate, to create [9].

Interview Grading A tool for effective instruction explored at various institutions is to give students an oral assessment of their work called "interview grading". In interview grading, students **evaluate** their master with an oral review of their written assignments potentially including novel questions on the content covered in the assignment. This works well both in traditional and ungraded methods described below. Interview grading has been shown to hold value for students being accountable to their own learning. It works best in a small class setting where the instructors can manage all of the interviews [10] [11]. However, it can be done at scale, by offloading the effort to support staff such as graduate teaching assistants and graders [12]. It is important to note that this doesn't continue to scale well as more students means more time for grading by expert course staff.

Ungrading Now a concept that can be very valuable in interview grading is the transition to an upgrading model for the assignments. Here we move away form a model

of grading out of one hundred points and toward a model of "X", "✓-", "✓", "✓+", or some other naming model to represent a distinction from work that is unacceptable (X,✓-) versus "good enough" (✓), or even exceptional (✓+). In various un-grading models such as reflective un-grading, contract grading and standards based grading we move the staff focus away from time obsessing over the difference in grade from an 85% to a 86%, and instead state, that's a "✓" and let's instead focus on providing substantive feedback to our students [13] [14] [15]. While this requires constant buy in from the course staff and students to ensure success across the term, the model has proven effective in many college courses including upper division topics [15] [16][17][18].

Reflective Learning In reflective learning, we ask students to have agency in their own education and constantly reflect on what they have learned, what they are struggling with, and how they could potentially apply what they have learned to reach their own goals. In fact, there is a model of ungrading built around this concept sometimes called "reflective un-grading" or "big-U Un-gradding" [13]. Here we develop a learning environment where students must author self reflections and even recommend their own grade for the course. We as course staff must then decide if the students self reflection and decided grade is accurate, or how it differs and discuss significant differences with the students.

2.3 Matrix

We propose the following matrix with effectiveness on the horizontal axis, scale-ability on the vertical axis and the goal being something that is highly effective even at larger scales as shown in figure 1.

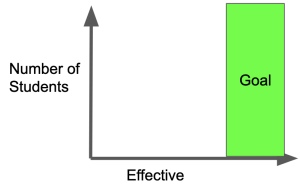

Fig. 1. caption

2.4 Peer Institutions

So we might ask, are any schools using the above methods of effective or scaled instruction for their equivalent to a Principles of Programming Languages (PPL) course? Looking internationally, the big schools tend to focus on grading solely based on exams, occasionally with leniency added in for completion of homeworks [19][20]. Looking domestically, the schools appear to implement a traditional grading structure with a variety of homework assignments being auto-graded and no indication that any of the student assessments are being given substantive feedback, followed by exams [21] [22]

[23] [24] [25] [26] [27]. This is not an exhaustive exploration of course offerings in PPL, and it's difficult to know the full course structure from the provided syllabus online. This is provides only a rough sketch of the landscape today.

2.5 A story of two courses

Now let us consider the story of two courses relative to the bloom taxonomy. Version 1 will refer to a project based course with a long history of success at smaller scales, while Version 2 will refer to a lecture based course that has appeared to succeed on a large scale over recent years.

In version 1, the following assessments are used to construct a course that in practice is shown to be highly effective but struggling to stay effective at increasing scale (challenges over 150 students):

1. Participation: students **analyze** information through discussions during class sessions.
2. Labs: students **analyze** topics of interest which serves as the basis of student learning. This is performed in teams to increase collaborative creativity and is ultimately auto-graded.
3. Grading Interviews: students **evaluate** their mastery on the lab material with one on one interviews with the course staff with an X/✓+ style score returned and limited personalized feedback.
4. Exams: students **create** novel solutions to relevant problems in a timed assessment that is manually graded and returned to students with limited feedback.

In version 2, the following assessments are used to construct a class that has been shown in practice to provide less value to students in the long run, but is consistent in its own efficacy at growing scales (up to 280 students):

1. Lectures: students **remember, understand and apply** ideas in a guided form that serves as the basis of student learning
2. Weekly Assignments: students **apply** mostly isolated topics with auto-grading applied and limited or automated feedback provided.
3. Mini Projects: students **analyze** multiple topics in connection to eachother on an auto-graded project.
4. Exams: students **understand, apply, and analyze** topics in a timed assessment that is manually graded and returned to students with limited feedback.

2.6 Courses on the Matrix

3 Proposal

FAKE: Below we detail one change to version 1 project based course structure that has demonstrated better effectiveness at larger enrollment sizes while keeping the value of interview grading for students.

