

# Principles of Programming Languages @ Scale: The Value of Student Collaboration

Spencer Wilson

University of Colorado Boulder

**Abstract.** The Principles of Programming Languages (CSCI 3155) course at the University of Colorado Boulder is experiencing growth in reported dissatisfaction in quality of education by students. This challenge is compounded as our student to staff ratio grows and our ability to hire highly qualified support staff for the course shrinks. Our students deserve a world class education through "learning by doing" around topics which are integral to success in the software industry and other related career fields. In this paper we design, detail, and test an approach of peer-interviews with graded self-reflections to re-prioritize learning over grading and reallocate staff time to course resource improvements contrasted against the existing staff-interview process whereby students are graded for accuracy and staff time is consumed by facilitation of these interviews.

**Keywords**— Education at Scale, Computer Science, Principles of Programming Languages, Peer Grading, Un-grading, Interview Grading

## 1 Introduction

The Principles of Programming Languages (CSCI 3155) course at the University of Colorado Boulder was developed with an assertion that the best way to learn is by doing. This course takes an approach of a project-style lecture by which students should perform work and readings on their own time and attend class ready to ask questions about what they do not understand in the content. This has been situated as rather unique course design within its department and accordingly has been met with resistance by students for many years. After a few years of hiatus during which a more traditional classroom was hosted for CSCI 3155, the project-style course was offered in Fall 2022 to around 150 students and met with such massive resistance by students including organized efforts by students against against the faculty of the course. The student concerns were centered on a feeling of discomfort around the structure of the course, specifically that the students wanted to be "taught" information. Many students reported that they did not find the textbook helpful for their learning and that the information presented at lecture was not sufficient to engage them at the level of understanding they would arrive to lecture with. Accordingly, the students performance on exams was poor relative to an idealized performance.

Industry demand for a workforce skilled in developing software drives increased enrollment in computer science and related fields at universities as well as coding boot-camps. For large public institutions like the University of Colorado Boulder, this continued growth in enrollment comes with larger format lecture halls between the

course taught today and the course taught a decade ago. Large institutions, such as these, often lag in their ability to hire additional qualified staff and construct sufficient physical spaces to continue to host these courses in multi-session, smaller, more intimate discussion spaces rather than 300-student lecture halls. Accordingly, the Fall 2022 course was offered in a 285 person lecture hall with around 150 students completing the course supported by 13 staff members. While the student-to-staff ratio of CSCI 3155 was lower than most courses in the department, the quality of staff was severely lacking this term as many staff members required training on the material and the duties of their roles within the course. This lack of quality staff was reported by students as they found that various members would not be able to provide sufficient assistants on the material or explain the related topics in a way that was relatable to the students level of mastery. This was even exploited by students as they discovered that some members of the staff would abandon trying to help and openly provide students with solutions to the assignments.

As students continue to express dissatisfaction with the structure of the course, the faculty has a responsibility to assess the viability of the current structure. CSCI 3155 hosts a variety of learning goals which are imperative for students to meet for success in industry positions. We assert with evidence that the project-style course to facilitate "learning by doing" provides the highest quality education in preparing students for success in their future careers. However, we recognize that elements of the course are leaving gaps in outcomes for students and encouraging unethical behaviors by students and staff alike.

With more students in the classroom, comes an opportunity to engage students in more peer-centered activities which increase students' sense of belonging to the community, soft skills for collaboration, and technical ability via the necessity to adequately discuss complex ideas with their peers. We define a method of peer-interviewing with graded reflective assessments to engage students in a highly scale-able manner that improves student-agency in learning while reallocate the time of the limited qualified staff in the course. In early 2024, we conducted research-interviews with students that completed CSCI 3155 in Fall 2022 and report on themes found through the research study. We explain how each theme suggests successes and shortcomings of the current course structure with the staff-interview model. We then report on the research participant's performance in the proposed peer-interview model via a simulation of the peer-interview.

## 2 Background

### 2.1 Current Course Structure

Let us frame the discussion of background around the current course structure last executed in Fall 2022 compared against what is described in the literature as effective scale-able instruction. We will detail each of the seven course components against the blooms taxonomy[1].

1. Class Participation (05%)
2. Class Preparations (00%)
3. Lab Assignments (06%)
4. Lab Surveys (01.5%)
5. Lab Exercises (10%)
6. Staff-Interviews (22.5%)
7. Exams (55%)

**Class Participation (05%)** Each lecture event includes formative assessments in which students build to an **analyze** level of cognition on the current learning goals through in-person discussions during class sessions. This implements a variant of the flipped classroom encouraging students to actively participate in the lecture by asking questions of the current prescribed learning goals to guide that days lecture discussion [2]. The course session begins with students asking questions about the lab which the instructor uses to construct a priority queue of topics for that days lecture. This is observed in the literature as an effective teaching method which is has seen success at many levels of college teaching provided that the learner is given adequate resources to support their self-directed learning prior to attending class [3]. This is a graded component of the course which awards student up to five percent of their grade for their labor. In addition to participating actively at lecture, students may participate using the online discussion forum (piazza) to be awarded some of their class participation credit [4].

