An Assessment of Oral Exams in Introductory CS

Peter Ohmann
College of Saint Benedict / Saint John's University, USA
pohmann001@csbsju.edu

ABSTRACT

Assessment of student knowledge is a crucial and challenging part of course design. Especially in computer science courses in the United States, written examinations are very common. While written exams offer a number of advantages in convenience and familiarity, they are also inflexible and prone to question misinterpretation. In contrast to written tests, oral exams offer the prospect of an interactive conversation where students can express their knowledge in a variety of ways while asking clarifying questions.

In this paper, we present and assess our implementation of oral exams in an introductory computer science course. We describe the motivation for and resulting features of our design, including a simplified rubric style for equitable, on-the-fly grading. We also perform an assessment relative to more traditional written exams. We find the time commitment for instructors to be manageable and comparable to traditional exams. Through post-semester surveys, students self-report spending slightly more time studying for oral exams, but rate the difficulty as similar to written exams. Both qualitative and quantitative student feedback indicates that oral exams can be effective and well-received.

CCS CONCEPTS

• **Social and professional topics → Computer science education**; **CS1**; **Student assessment**; Computational thinking;

KEYWORDS

Oral exam, written exam, student survey

ACM Reference Format:
Peter Ohmann. 2019. An Assessment of Oral Exams in Introductory CS. In *SIGCSE '19: 50th ACM Technical Symposium on Computer Science Education, February 27–March 2, 2019, Minneapolis, MN, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3287324.3287489

1 INTRODUCTION

This work was undertaken while the author was a faculty member at Xavier University in Cincinnati, OH, USA. All discussion of curriculum and course design refer to the Computer Science program and course offerings of Xavier University during the 2017–2018 academic year.

Individual assessment of student knowledge is a crucial component of any course design. Rigorous in-class examinations (exams) are typically a core method. While exams vary immensely across universities and cultures, two common types of exams are well-known: the written exam and the oral exam.

In a written exam, students are given questions on paper and work alone to write answers to the questions. Usually, written exams have a fixed time limit. When a student has finished with the exam (or at the end of the time limit), they turn in the written copy for later grading by the instructor or course grader. Written exams typically involve little to no direct interaction between the student and instructor, aside from short clarification questions.

Oral exams are, instead, intended to foster unscripted discussion between the evaluator and student. An individual student (or, sometimes, a group of students) meets face-to-face with the evaluator at their own unique time, separate from the rest of the class. The student(s) are then asked a smaller set of questions, which they answer orally, perhaps with the occasional use of paper or whiteboard. Oral exams are typically graded during or immediately after the exam ("on-the-fly" in some fashion).

This paper describes our experience in implementing oral exams as the final examination for an introductory computer science course in a university setting. Through oral exams, we hoped to better assess the depth of student knowledge on exam topics (without question wording or guessing getting in the way). Ideally, we also hoped to reduce stress associated with writing under time pressure. We describe our design for the exam, our goals, and how we attempted to minimize some commonly-cited negative aspects of oral exams. We also include sample questions that we used in our exam implementation, and describe our use of pre-scripted prompts to ensure consistency across students' exams. These samples also help to demonstrate our simplified "check-box" rubric format, which is designed to reduce unintended bias during examination. We found this rubric format to be easy to use and efficient for on-the-fly grading; it may be useful to others hoping to implement oral exams into an introductory computer science course.

We also perform a preliminary evaluation of our design by comparing student reactions and our own perceptions to those when students took a traditional written exam in the same course. While our sample sizes are insufficient for an extensive experiment at this stage, we nevertheless glean encouraging indicators from a post-semester student survey. Students find the oral exams to be approximately as difficult as written exams, but report spending about 50% more time studying for oral exams. Student comments indicate that some students found the oral exam unique and forgiving, and some felt more relaxed (somewhat contrary to prior research on oral exam anxiety). From an instructor standpoint, we found the exam organization, scheduling, and grading to be as manageable as a traditional written exam, though we raise some concerns with scalability to large class sizes. We also saw positive signs in student performance on the exam. Overall, both instructor and student reactions to the oral exam format were positive and encouraging.