REAL: Considering that scaled education is best facilitated through peer to peer interactions and methods of informed peer grading, we explored a change to the interview grading process for the project based version 1 of the course. Grading interviews facilitate by course staff are horrid to scale and we've seen this overtime with grading interviews shrinking from 20 minutes down to 15 or even 10 minutes in certain

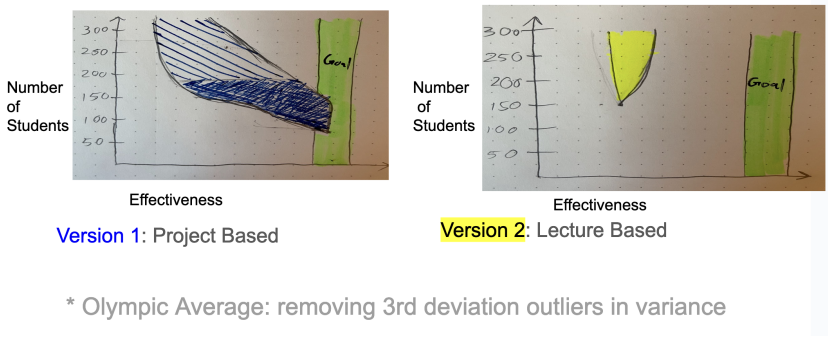

Fig. 2. version 1 and 2 of the principles of programming languages course relative to the goal of effective and scale-able education

courses. Now, students interview each other instead of interviewing with the course staff. Students then grade themselves and reflect on what they learned and where they have gaps in mastery. The interview and reflection are submitted to the course staff for final grading decisions. While course staff is still in the loop and spends time on grading students work, they employ an "X/✓+" grading system and re-prioritize time to giving students constructive feedback on their gaps in knowledge and connecting students to appropriate resources for further learning.

4 Implementation

The peer to peer interview process tested has 4 phases

1. Training Phase
2. Interview Phase: preparation, recording, and submission
3. Reflection Phase: self reflection, conference, write-up, and submission
4. Feedback Phase: review of submission, feedback, interventions

4.1 Training Phase

The training phase may happen only at the beginning of the semester, but can be reassigned to students throughout the term to re-commit the student to the interview grading process. In the training phase, students are given a series of videos on mock-interviews with a grading rubric for the interview using an "X, ✓-, ✓, ✓+" grading system for the topics in the interview. Students are asked to grade the interviewee against the rubric and submit their solutions to an automated grading tool which compares the students proposed grades to the known grade of the mock interview. While this effort had large upfront cost, this sample of the grading process has been shown in other studies to provide great value in reducing overhead throughout the semester by helping to set clear expectations for students on the peer interview process for this course [7].

4.2 Interview Phase

Preparation Consider students Bilbo and Frodo have just completed a lab as a student team. They first start by preparing for the interview. They meet in-person, or over zoom, and discuss what they learned during the lab. They then download the interview question set for the lab and prepare to record their oral interview.

Recording Next, Bilbo and Frodo will record their peer to peer interview. This is designed to take about 15 minutes. Bilbo will ask question 1 (provided from the course staff) to Frodo and listen to Frodo's response, if needed, Bilbo can assist in answering the question. Next, Frodo will ask question 2 to Bilbo, and so on until the questions are complete or the 15 minutes are over. This should be a single take video without edits requiring limited time for the students.

Submission Ending the interview phase, these students submit an mp4 file of their peer to peer interview to canvas.

4.3 Reflection Phase

Self Reflections Having submitted the interview, Bilbo and Frodo now have access to a rubric for the interview assignment on an [X, ✓+] scale. Bilbo will go off and reflect on the interview with respect to this provided rubric. Bilbo will then write up his perspective on how he performed on each question, how Frodo performed on each questions, and how the team worked together. Bilbo must also answer some questions about what he feels he understands about the lab and where he identifies room for personal growth. Meanwhile, Frodo is completing this same exercise separately. Recall that these students have training on how to assess a grading interview and understand the purpose of the rubric in guiding their own self assessment.

Conference Now that Bilbo and Frodo have their self reflection complete, they meet together again. They discuss their results and see if they came to the same grades. If their grades do not align with each-other, then they must pause to have a respectful and constructive discussion on better assessing their performance on the interview. This also serves as an opportunity for the students to discuss those pain points and additional resources to perform better in the future.

Write-up Now Bilbo and Frodo can complete a write-up together summarizing the individual confered scores, what they've done well, where they need more practice, and any additional thoughts or questions they have on the interview process and course content.

Submission Finally, Bilbo and Frodo submit this single write-up to canvas.