**Class Preparations (00%)** Each lecture event is hosted in the flipped classroom model where students are expected to come to class having attempted some amount of work preparing them for active discussions during that course session [2]. Preparation for lecture includes attempting portions of the lab assignment, required and recommended readings, or watching video lectures on key topics. Lecture preparation for each student is the key means for students to build their mastery through **remember** and **understand** toward an **apply** level of cognition of the related learning goals prior to attending lecture where we expected to build to an **analyze** level of cognition. In an ideal world, student come to lecture having attempted some amount of their lab assignment using available readings and internet resources to get as far as they can, and to reach some blocking point where they have a question to discuss at lecture. Presently, this labor is not directly rewarded with course credit, but instead rewarded by outcomes on the related formative assessments.

**Lab Assignments (06%)** The course is composed of six lab assignments which are assigned in two week blocks. Each lab assignment is a formative assessment where students build toward a **create** level of cognition of the learning goals in designing solutions to the assignment which require connecting many prerequisite ideas to an optimized solution for the problem. Although students can complete the lab with a lower level of cognition by stopping once they have a working solution without considering optimizations, or by plagiarizing others work as their own. Successful completion of each lab expects that students take on additional readings and research to grow the early levels of cognition prior to actively engaging at lectures to grow their mid-level cognition for the learning goals. All students complete the same lab synchronously in teams of two or three students. The lab is auto-graded for correctness against a set of predefined tests which are partially shared with the students. This use of auto-grading is critical to allow for a highly scale-able resource for immediate feedback to the students as course enrollment continues to grow.

**Lab Surveys (01.5%)** After completing each lab assignment and before completing their staff-interview, each student completes a survey form where they reflect on their learning from the lab assignment, document the time spent, and document their successes and challenges from the lab. This lab survey provides the course staff with some

timely feedback about how well the students are performing against the learning objectives. While the lab survey does receive some credit toward the student grade (1.5% of the full course grade), this has limited credit for the labor as described in labor-based models of grading [5]. This also leverages some reflective learning techniques whereby we ask students to reflect on what they have learned, state what went well and what did not [6]. In practice this allows the course staff to gain insights on which topics students are succeeding with, and which topics warrant some deeper discussion early in the next lab as prerequisite knowledge which they should have mastered in the previous lab but fell short.

**Lab Exercises (10%)** Lab exercises are composed of three components:

1. Lab Code Checkpoint
2. Lab Checkpoint Online Quiz
3. Post-Lab In Class Quiz

*Lab Code Checkpoint* To encourage consistent progress on their code a lab checkpoint is assigned, requiring students to complete some minimal number of auto-graded tests on their written code in the first week of the two week lab session. This checkpoint is confirmed and entered manually by the course staff requiring minimal effort. This rarely provides students with feedback on their progress toward the complete lab.

*Lab Checkpoint Online Quiz* To encourage consistent progress on the labs theoretical concepts and terminology, an auto-graded quiz is administered online giving students multiple attempts to answer questions which assesses up to an **apply** level of cognition on the relevant topics.

*Post-Lab In Class Quiz* The post-lab in class quiz is taken by all students at the first class period after the lab has been completed. In this formative assessment students are asked to demonstrate up to a **create** level of cognition on the work from their recently completed lab assignment. This particular event is peer graded, having students grade each others quiz live during lecture to assist in the scale-ability of providing timely feedback on this assessment. 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 in post-secondary learning [4] [7]. This scales infinitely, as more students yields more people to perform the reviews. However, perhaps the most important aspect of doing this effectively at scale is to have a way of assessing the students grading 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 [8]. This method has been employed extensively in the online learning environment where scale is nearly limitless. Today we do not implement any such assessments to assist the students in growing their ability as peer-graders and presents an opportunity for improvement.

**Staff-Interviews (22.5%)** Staff-interviews are another formative assessment in which students are asked to demonstrate up to a **create** level of cognition on the learning outcomes. In Fall 2022 the staff-interviews held the following format. The results section of this document will further detail the students perception of this structure.

Each member of the course staff conducting interviews has an assured **analyze** level of cognition on the learning outcomes. After completing each lab assignment, each student signs up for a single staff-interview one-on-one with a member of the course staff. The student attends the interview without prior knowledge of the questions that will be asked, and performs the interview in a twelve minute slot. At the end of the interview, the course staff, informs the student how they performed on each question in a 4 point mastery grading scale, then submits the grade and brief performance feedback for the student in the time before the next interview begins. This interview is graded on the basis of student's ability to correctly answer the interview questions within the time provided.