SIGCSE'19, February 27–March 2, 2019, Minneapolis, MN, USA

ACM ISBN 978-1-4503-5890-3/19/02…$15.00
https://doi.org/10.1145/3287324.3287489

1.1 Paper Organization and Contributions

Our primary contributions are as follows:

- We describe an instantiation of final oral exams into an introductory CS course.
- We provide sample questions and formatting/grading rubric styles to allow others to implement similar exams.
- We present preliminary qualitative and quantitative student and instructor feedback on the effectiveness of the oral exams. We especially focus on comparisons in student and instructor reactions to oral versus standard written exams.

This paper is organized as follows. Section 2 provides background on written versus oral exams in computer science, and discusses related work. Section 3 discusses our implementation of oral examinations into an introductory computer science course, our design goals, and includes sample questions from the exam. In section 4, we discuss preliminary comparisons of student and instructor feedback for the oral and written exam styles. Section 5 concludes.

2 BACKGROUND AND RELATED WORK

Prior work suggests that oral exams have both advantages and disadvantages (when compared with written exams). On the positive side, Asklund and Bendix [1] and Gharibyan [5] note that oral exams allow the instructor to more directly assess student levels of understanding, and provide prompt feedback on student performance. Further, Roecker [13] finds that oral exams can lead to improved student performance, allow freedom in students' problem-solving approaches, and can even expose flaws in course instruction. However, on the negative side, oral exams can be very time-consuming for larger class sizes [13], and give rise to concerns about reliability and fairness (since each student will, inevitably, have a slightly different exam) [6, 12].

Oral exams have historically been common outside of the United States (especially in the German education system [11]). However, written exams are by far the more popular style of exam in computer science courses within the United States. In fact, oral exams are uncommon in most disciplines at US universities [6, 10], outside of foreign languages [3, 7]. In computer science, oral exams are extremely rare outside of evaluating Ph.D. candidates [14]. Further, studies and reflections on their effectiveness are limited in the literature. Oral exams have been studied in other STEM areas, including physics [15], biology [8], and mathematics [9]. Prior work in computer science, however, is limited to older reports [4] and those offering general recommendations for implementation of oral exams (rather than comparisons between oral and written exams in the same course) [2, 5]. For example, Gharibyan [5] assesses the pros and cons of oral versus written exams, and offers advice on how to administer an oral exam for a theoretical computer science course. Gharibyan also describes written exam disadvantages, including the drastic effects of time pressure, the lack of flexibility (e.g., question wording becomes all-important), and the possibility of cascading minor errors. Other related work by Asklund and Bendix [1, 2] describes a design for group oral evaluations in a software configuration management course.

In the following sections, we extend this prior work in two key ways. First, we present an oral exam design used in introductory (rather than upper-division) computer science courses. Second, we present preliminary data comparing our implementation to a parallel offering of the same course with traditional written exams.

3 APPROACH

The first course in the Xavier University computer science sequence is a broad introduction to the field (traditionally called a CS0 course). For the course offerings/sections referenced in this paper, approximately 50% of topics are directly taught via programming in Python (traditionally associated with a CS1 course), 25% of topics involve no programming whatsoever, and the remaining 25% (related to algorithmic complexity) use programming in a smaller support role. Thus, throughout this paper, we use the term "CS 0/1 course."

Most prior offerings of this course have given a final, cumulative, written exam. In this section, we detail our design for a new final, cumulative, *oral* exam.