4.4 Feedback Phase

Review of Submission Enter course staff in the loop. Our course staff now review all submitted videos and associated write-up. The course staff decides if they agree or disagree with the self reflections.

Feedback The course staff now provides the student their respective grades on the interview and proceeds to prioritize individual feedback for the students. e.g. Frodo, you identify topic B as a challenge for you, but in review of the video I would recommend starting with the precursory topic A in your studies.

Interventions On occasion, students will not assess themselves correctly. Sometimes students will inflate their grade in an attempt to score higher than earned, but far more common, students will list their score much lower than what is earned [13]. In both cases, the course staff is responsible to intervene and work with the student to help them understand what their true mastery level on the content is and the personal reasons they may be inflating or deflating their personal grade.

5 Experiment

To better understand the potential value and pitfalls to this method of interview grading, we took informational interviews with four current teaching assistance in the department of Computer Science with varying levels of prior experience in interview grading. In each interview we explained the current common interviewing process, then the proposed peer interviewing process. We then posed the following questions:

1. What clarifying questions do you have about the concept?
2. What do you see as potentially valuable in this concept?
3. What concerns do you have about this concept?
4. How would it match your course?

FAKE: ...We ran a control in Spring 2024 with 190 students, then ran a pilot of peer interview grading with in Summer 2024. We made the following changes to the process based on summer findings... Finally ran the course at scale with peer interview grading with 200 students in Fall 2024...

6 Results

The data found in interview of current teaching assistance can be found here.

6.1 Value for Students

In interviews, current teaching assistance have expressed excitement in the potential for this system to provide students with more opportunities for collaboration and self reflection. The concept that this will provide more actionable feedback and assistance to students in furthering their studies is what we find most interesting. By moving time toward feedback we have the ability to target student individual challenges and construct new practice material that students can leverage to hone their mastery of the course content.

6.2 Value for Staff

As for the value potential for current staff, there is significant interest in the opportunity to spend less time fine tuning a students grade and instead focus on working with students in improving their own mastery of the material, their presentation abilities, and their own ability to assess the quality of their own submissions. One TA notes with a smile on their face "I'm really excited for the opportunity to help a student work through why they might undervalue themselves and potentially work through 'imposter syndrome'". We also note a potential equity value in this process to help women who historically disproportionately undervalue their solutions have important discussions in properly valuing the merit of their own solutions. With proper training the teaching staff can increase their empathy for student experiences and help the students realize their full potential.

6.3 FAKE RESULTS

In running the course we actually found... Replacing staff lead interview grading with peer to peer interview grading has demonstrated an opportunity to teach more students more effectively with the same number of course staff as shown in figure 3. In addition, students have reported a better sense of confidence in their performance as shown in figure 4.

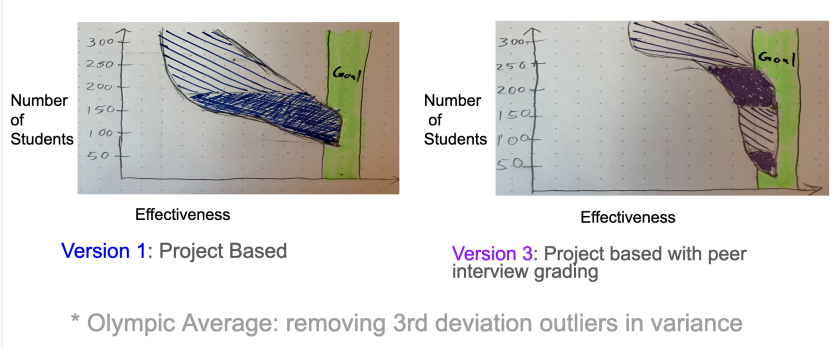

Fig. 3. version 1 and 3 of the principles of programming languages course relative to the goal of effective and scale-able education, demonstrating better scalability with the addition of peer to peer interview grading in the project based course

7 Future Work

Many fair criticism are raised through the experiment process that require further exploration. Consider, what are the FERPA concerns around students peer grading? Would this work in the real world where not all students behave with high standards of ethics? What if students are not paired well and are sabotaging each other or lying to protect one another? How will the staff be trained to facilitate their transition to the

Topic	Term 1 (0 - 10)	Term 2 (0 - 10)
Student Confidence	2	7
Student Performance	4	6
Staff Availability	1	5
Staff Emotional State	0	8

Fig. 4. demonstration of student confidence

new role of having potentially challenging discussions with students on their inflating or deflating of personal assessment?