The current structure is not unique in the department of Computer Science in providing staff-interviews, however it is somewhat unique in assessing student performance at the interview in the 4 point grading scale of "Novice", "Approaching", "Proficient", and "Advanced Understanding" whereby the staff can more quickly give students important feedback on how they are performing relative to our expectations and re-prioritize grading time toward giving students targeted feedback [6] [9] [10]. While this requires constant buy-in from the course staff and students to ensure success across the semester as students become co-conspirators in this different educational model, the model has proven effective in many college courses including upper division topics [10] [11] [12] [13] [14].

Where possible, the course staff also takes time to celebrate what the students have already mastered and encourage their continued success. Ideally, the member of the course staff are able to pivot the interview as needed to ask follow-up questions of the student in the Socratic method that encourages the student to create a more comprehensive understanding of the related topics. However, this is variable in practice as each member of the grading staff have a different level of ability and interest in hosting an effective interview [15].

Over the past decade we have documented growth in enrollment to computer science and related college fields of study [16] [17] [18]. Interview grading has seen many successes in lower division courses with larger enrollments [15]. There have also been successes documented for similar in-person assessments for smaller course settings hosted by the course lecturers [19] [20]. However, as course enrollment increases and the required staff to hire increases, it has become increasingly difficult to hire sufficiently skilled hourly and stipend labor in the graduate community to effectively host these grading interviews with students in this specialized upper-division course at scale [21].

Having completed all course work related to a lab assignment, if the student believes that their grade does not adequately represent their mastery of the material, then each student may request a regrade of the lab. In a regrade event, the student must formally request the regrade using our online communication forum (piazza), then schedule a grading interview directly with the course professor. This practice allows students an opportunity to makeup credit when they believe that existing grade is inaccurate. As this places the grading burden directly on the professor of the course, this does not scale well in the event that all students might want a regrade. However, in practice it is rare that students will request a regrade with only around one percent of students requesting a regrade at any point in the semester.

**Exams (55%)** The course is composed of one midterm exam (25%) and one final exam (30%) as the only summative assessments of the course accounting for 55% of the students total course grade. In these exams students are typically asked to

demonstrate an **evaluate** or **create** level of cognition on the related learning goals during a timed in-person assessment. These assessments are manually graded by the course staff to include partial credit aligning with the spirit of the four point scale used for grading interviews. The midterm exam is returned to students with limited qualitative feedback beyond what is embedded in the exam rubric. The final exam is not returned to students.

## 2.2 Proposal

In our analysis of the existing course structure against the literature, we believe that one weak point of the existing course structure lies in the execution of the grading interviews. The grading interviews have begun to fail our students in that they are hosted by inadequately prepared interviewers in time windows that are too small to host an effective interview relative to the students common level of mastery with the material. Many institutions have implemented staff-grading interviews at small scale for upper divisions and large scale for lower division courses. Many schools have also implemented reflective grading for a small scale classroom. However, there is a gap in the literature on how to support our large scale upper division classrooms with in-person oral assessments like staff-interviews, or with the use of reflective grading. This paper proposes a new model of interview grading completed in a peer-to-peer model with reflective grading procedures. This paper gathers data on how students might succeed or fail in the new proposed structure based on themes observed from past students performance in the existing structure.

## 3 Experiment

The experiment section is composed of two subsections. First we will describe our proposal for changing the course structure from a performance based staff-interview to a completion based peer-interview with graded self-reflections. Second we will describe the research study designed to solicit information on how past students experienced CSCI 3155 in Fall 2022 and how they experience a sample of the proposed peer-interview model.

### 3.1 Proposal

In this paper we propose a method of peer-interviews which would not be graded based on the students' correct answers to the interview questions, but instead on the quality of a self-reflection of their peer-interview process and their current understanding of the material. The course staff will only review the students submitted self reflection for grading and feedback. Through this change we emphasize the formative nature of the interview and focus on giving students qualitative feedback on their performance in an ungraded model, leveraging peer interactions as a more scale-able resource than hiring sufficiently trained course staff to host each interview in the staff-interview model. This proposed peer-interview process contains four phases as follows:

1. Training Phase
2. Interview Phase
3. Reflection Phase
4. Action Phase

**Training Phase** The training phase is required at the beginning of the semester and is reassigned as needed to students throughout the semester to re-commit the student to this interview grading process. In the training phase, students are given a detailed introduction to a four point grading scale of "Novice", "Approaching", "Proficient", and "Advanced Understanding" along with an explanation of the reasoning for using this method of grading. Students are then given a series of videos on mock-interviews with a rubric using this four point grading scale. 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 would have large upfront cost, this effort has been shown in other studies to provide great value in reducing overhead throughout the semester by setting clear expectations for students early in the semester [8].