3.1 Exam Format and Goals

Both the paper and oral exams that we evaluate in this paper primarily cover four high-level topics from the semester:

- Sorting Algorithms; implementation is not expected
- Data Representation; specifically, binary and hexadecimal representations for positive and negative integers
- Algorithms and Complexity; specifically, analyzing complexity for an algorithm that the student designs
- Abstraction; specifically, its purpose and relevance to various course topics

Of course, the types of questions differ greatly between the two styles of exam. In designing the oral exam, we targeted the following core goals.

Flexibility Students should be given as much freedom as possible in demonstrating their knowledge of each topic. Of course, this freedom must be balanced with ensuring that all students receive substantively the same examination.

Consistency The exam should be a similar experience for each student. This is a major challenge for an exam based on free-form conversation.

Equity The exam should be equally accessible to students of various backgrounds. This is a continual challenge in any exam design, whether oral or written.

Overall, the exam proceeds like a *conversation*, wherein the instructor prompts for a particular topic, allows the student to demonstrate knowledge, and asks questions along the way. Students are graded on their ability to say or cover specific aspects of the topic; for example, a student might be expected to recognize that computers represent numbers in *binary* (and note that this is an alternative to *decimal*). Students earn points by satisfying the expected prompts, which the examiner has listed as check-boxes. In general, we did not expect a student to satisfy all check-boxes in order to earn full points; we thus had more boxes than the maximum possible exam score. This is especially important for open-ended questions, where the student is not expected to mechanically list every detail of an entire topic area. For example, in the binary number example, the examiner might have a check-box for "Mention binary numbers" and another for "Mention alternative to decimal."

Of course, not all students will naturally be driven to the expected sub-topics, and students may not always know how to proceed with

We covered "complexity" these past few weeks. What can you tell me about complexity?

- [ ] How long an algorithm will take
- [ ] It is based on input size
- [ ] Operations are counted (rather than time)

(a) Free topic discussion

How might we represent a decimal number as a binary number?

- [ ] Correct description given

(Could you give me an example?)

- [ ] Example partially correct
- [ ] Example completely correct

(Can we represent negative numbers too?)

- [ ] Basics of how negatives can be represented
- [ ] Some specific representation (sign-magnitude, 2s complement)

(Can you give me an example of how to represent a negative number?)

- [ ] Example partially correct
- [ ] Example completely correct

(b) Discussion of an example

What is "functional" abstraction?

- [ ] Understanding of abstraction, generally
- [ ] Definitional / Understanding of functional abstraction

(No answer: Functional abstraction is giving a name to a series of operations)

Why is functional abstraction useful?

- [ ] Use of name, abstraction of implementation details

(How about in Python? How about outside of coding?)

- [ ] Example from Python
- [ ] Example outside Python

(c) Discussion on the purpose of a topic

Figure 1: Sample Oral Exam Questions and Rubric

a question. The examiner needs some way to help orient the student, or give a subtle "hint." In our design, the exam rubric contains pre-scripted prompts. This helps ensure consistency for each student, i.e., that no particular student receives more extensive guidance. One could consider entirely restricting examiners to the prescribed prompts; in our case, we allowed ourselves, as instructors, some flexibility to stray from prompts when the conversation involved question clarity (rather than hints).

3.2 Question Design

Next, we shift our attention to the types of questions that appear on our exam. Loosely, we used three high-level types of questions. Each of the question parts in figure 1 demonstrates a question type, and is a (sometimes simplified) version of an actual question from our CS0/1 exam. Of course, our actual exam consisted of more and longer questions.

The first type of question initiates a free-form conversation about an important course topic. These questions are of the general form "tell me about X." An example is shown in figure 1a. Note that the student is expected to hit three important aspects of complexity learned in class: that it relates to how long an algorithm will take to run (we cover only time complexity), that it is an equation over input size, and that we count operations rather than seconds. Pre-scripted prompts would be given in parentheses (), so no hints are given here. There are, of course, follow-ups to this initial discussion (including having the student analyze a complete algorithm), but it is clear that this specific question part is worth up to 3 points.