8 Conclusion

We have demonstrate the value added proposition by changing the current method of interview grading in one course from un-grading non-reflective model managed by course staff to a reflective un-grading model manged by peer to peer assessment. In creating a framework where students can reflect on their own understanding of the material, we have encouraged our students to challenge their own mastery of course topics and go deeper than they otherwise would have. Additionally, this model has allowed our course staff to re-allocate time away from tediously grading our students performance on percentage basis, and instead focus on providing targeted feedback to our students based on their needs and personal ability. While there is more work to be done, we hope that this structure continues to see adaptations that better enable our students success, while improving the experience for the course staff as well.

References

1. Zweben, S., Bizot, B.: 2015 taulbee survey continued booming undergraduate cs enrollment; doctoral degree production dips slightly Accessed: 2023-09-20.
2. Zweben, S., Bizot, B.: 2018 taulbee survey undergrad enrollment continues upward; doctoral degree production declines but doctoral enrollment rises Accessed: 2023-09-20.
3. Zweben, S., Bizot, B.: Cra 2022 taulbee survey: Record doctoral degree production; more increases in undergrad enrollment despite increased degree production Accessed: 2023-09-20.
4. Martin, F., Ritzhaupt, A., Kumar, S., Budhrani, K.: Award-winning faculty online teaching practices: Course design, assessment and evaluation, and facilitation. The Internet and Higher Education **42** (2019) 34–43
5. Berge, Z.L.: Changing instructor's roles in virtual worlds. Quarterly Review of Distance Education **9**(4) (2008) 407–414
6. Alam, A.: Employing adaptive learning and intelligent tutoring robots for virtual classrooms and smart campuses: Reforming education in the age of artificial intelligence. In Shaw, R.N., Das, S., Piuri, V., Bianchini, M., eds.: Advanced Computing and Intelligent Technologies, Singapore, Springer Nature Singapore (2022) 395–406
7. Gehringer, E.F.: Electronic peer review and peer grading in computer-science courses. SIGCSE Bull. **33**(1) (feb 2001) 139–143
8. Smith, A.J., Boyer, K.E., Forbes, J., Heckman, S., Mayer-Patel, K.: My digital hand: A tool for scaling up one-to-one peer teaching in support of computer science learning. In: Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education. SIGCSE '17, New York, NY, USA, Association for Computing Machinery (2017) 549–554
9. Unkown: Bloom's taxonomy Accessed: 2023-11-02.
10. East, J.P., Schafer, J.B.: In-person grading: An evaluative experiment. SIGCSE Bull. **37**(1) (feb 2005) 378–382
11. Ruehr, F., Orr, G.: Interactive program demonstration as a form of student program assessment. J. Comput. Sci. Coll. **18**(2) (dec 2002) 65–78
12. Grunwald, D., Boese, E., Hoenigman, R., Sayler, A., Stafford, J.: Personalized attention @ scale: Talk isn't cheap, but it's effective. In: Proceedings of the 46th ACM Technical Symposium on Computer Science Education. SIGCSE '15, New York, NY, USA, Association for Computing Machinery (2015) 610–615
13. Flaherty, C.: When grading less is more Accessed: 2023-09-15.
14. Stommel, J.: How to ungrade Accessed: 2023-09-15.
15. Owens, K.: A beginner's guide to standards based grading Accessed: 2023-09-06.
16. Chen, L., Grochow, J.A., Layer, R., Levet, M.: Experience report. In: Proceedings of the 27th ACM Conference on on Innovation and Technology in Computer Science Education Vol. 1, ACM (jul 2022)
17. Mittell, J.: Rethinking grading: An in-progress experiment Accessed: 2023-09-15.
18. Stommel, J.: Ungrading: A bibliography Accessed: 2023-09-15.
19. Scherer, G., Dagand, P.E., Jourdan, J.H., Pottier, F.: Functional programming and type systems Accessed: 2023-10-24.
20. Nipkow, T.: Functional programming and verification Accessed: 2023-10-24.
21. Aiken, A.: Cs242 website Accessed: 2023-10-24.
22. Chong, S.: Cs152 - harvard university - programming languages Accessed: 2023-10-24.

23. Chlipala, A.: Mit 6.512, spring 2023: Formal reasoning about programs Accessed: 2023-10-24.
24. Foster, N.: Cs 4110 (fall 2012) programming languages and logics Accessed: 2023-10-24.
25. Harper, R.: 15-312 principles of programming languages Accessed: 2023-10-24.
26. Wortzman, B.: Cse 341: Syllabus:: Programming languages Accessed: 2023-10-29.
27. Chasins, S.E.: Cs 164, fall 2022 — syllabus: Programming languages and compilers Accessed: 2023-10-29.