**Interview Phase** For each lab, the course staff prepares an interview document containing questions, alternate questions, hints, solutions, and grading guidelines. Consider hypothetical students Ethan and Ayden have just completed the lab as student team. Ethan and Ayden each play the role of interviewer and interviewee and should take turns asking each a unique question from the interview document of various levels of difficulty.

If the interviewee cannot answer the question sufficiently, then interviewer should review the hints for the question and attempt to provide scaffolding that supports the interviewee to a correct solution. If the neither member of the team can understand the question hints or solutions, then the student team should study before continuing the interview. In studying, Ethan and Ayden might collaborate or study individually based on their unique learning styles. This study session might include resources such as conversational Ai, reviewing the provided solutions to the question, and contacting peers or course staff through resources such as the online discussion forum (piazza) or back-channel student forum (discord). When recommencing the interview, the student team should consider the alternate question provided rather than the original question as they have already learned the solution to this question effectively forfeiting that learning opportunity.

In completing the interview, the students should agree on grading for each participant based on their total understanding of the course content exposed through the interview and not only the students initial answer to the question. Here, we encourage student collaboration in learning and increase student autonomy in resourcing their learning when compared to a 12 minute staff-interview that are closed resources beyond the individual student and the staff member conducting the interview. Additionally, as the staff is no longer conducting the staff-interviews, they have more time to dedicate toward supporting student learning in other ways such as additional office hours or construction of improved readings and videos that help in students early growth through the **remember**, **understand**, and even **apply** levels of cognition.

## Reflection Phase

*Student Reflection and Action Planning* Ethan and Ayden now meet to review their performance on the peer-interview and the lab content as a whole. Students are encouraged to spend about thirty minutes on this exercise. They identify their performance on a selection of learning outcomes for the lab and develop a personal action plan for what they might focus their efforts on in the next lab, taking advantage



of the benefits of reflective learning. While the action plan is personal to the individual, students are encouraged to collaborate in generating ideas for their action plan. Each student submits this via a survey form that allows for the aggregation of student data.

*Grading and Feedback* The course staff allocate time to review each students reflection, ideally with each teaching assistant grading the reflection of students enrolled in their recitation. The teaching assistant grades the student on the quality of their self reflection against a defined rubric which is shared with the students require specific examples over vague explanations. Each teaching assistant is also required to provide constructive feedback to the student where possible such as linking to relevant supplementary material to fill in the students self-identified gaps in mastery.

*Staff Reflection and Action Plan* While each member of the course staff has access to the student reflections, one delegate of the course staff analyzes the student performance as described by the students in their self-reflections. This data analysis is targeted at understanding current student strengths and opportunities for improvement. The delegate presents their finding to the rest of course the staff and the team collaborates on a plan of action that will build on those stores of knowledge to scaffold learning on the topics where students have the most opportunity to grow in their learning journey. In practice, we expect this will take a few hours by the delegate staff member for each lab as well as a one-hour meeting with the full course staff after the completion of the interview phase for the lab. For best results, we recommend that this work is completed as early as possible during the next lab so that the feedback can be leveraged in a timely fashion.

**Action Phase** In the action phase, the course staff executes on their plan for improving the course lectures based on common findings in students' gaps in knowledge. In an attempt to increase transparency of the process and build our students as conspirators to the method of reflective learning, the course instruction includes anonymous quotes from the student reflections and openly recognizes why we are covering certain topics in more depth. The students are also encouraged to act on their own action plans and seek whatever assistant or materials they may need. Toward enabling the students' success, the course staff is listening to students and taking note of what roadblocks exist for the students and actively working at removing those roadblocks wherever staff intervention is necessary while being careful not to remove those critical speed-bumps that students need in order to have autonomy in their own learning to foster a sense of accomplishment.

### 3.2 Research Study Design

In this experiment we conducted a research study with Fall 2022 students of CSCI 3155 at the University of Colorado Boulder composed of three interviews one-on-one with the researcher and one focus group in teams of two participants with the researcher centered on their lived experience in the course and topics related to CSCI 3155, interview grading, and peer interactions. Due to the position of the researcher as a trusted member of the computer science undergraduate community, research participants shared information that would otherwise likely not have been shared which could impact their perceived or actual reputation. As is standard practice, the research study was designed with an informed consent process following the guidelines set forth



by CITI "human research" training to minimize potential risks to the six research participants that agreed to participate in the study.

**Consent Process** The consent process was conducted prior to the first interview including a detailed description of the format of the research, the data collection and review process, the voluntary nature of the research, the potential risk to participants, mitigation strategies for these risks, and the potential benefit of this research study. This took about 15 minutes for each participant.

**First Interview** The first interview focused on understanding the students experience in completing CSCI 3155 at the University of Colorado Boulder in Fall 2022. We sought to understand the students motivations in the course and how various components of the Fall 2022 course impacted their confidence to succeed in learning the requisite material. This interview took around 30 minutes per participant.