The second type of question is of the general form "please give me an example of X." Figure 1b shows an example. Here, the initial question simply asks *how* to represent decimal numbers in a computer; however, the instructor has numerous prompts to use, all directed at getting the student up to the board, working examples.

The final type of question targets the core of many breadth-based introductory courses: the "why" questions. These questions are of the general form "explain why X is good/better/useful." Figure 1c shows an example. Here, we see a case where a pre-scripted instructor prompt gives the answer to a previous question (in this case: the definition of functional abstraction) so that the student is able to continue, even if they were unable to give a satisfactory answer to the initial question.

3.3 Running the Exam

Our goal was for oral exams to last approximately 20 minutes for an average student. To allow for struggling students, each student

Exam Type	Total Students	Respondents	Course Grade Difficulty Scaled: −5 to +5	Relative Exam Difficulty Scaled: −3 to +3	Study Hours
Written	24	5	−1.00	−0.40	4.00
Oral	53	17	−1.00	−0.44	6.13

Table 1: Survey Result Data

scheduled a 30-minute slot to take the exam; we cut off the discussion at that time. To help students manage time, we estimated maximum times for each topic (varying from 7–10 minutes) and warned students if they neared this threshold. We also provided the four exam topics (listed in section 3.1) at the start of the exam, and allowed students to reorder them (though no one did so). Students with special accommodations (e.g., due to documented disabilities) received extra time (e.g., 1.5× or 2×) just as on written exams.

Scheduling oral exams can be a challenge. In our case, students were given a random 2-hour time window in which their registration survey for final exam times "opened" (much like class registration at many universities). Slots included morning, daytime, and evening times, to accommodate students with varying commitments (e.g., work, family, class, etc.), based on the instructor's availability during the week of final exams.

As instructors, we need to determine when a student has satisfied a check-box in the rubric. This can be challenging when students give multiple, contradictory responses. Whenever we encountered this situation, we explicitly pointed it out to students, and forced them to take a stand. This is also slightly complicated by scripted prompts, which might help a muddled response get back on track. We planned for this inevitability: check-boxes that appear *before* the prompt in the rubric are generally "lost" once the hint is given. Figure 1c is a perfect example of this: once the student is given the definition of functional abstraction, they can no longer earn points for the initial "What is functional abstraction?" question.

Because of the "on-the-fly" check-box grading, we were able to give students a reasonable estimate for their grade immediately following the exam. We first asked students what grade *they* thought they earned, and then quickly counted check-boxes to determine their actual estimated grade. (The grade was an estimate to leave flexibility for scaling, though we did not need to do so.)

4 FEEDBACK AND RESULTS

We performed a preliminary evaluation of our oral exam design from section 3. To do so, we investigated two offerings of our CS0/1 course, taught in the Fall 2017 and Spring 2018 semesters. Both offerings were taught by the same instructor (the author of this paper), and covered the same material in the same order. In Fall, the final exam was given as an oral exam; in Spring, the final exam was written. The Fall semester had two sections totaling 53 students; the Spring semester had one section of 24 students.

Our assessment is in three parts. We begin by examining results from a survey sent to students following the completion of both semesters. First, section 4.1 investigates quantitative measures from the survey, focusing on perceptions of exam difficulty. Second, section 4.2 analyzes free-form survey comments, with an emphasis on student reactions to oral exams (from the Fall semester). Lastly, section 4.3 provides our own reflections, and a break-down of instructor time for the two styles of exam.

The results in this section are not statistically significant, due to the small sample size and low survey response rate (noted in section 4.1). Nevertheless, we find the results encouraging, and glean positive indicators from student comments and our reflections.

4.1 Survey

To begin, we examined quantitative data from student surveys. Some of this data is summarized in table 1. Here we see the total number of students who took each type of exam, and the number of students who responded to our post-semester survey. Clearly, response rates were low (e.g., 21% for Spring 2018 students who took the written exam); as noted above, the results cannot be considered statistically significant. We instead view them as preliminary indicators worthy of further study.

The next column, "Course Grade Difficulty," asked students to rate the difficulty of obtaining their desired course grade (compared with other courses). Specifically, we asked:

> "On a scale from −5 to +5, how difficult was earning the grade you desired (or needed) in this class?"

Negative numbers indicated that our CS0/1 course was easier than other courses, while positive numbers indicated that it was harder. The absolute numbers here are of lesser importance. Our key observation is that final exam style (which is the final thing that students will remember about a course) did not impact student perception of overall course difficulty for our small sample of students.

The column labeled "Relative Exam Difficulty" asked students to rate the difficulty of all exams (not just the final) relative to homework assignments. Specifically, we asked:

> "How relatively difficult were the homeworks and exams?"

Negative numbers indicated that homework was harder, while positive numbers indicated that exams were harder. Note that both groups completed the same set of homeworks (differing only in their final exam style). Once again, we see similar responses regardless of exam style.

We also asked students to rate the difficulty of their final exam relative to the difficulty of prior exams; students could respond that the exam was "Easier than the first two exams, " "Harder than the first two exams," or "About the same." Note that mid-semester exams were written (not oral) in both semesters. Figure 2 shows the results, given as the percentage of respondents who selected each option. The vast majority of students indicated that their final exam was either easier or about the same difficulty as the prior exams, with oral exams being marked as "easier" slightly more often. Of course, with our small sample size, these results may not be entirely representative of all students. Nevertheless, we consider this result encouraging, especially when coupled with the data on

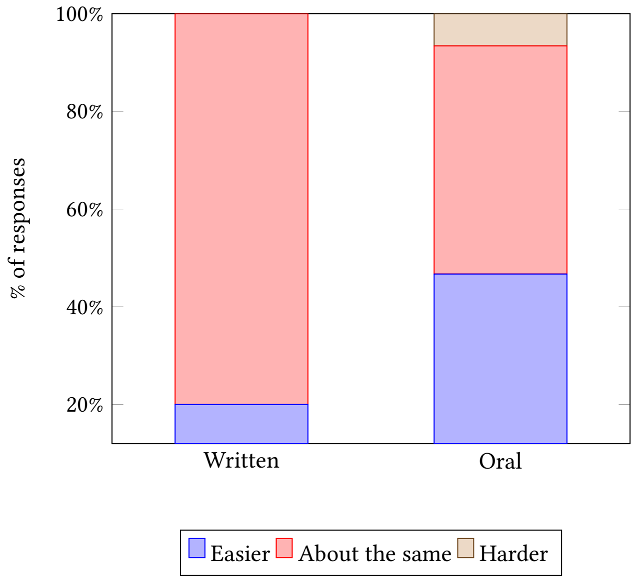

Figure 2: Relative difficulty of final vs. mid-semester exams

course difficulty and exams relative to homework. Our goal was not to make oral exams easier or harder than their written counterparts; rather, we hoped oral exams might spur deeper engagement with (and evaluation of) course material.

Finally, returning to the last column of table 1, we asked students to estimate the total number of hours they studied for the final exam (column "Study Hours"). Students who had an oral exam self-reported studying over 50% more than those who took another written exam. We again take this as a positive indicator. One feasible explanation is that students felt that they required a deeper understanding in order to have a *conversation* about a course topic.

4.2 Student Feedback

To better understand student reactions to oral versus written exams, we also asked students to provide free-form comments with their survey. Some common themes emerged. Students found the oral exam unique, and enjoyed the conversational style.

> "It was unlike any final I have ever taken before [. . .] I think that it really showed if you learned the material or not."

> "I liked the oral exam [. . .] it allows students to talk out their thinking process which I believe to be helpful."

> "I liked how exams were done as reflections of the examples we did in class."

> "It kind of made me feel like I was the teacher, which was pretty fun!"