**Second Interview** The second interview focused on understanding the students preferences in course design, topics, and class size. We sought to understand what makes a course less enjoyable for this student, and what makes a course enjoyable for this student. In particular we focused on the students experience in larger lecture hall style classes involving over two-hundred students enrolled. This interview took around 30 minutes per participant.

**Focus Group** The focus groups for this research study were indented to host all six research participants at once to create an environment with less social pressures for agreement between participants. However, due to scheduling conflicts, the focus groups were instead conducted in teams of two research participants per focus group. The intention of these focus groups were to allow the research participants to experience a flavor of the interview phase and reflection phase of the proposed peer-interviews and to encourage a group discussion on their experience with the proposed model. The focus group were scheduled for one hour each.

*Consent Process: 2 minutes:* During the consent process, research participants were briefed on the scope and plan for this focus group. Then each participant was required to verbally consent to participate in the focus group prior to execution, or to leave the event. In accordance with CITI training, each participant was advised prior to the focus group, that while the researcher will maintain their anonymity in the study and request that other participants also retain the anonymity of their peers in the study. However, the researcher could not impose such a restriction on the participants in the study.

*Lecture: 8 minutes:* The researcher then provided a brief lecture on the nature of the stack data-structure as an inductively defined data structure that could be represented with Backus-Naur Form (BNF) grammars. This brief lecture included time for the participants to ask questions about how BNF grammars define all sentences that can exist in an infinitely sized language.

*Peer-Interview: 15 minutes:* All participants were formed into teams of two and provided with a physical print out of the interview questions for this event including an ice breaker, one question testing an **analyze** level of understanding to apply BNF grammars to a linked list, and one questions testing an **evaluate** level of cognition on applying BNF grammars to a binary tree. The provided interview documents supported the interviewee by providing space to document their solutions while explaining their thought process verbally. The provided interview documents supported the interviewer by providing followup questions for the interviewee as well as detailed solutions to the question including examples of what an ideal solution could look like and common mistakes made by learners on this topic. Finally, the provided interview document supported the student team by detailing grading recommendations for the questions in an ungraded model defined as follows:

1. Advanced Understanding: clearly and concisely articulates their understanding of the topic, states their assumptions, and uses strong vocabulary correctly in their explanation without excessive need for follow up questions.
2. Proficient: articulates their understanding of the topic with or without some need for follow-up questions from the interviewer. Demonstrates that they understand the topic well enough as pre-requisite knowledge for the next module of the course.
3. Approaching: articulates their understanding in a way which demonstrates their mastery of the underlying information required as pre-requisite knowledge for the next module of the course, but needs to spend some additional time reviewing and learning the topic when possible.
4. Novice: Fails to demonstrate any sufficient mastery of the material. The explanations are vague or unclear. They need excessive follow up questions from the interviewer. They need to spend significant time reviewing and learning the topic when possible.

*Self-Reflection: 5 minutes:* Each student was given access to an online survey to guide their self reflection on their process and learning outcomes from the peer-interview. Each participant was given time to complete this reflection for later review by the researcher.

*Discussion: 30 minutes:* Finally, the focus group concluded with a discussion phase which sought to prioritize student collaboration and ideation in discussing their experience in the provided simulation of the peer-interview process. Additionally, this attempts to understand how this peer-interview process compares to their past experiences with TA-interviews.

**Third Interview** The third interview completed the research study with the participants, seeking to gather more individual data on the participants experience in the focus group. Additionally, this interview debriefs the participants on the research study and asks them about the impact that they would expect to see from the application of this peer-interviewing model.

## 4 Results

The results section begins with an exploration of themes observed in the data collected from research participants which inform the viability peer-interviewing. The results section then concludes with an explanation of the data collected around the proposed peer-interview model.

## 4.1 Themes

Below we identify themes found in the data collected as they relate to the goal of improving the course structure using peer-interviewing to reallocate staff time to other necessary tasks which support student learning.

**Lab Partner Variability** Each student notes that their experience working with a lab partner is quite variable, as would be expected [22]. Interesting in the data is some insights into the unique challenges faced by each participant. When looking at the experience of Participant 13, we see the story of a student who would start each lab as early as they could. They would attempt to engage their lab partner to start with them and be ahead on the material, but often found that their lab partner was not willing to start yet. To see this in an extreme, we look to participant 97 as they detail an experience where they started the lab early in the first week, but their lab partner had already completed the lab in its totality. They asked their lab partner to explain the material and they happily accepted. Once the partner had explained the entire lab, participant 97 didn't feel comfortable asking them for additional help. "I would feel bad saying I don't understand it after he went through it line by line."

While participant 93 actually cannot recall ever specifically having a lab partner assigned for any lab in this course, they would regularly work with a select study group of friends that they had in the course. Participants 58, 93, and 97 detail how they would often try to attend office hours to get help with each lab assignments. Each would attempt to go with their assigned lab partner and their chosen study group for the course. However, when this was not possible they would attend office hours alone to get the help that they could. They would then share what they learned with others when possible.

This presents a challenge for implementing peer-interviewing with lab partners as there will, on occasion be poorly matched teams. For this transition to peer-interviewing to be successful, we must carefully consider how to form lab teams throughout the semester [22] [23]. In general, the feedback around student teaming was positive with each participant noting that they had a few lab partners who they worked well with. They found it nice to meet new "cool people", and to have others with an expected similar background with the material to help them through the assignments. Participant 15 explicitly details how working with a peer aided in their learning as they discussed the material void of the theoretical jargon of the course, leveraging common language from pre-requisite courses.

**Interviewer Variability** The ability of interview staff to host an effective interview varied widely, as is expected from the literature [15]. In this study we observe many students reactions to this variability. Participant 72 and 97 would select interviewers based on availability rather than the individual. They each report variability in how the interviews were conducted. Participant 72 details how they were frustrated as some of the interviewers were willing to give hints and encourage the students growth in learning live at the interview while others would say "if you don't know how to answer this, then we'll move on to the next question" without giving the student any chance to receive credit on the question or learn from their mistakes.

The remaining participants had different rationals for selecting a consistent interviewer to use across the semester when possible. Participant 13 felt that interviews were a "check" that they understood the content of an assignment. They performed well in

the course and quickly found it best to select one interviewer that they found was easier to work with, so that they could complete the interview with as little "frustration" as possible. Meanwhile Participant 58, seeking to achieve the highest grade possible, would intentionally work with the interviewer that they learned was the most willing to give out a high score with limited interest in assessing the students true mastery of the material. Finally, participants 93 and 15 found interviewers that they connected well with on a personal level. They would choose the same interviewer consistently, who would then take time to help explain the course material to them whenever they were under-performing. They each began to form a mentoring bond with their selected interviewer. As participant 15 states, "even if I couldn't do it for myself, I just wanted to do it for [my TA]."

**Crowdsourcing** Each research participant reports different values from the level of collaboration encouraged in the course. Participant 15 describes some nuance of the typical experience in crowd-sourcing questions and answers for the grading interview. Students would openly share the interview questions, and occasionally share how they answered the question. However, these students were uncertain if their answer was correct, so students felt this was not a violation of the course policy as collaboration was encouraged and "solicitation was never truly helpful". Students would then enter the interview with privileged knowledge of the interview questions and still under-perform. Participant 58 took it upon themselves to ask their peers for the questions and answers to interview questions "[It's not because I wanted to cheat but rather because I did not understand the content well enough.]" Meanwhile participants 15 and 93 would consistently ask their peers for the questions of the interviews, but not necessarily ask for the answers as they first wanted to attempt to learn the material themselves. This process was facilitated by other students as participant 97 states "I definitely told other people what was on my interview... We [would share the questions] especially if it felt like a 'wildcard' to students."

## 4.2 Peer Interviews

Below we detail our findings from simulating peer-interviews with research participants. We find that students are able to perform well in peer-interviews as described in the "focus group" subsection of the "experiment" section of this document, but express some concerns about how this would be implemented in practice.

Research participants do express some concerns about this structure however. As one subject group explain that they enjoyed working with each other, but they would be "scared" to do this for a grade with a peer. Accordingly we do not recommend that students are graded on the accuracy of performance during the interview. They also note that they would be quite concerned about doing this with one of the really intelligent peers in the course. Stating, "I don't want to look dumb" in front of another student. This supposes that their peers should have a similar level of mastery of the course material and that the staff should have a higher level mastery (which is not necessarily true in practice). Accordingly these research participants feel less social pressures to perform well on interviews with members of the course staff.

This group also expressed concerns about performing peer-interviews with a close friend. They discuss the tension of it not only being "kind of scary" but also likely to be taken less seriously with a friend. They detail how they would be more likely to give each other the answers directly, where-as a TA will be more willing to wait for them to

state an attempted solution before prompting them toward a correct solution without giving away the answer. In this groups peer-interview they did not give each other the answer to the interview questions, however they also had limited verbal interactions during the interview. The interviewee would answer the question and be seeking more advise, the interviewer would say something to the effect "I think that is correct" or they would ask the interviewee to say more.

Contrary to this conjecture from the first focus group, the other focus group in the study centers on two friends. These participants interviewed each other in a very professional manner. In fact, they were less willing to look at the solutions together. They were coaxing information out of each other using the Socratic method. Their collaboration appeared more effortless as they already understood their peers preferred communication methods and how to encourage their friend to grow in their mastery of the material, as stated by participant 93 "It felt like I was guiding participant 15 into [their] exploration of the question." Separately participant 15 states that "Being coached by someone at your own level of understanding is nice as they use easier to understand terms." This suggests that the concerns of friends peer-interviewing each other would depend on the dynamics of that specific social group.