The fact that the instructor could prompt students, who were given the freedom to explain and talk through answers, was also positive.

> "I liked the not having to write. Also being able to personally explain answers."

> "I liked that [the instructor] helped you come to an answer if you were close. When he dropped key words it helped me remember the topic better and give a better answer."

Surprisingly, and contrary to some prior research [7, 8], two students explicitly noted feeling more comfortable and/or experiencing less anxiety in the oral exam setting.

> "I liked the oral exam more so than the paper exam because I felt more relaxed and was having a conversation that highlighted what I had learned from the class rather than an extremely rigid exam where I tend to second guess myself."

> "It was more forgiving. I could take my time. I also enjoyed the immediate feedback."

Of course, not all comments were positive. In contrast to the above quotes, one student explicitly noted:

> "I did not like the oral exam as it made me nervous"

Written exam comments were less insightful, likely because the format was so familiar to students. This comfort, though, should not be dismissed: it can be both an advantage (in reducing anxiety) and a disadvantage (students who are simply "good test takers").

> "I liked that the exam was very similarly laid out to previous exams so we were comfortable with it."

> "The exams were well presented & straightforward."

4.3 Instructor Impacts

As instructors, we noted both positive and negative aspects to our oral exam design. In preparing the exam, we found that we could focus on question content, and spent much less time worrying about question wording, since we could always clarify during the exam. Instead, we shifted our focus to the rubric, which (per section 3) had to be written ahead of time. This was a major challenge, though we suspect that our skill would improve with further practice, and as questions are re-used and refined.

We were also pleasantly surprised by some students who had fared poorly on prior written exams. We suspect that encouraging students to use examples to explain their answers was a major benefit. Multiple times during the exam, we witnessed a student struggle initially with a question, but, through giving a free-form example, recover from an earlier misstep. These are situations where, on a written exam, the student may have been unable to make progress.

Our class size (53 students) was large by oral exam standards, and we were initially concerned about our ability to schedule all students within our final exam week. However, scheduling all students in limited time slots was not an issue, despite the large number of students and single instructor. In one case, we had to make special arrangements with a student who had a late registration slot; the student was able to swap exam times with a classmate.

We also tracked instructor time commitment for the Fall and Spring semesters; figure 3 shows the results. The plot shows time broken down into pre-exam preparation, exam delivery (proctoring for the written exam, or individual meetings for the oral exam), and post-exam grading (for oral exams, this essentially consisted of verifying and entering grades). While we did track work time, the times shown are approximate: most preparation and grading time overlapped with other unrelated duties.

For direct comparison, we scaled the oral exam delivery time so as to match the (smaller) number of students who took the written exam. This was feasible because each student was given a standard, fixed amount of time (30 minutes) for the oral exam. Naturally, since

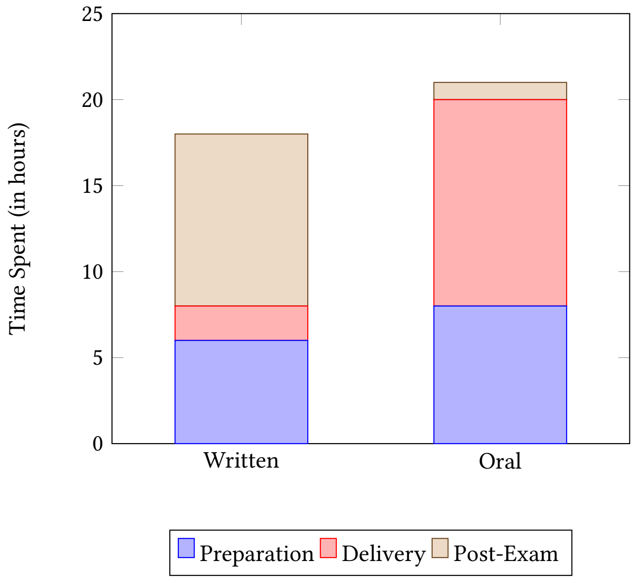

Figure 3: Instructor time spent on the exam (scaled to a 24 student class size)

we have only one semester of data for each exam type, we cannot compare their scalability. (For example, while adding one student to the oral exam adds exactly 30 minutes, it would be reasonable to assume that one additional written exam student would add significantly less post-exam grading time.) In all cases, we excluded additional time given to students with disabilities.