Participant 97 expresses some concern about structure of peer-interviews. In particular, they explain how it is helpful for them to know if their answer to a question is exactly correct or not. They express that they believe a peer in the course would be less capable of discerning this compared against a member of the course staff. Conversely participant 13 explains how in a TA interview, the TA is supposed to help the student learn the material if they cannot answer the posed question. But that isn't how it worked in their personal experience. Many, perhaps most TAs, would take the following approach. They ask the question to the student. If the student cannot answer the question, then they just move on to the next question and do not support the student in their learning. This issue with TAs not providing substantive feedback was also echoed by participant 72. Combined, this suggests that peer-interviewing might not be better than staff-interviewing, but could be equally as effective at a much lower cost allowing staff time to be reallocated to other tasking while still helping students identify their current level of mastery with the material.

At the peer-interviewing focus group, each participant was asked about how well they understood the topic of BNF grammars before and after their peer-interview. In this limited sample each research participant reported improved confidence with the material comparing before and after the peer-interview as detailed in table 1, presenting evidence that this proposed method of peer-interviewing can improve student learning compared against no interviewing at all.

Participant ID	Before	After	Change
13	3	4	+1
15	2	4	+2
58	2	3	+1
93	1	4	+3
97	2	4	+2

**Table 1.** Table 1: Comparison of participant self reported level of understanding before and after the peer-interview on a Likert one through five scale, followed by the improvement level change. One being "very bad" and five being "very good".

## 5 Future Work

This initial study shows some promise for the viability and value of peer-interviews, but leaves us with many more questions to answer.

### 5.1 Does it work?

The current proposed method of peer-interviewing suggests value when implemented correctly, but only with limited samples. Would this actually work when executed in a class of 150 or even 300 students? What aspects of the student learning environment exist in this modified course structure that could be leveraged in other aspects of the course? How would that change be implemented? When would it not be wise to make such a change? Does it work as well in lower division courses as it does in upper division courses?

**Execution and Measurement:** One next step in this research is to suggest how to measure the success of the proposed peer-interviews, then design and execute the experiment to collect and analyze data while summarizing the benefits and challenges of the model that should be considered prior to future iterations. This will also have to explore if the peer-interviews actually work, considering: what conditions need to be set for the student interviewing team to be successful; what conditions lead to the peer-interview failing to improve the student experience or even hurting their learning; and what conditions lead to the interview not actually occurring and students lying in their self reflection.

**Staff Reflection Phase:** One important advantage of the peer-interviews for the course staff is that the interview data is returned to course staff about one week earlier than it is with staff-interviews. Accordingly, if no students complete staff-interviews, then the course staff would be able to construct their action plan for course improvement based on the student reflections earlier and be able to deliver effective change to the classroom more rapidly. But what impacts would this have on the effectiveness of the review process if course staff had not actually completed an interview with a student and directly observed where students are struggling? Will the staff reflection phase still be as effective?

**How to Re-purpose Staff Time:** This proposal claims that moving from staff-interviews to peer-interviews will increase the time available for staff members to provide targeted support for students. However, it does not specifically detail how that time should be spent to support student learning. In executing the experiment it should be considered how to re-purpose staff time to improve student learning. This might even include some form of regading interview opportunities for students to take by request with a member of the course staff.

### 5.2 Additional Themes

Through this study, additional themes in student successes and challenges were exposed which can inform course corrections which did not appear relevant tot he current study.

**Grading** Students expressed some confusion over the course grading structure as we had two grading systems by the end of the term, described by Participant 72 "I just wanted to pass the class because I was in danger of failing (I think). I remember there being two grading systems and the reason I chose one of them was because there was a possibility of failing the class with the other one." This dual grading system was in name only with an expectation that student grades would not change based on which option they selected. Underlying these concerns from students was the perception of a course curve as participant 97 explains that grading felt subjective relative to their peers rather than objective to a level of understanding of the material: "Honestly, I don't remember anything from this class... PPL was curved a lot. I did not understand an [X] amount of material."

**Office Hours** A recurring theme from students without explicit prompting was the matter of office hours. CSCI 3155 offered forty to sixty hours of dedicated assistance on a first-come first-serve basis in the department's open office space each week. This was an uncommonly large number of office hours for a course in this department, but it coincided with there being too many students at each office hours. Additionally, Participant 97 detailed their personal strategy for attending office hours which exposes some challenges of the current staffing. They would have to choose who to meet with carefully. When they worked with an undergraduates staff member, the undergraduate would often not be able to help sufficiently and would often violate the integrity of the course material by allowing the current students to view the private reference solutions for the course. Sometimes this helped the student and their peers complete the lab, but this student would then make a point to understand the solution and typically also attend office hours with a member of the graduate staff who would be able to help them understand why a solution was correct and how one might arrive at this approach. They would carefully consider which office hours to attend based on what kind of information they needed. They found that these smaller group discussions at office hours with their peers and course staff helped to solidify their understanding of the material. In particular, they would attempt the lab, arrive at a working solution, not understand why that solution works, and then be able to ask a specific question to graduate staff about some section of their solution for a more nuanced understanding of why the solution is valid.