For our class sizes, our results are largely consistent with Roecker [13], who notes that overall time commitment for oral exams does not significantly exceed that of paper exams, but the time instead shifts from post-exam grading to time spent directly examining students. Overall, the oral exam required about 21 hours, while the written exam required 18. Preparation time was a significant factor, requiring 8 hours over the 6 required for the written exam. As noted previously, designing the check-box rubric was the primary time contributor prior to exam delivery.

4.4 Discussion and Practical Implications

Overall, we were pleased with our implementation of oral exams. Student and instructor reactions were positive, and we genuinely felt that students demonstrated a deeper level of engagement with the material.

We felt confident in our assessment of students' knowledge. As previously noted, we were able to directly interact with students to avoid question wording issues, and provide scripted hints so that students were not blocked from fully demonstrating the depth of their understanding. Rubric scores seemed to correspond with our perception of a student's performance during the exam. The only major exception was that the difference between B and C students often felt artificial (though this is often true of written exams as well). We attribute this, partially, to the rigidity of the check-box rubric; some method of adding more flexibility for the instructor would be welcome. Balancing this with our stated goals of equity and consistency, however, poses an interesting challenge.

On the positive side, our check-box rubric allowed the instructor to remain focused on discussion during the exam. It also gave a clear-cut argument for grade assignments; though, in our case, no students challenged the grading afterward. Ensuring equity in rubric design is an ever-present difficulty. Involving more instructors in rubric design may help to decrease bias (though this struggle is certainly not unique to oral exams).

Another challenge involves students discovering exam questions prior to their time slot. While we have no evidence that this occurred, we also had no way to prevent students from discussing the exam between sessions. In fact, our exams were given over the course of 5 days, so we consider this to be a major risk. In smaller courses, the danger could be mitigated by scheduling all exams in sequence, but this was not feasible for us (with 53 students).

Instructor time commitment, though not wholly unreasonable, was significant. For comparisons in section 4.3, we calculated as if both sections had 24 students. In actuality, the oral exam group of 54 students required 27 hours of dedicated examination time (one student was allowed double time). This was a burden for a single instructor, who, fortunately, had few other commitments during finals week. Regularly scaling to class sizes larger than 20 or 30 students appears challenging. With sufficiently-detailed rubrics, and well-scripted prompts, using multiple instructors (or TAs) may be an option.

5 CONCLUSION

Through our design, implementation, and evaluation, we find positive signs of effectiveness for oral exams in an introductory computer science course. Both qualitative and quantitative student reactions were largely positive. Students enjoyed the unique conversational style, and indicated that the oral exam was of comparable difficulty to a written exam, though they spent more time preparing. While instructor time commitments did increase slightly, we found ourselves better able to assess student knowledge at exam time.

Perhaps most importantly, though, our experiences and findings support the need for more research into oral exams in undergraduate computer science curricula. Scaling to larger class sizes while continuing to keep instructor time commitments reasonable and maintain consistency across students are open challenges. Further rigorous assessment of student reactions, student performance and grade impacts, and instructor attitudes would also be valuable, and would help pave the way for ever-improving assessment in an ever-changing discipline.

ACKNOWLEDGMENTS

The author would like to thank Dr. Kathleen Timmerman and Dr. Michael Goldweber for their advice and assistance with this paper and oral exam designs. He would also like to thank the members of the Xavier University *Center for Teaching Excellence* for their support. As previously noted, this work was undertaken while the author was a faculty member at Xavier University. Opinions, findings, conclusions, or recommendations expressed herein do not necessarily reflect the views of Xavier University or the author's present affiliated university.

REFERENCES

[1] Ulf Asklund and Lars Bendix. 2003. Oral vs. written evaluation of students. *LTHs 1: Pedagogisk inspirationskonferens, Lund, Sweden* (2003).

[2] Ulf Asklund and Lars Bendix. 2003. A Software Configuration Management Course. In *Software Configuration Management, ICSE Workshops SCM 2001 and SCM 2003 Toronto, Canada, May 14-15, 2001 and Portland, OR, USA, May 9-10, 2003. Selected Papers (Lecture Notes in Computer Science)*, Bernhard Westfechtel and André van der Hoek (Eds.), Vol. 2649. Springer, 245–258. https://doi.org/10.1007/3-540-39195-9_18

[3] Barbara Gross Davis. 2009. *Tools for teaching.* John Wiley & Sons.

[4] Ian Gent, Bill Johnston, and Patrick Prosser. 1999. Thinking on your feet in undergraduate computer science: A constructivist approach to developing and assessing critical thinking. *Teaching in Higher Education* 4, 4 (1999), 511–522.

[5] Hasmik Gharibyan. 2005. Assessing students' knowledge: oral exams vs. written tests. In *Proceedings of the 10th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education, ITiCSE 2005, Caparica, Portugal, June 27-29, 2005*, José C. Cunha, William M. Fleischman, Viera K. Proulx, and João Lourenço (Eds.). ACM, 143–147. https://doi.org/10.1145/1067445.1067487

[6] M Hairer. approx 2007. Oral examinations/presentations. *Mathematics Research Centre, University of Warwick, UK* (approx 2007). http://homepages.warwick.ac.uk/~masbm/PCAPP/Projects/Oral_exams.pdf

[7] Elaine Hewitt and Jean Stephenson. 2012. Foreign language anxiety and oral exam performance: A replication of Phillips's MLJ study. *The Modern Language Journal* 96, 2 (2012), 170–189.

[8] Mark Huxham, Fiona Campbell, and Jenny Westwood. 2012. Oral versus written assessments: a test of student performance and attitudes. *Assessment & Evaluation in Higher Education* 37, 1 (2012), 125–136.

[9] Paola Iannone and A Simpson. 2012. Oral assessment in mathematics: implementation and outcomes. *Teaching Mathematics and Its Applications: International Journal of the IMA* 31, 4 (2012), 179–190.

[10] Gordon Joughin. 2010. *A short guide to oral assessment.* Leeds Met Press in association with University of Wollongong.

[11] Barbara M Kehm. 2001. Oral examinations at German universities. *Assessment in Education: Principles, Policy & Practice* 8, 1 (2001), 25–31.

[12] Muhammed Ashraf Memon, Gordon Rowland Joughin, and Breda Memon. 2010. Oral assessment and postgraduate medical examinations: establishing conditions for validity, reliability and fairness. *Advances in health sciences education* 15, 2 (2010), 277–289.

[13] Lee Roecker. 2007. Using oral examination as a technique to assess student understanding and teaching effectiveness. *Journal of Chemical Education* 84, 10 (2007), 1663.

[14] Kathryn E. Sanders and Robert McCartney. 2003. Program assessment tools in computer science: a report from the trenches. In *Proceedings of the 34th SIGCSE Technical Symposium on Computer Science Education, 2003, Reno, Nevada, USA, February 19-23, 2003*, Scott Grissom, Deborah Knox, Daniel T. Joyce, and Wanda Dann (Eds.). ACM, 31–35. https://doi.org/10.1145/611892.611926

[15] Yitong Zhao. 2018. Impact of Oral Exams on a Thermodynamics Course Performance. In *ASEE Zone IV Conference.* ASEE Conferences, Boulder, Colorado. https://peer.asee.org/29617.