**Not knowing how to prepare for lecture** Each participant made some comment about how the method of lecturing used in the course was new to them and "weird" or "challenging", with many expressing frustrations about how they needed to be "taught". The course intention is that each student takes time to prepare for the lecture. However, many participants suggest that the provided readings did not help them **remember**, **understand**, and **apply** knowledge so that they were adequately prepared for lecture to **analyze** the material, we see one opportunity to improve the course structure. One resource made available to students in Fall 2022 that no research participant mentioned was an access to post-baccalaureate lecture videos on the related learning objectives of each lab on YouTube. As students were aware of the expected readings of the course and not the optional videos which supplemented the readings, perhaps more work should be considered in assigning these videos on a timely schedule as supplemental or even required material along side the expected readings of the course for students who lack the learning processes necessary to gain the requisite information from the required readings.



## 6 Conclusion

In this paper we have demonstrated examples of the growing reports of dissatisfaction from students with their ability to meet their learning goals for the course contextualized by their lived experience. This highlighted challenges faced by students as the quality of staff available to support them was variable and, at times, lacking. Building on the literature we have designed a method for peer-interviewing to support large format upper division courses when the available staff is not qualified to do so effectively. This concluded with a report of how peer-interviews conducted with research participants demonstrates potential to aid in student learning while allowing staff time to be reallocated to other tasking that supports students mastery of the learning objectives of the course.

# References

1. Bloom, B.S., Krathwohl, D.R.: Taxonomy of educational objectives: The classification of educational goals. Book 1, Cognitive domain. longman (2020)
2. Bishop, J., Verleger, M.A.: The flipped classroom: A survey of the research. In ASEE National Conference Proceedings (2013)
3. Halili, S.H., Zainuddin, Z.: Flipping the classroom: What we know and what we don't. The online Journal of Distance Education and E-learning **3**(1) (2015) 15–22
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. Stommel, J.: How to ungrade Accessed: 2023-09-15.
6. Flaherty, C.: When grading less is more Accessed: 2023-09-15.
7. Luca de Alfaro, M.S.: Dynamics of peer grading: An empirical study. In: Proceedings of the 9th international conference on educational data mining. (2016) 62–69
8. Gehringer, E.F.: Electronic peer review and peer grading in computer-science courses. SIGCSE Bull. **33**(1) (feb 2001) 139–143
9. Stommel, J.: Ungrading: A bibliography Accessed: 2023-09-15.
10. Owens, K.: A beginner's guide to standards based grading Accessed: 2023-09-06.
11. Chen, L., Grochow, J.A., Layer, R., Levet, M.: Experience report: Standards-based grading at scale in algorithms. In: Proceedings of the 27th ACM Conference on on Innovation and Technology in Computer Science Education Vol. 1, ACM (jul 2022)
12. Mittell, J.: Rethinking grading: An in-progress experiment Accessed: 2023-09-15.
13. Kohn, A.: The case against grades Accessed: 2023-09-15.
14. Nilson, L.B.: Yes, virginia, there's a better way to grade Accessed: 2023-09-15.
15. Grunwald, D., Boese, E., Hoenigman, R., Saylor, 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
16. Zweben, S., Bizot, B.: 2015 taulbee survey continued booming undergraduate cs enrollment; doctoral degree production dips slightly Accessed: 2023-09-20.
17. Zweben, S., Bizot, B.: 2018 taulbee survey undergrad enrollment continues upward; doctoral degree production declines but doctoral enrollment rises Accessed: 2023-09-20.
18. 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.
19. East, J.P., Schafer, J.B.: In-person grading: An evaluative experiment. SIGCSE Bull. **37**(1) (feb 2005) 378–382
20. Ruehr, F., Orr, G.: Interactive program demonstration as a form of student program assessment. J. Comput. Sci. Coll. **18**(2) (dec 2002) 65–78
21. Cuseo, J.: The empirical case against large class size: Adverse effects on the teaching, learning, and retention of first-year students. The Journal of Faculty Development **21**(1) (2007) 5–21
22. 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

23. Beichner, R.J., Saul, J.M., Abbott, D.S., Morse, J.J., Deardorff, D., Allain, R.J., Bonham, S.W., Dancy, M.H., Risley, J.S.: The student-centered activities for large enrollment undergraduate programs (scale-up) project. Research-based reform of university physics 1(1) (2007) 2–39