Oral Exams: A More Meaningful Assessment of Students’ Understanding

Allison S. Theobold

Department of Statistics, California Polytechnic University, San Luis Obispo, CA

ABSTRACT

Compared to their written counterpart, oral assessments provide a wealth of information about student understanding. Instead of deciphering a static response, oral assessments provide instructors the opportunity to probe student explanations, obtaining a more complete picture of their understanding. Moreover, students explaining their conceptual reasoning is advocated in the 2016 GAISE guidelines. Additionally, oral exams allow for flexibility in how students can explain their thinking, potentially helping build students’ identities as statistical thinkers and speakers. Despite the facilities these assessments provide, oral assessments are rarely used in the statistics classroom. In this article I describe the important considerations to be attended to when implementing oral exams in the classroom, my experiences facilitating oral exams in my statistics courses, and some lessons I learned along the way. Supplementary materials for this article are available online.

“A p-value is a value showing the probability. It is the probability that you get a value that is as or more extreme than your observed statistic, just by chance or by natural variance of your dataset.”

A response such as this is likely familiar to many instructors of introductory and intermediate statistics courses. When presented as a static, written response on an exam instructors are forced to discern the student’s true understanding of a $p$-value. The instructor may be left wondering what the student means by “just by chance,” or if the student understands that this probability assumes the null hypothesis is true. Oral exams address these questions by allowing instructors to probe students’ understanding through clarifying questions, gaining a more complete picture of each student’s understanding.

1. Motivation

The 2016 Guidelines for Assessment and Instruction in Statistics Education (GAISE) emphasizes the importance of students’ conceptual understanding (ASA GAISE College Group 2016). However, the way in which big picture concepts are often taught in introductory and intermediate statistics courses is through a set of “recipes,” where students are taught how to apply each recipe to a given problem. This process potentially results in an abundance of students leaving their statistics course without truly understanding what the contents of each of these recipes mean. Rather than expecting a regurgitated recipe, asking a student to explain what a concept means necessitates a deeper understanding of the concept (Huxham, Campbell, and Westwood 2010; Iannone, Czichowsky, and Ruf 2020).

A student’s ability to articulate their statistical thinking could be thought of as the “apex” of their understanding. Asking students to explain concepts ties together four of the six facets of understanding (Wiggins and McTighe 2005)—explanation, interpretation, application, and perspective. Furthermore, the GAISE guidelines (2016) recommend assessments where students are asked to explain their reasoning behind key concepts. Oral assessments allow for students to articulate their thinking and provide a large number of benefits (Huxham, Campbell, and Westwood 2010).

First, oral exams aid in developing the oral communication skills expected of statistics majors and students from other scientific disciplines (ASA Undergraduate Guidelines Workgroup 2014). Second, oral exams provide a more authentic experience, as students will need to defend their thinking after graduation, but will likely never sit for a written exam. Third, this type of assessment is a powerful way to gauge student understanding and/or prolonged misconceptions, allowing for genuine conversations about students’ understandings of statistical concepts. Finally, oral exams are resistant to plagiarism, as students must articulate their understanding in their own words. Despite the potential benefits of this type of assessment for both students and instructors, it is likely that most undergraduate students in STEM have never encountered oral exams during their education (Goodman 2020, p. 3441; Huxham, Campbell, and Westwood 2010, p. 1; Iannone, Czichowsky, and Ruf 2020, p. 313), as the “standard” assessment diet in collegiate STEM courses consists of written, timed exams.

Due to the students’ lack of exposure to oral assessments, they may feel a great deal of anxiety surrounding this type of assessment. However, educators in biology found that the
anxiety induced by oral exams was not necessarily negative, as it helped students “prepare more thoroughly than a ‘standard’ assessment” (Huxham, Campbell, and Westwood 2010, p. 8) and many of these students stated that, despite this anxiety, they preferred oral exams to written tests. Mathematics educators administering oral exams found that students realized they could not succeed by simply memorizing or reworking old problems (Iannone, Czichowsky, and Ruf 2020), and that these exams encouraged “a focus on understanding” (Iannone and Simpson 2015, p. 971).

2. Considerations

Now that I’ve set the stage for the importance of oral exams, let me discuss some of the finer details of implementing this assessment practice in the classroom. Oral examinations can be characterized in six dimensions: content, interaction, authenticity, structure, examiners, and orality (Joughin 1998, p. 368). Both oral presentations and oral exams in STEM typically have a “knowledge and understanding” focus as the content type, but could also address problem solving ability or inter/intrapersonal competencies. Where oral presentations have a “presentation” interaction, oral exams use a “dialogue” (viva voice) interaction. The “authenticity” of the exam refers to “the extent to which assessment replicates the context of professional practice or ‘real life’” (Joughin 1998, p. 371). A “contextualized” exam is conducted in the context of the practice, and a “decontextualized” exam focuses on the ideas abstracted from their context. Oral exams can take on a “closed” or “open” format, depending on how the conversation is permitted to flow. In the case of an instructor-lead exam, the assessment takes an “authority-based” examiner structure. Finally, an oral exam given in substitution for a written exam has a “purely oral” orality, which differs from an exam given as a secondary component.

Each of these dimensions should be considered when planning for an oral exam. However, these are not the sole tasks necessary when preparing to give an oral exam. After deciding on the dimensions of the exam, the instructor must consider four additional tasks: (i) provide students with practice before the exam, (ii) decide the time commitment you are willing to make, (iii) determine how students should prepare for the exam, and (iv) resolve how the exams will be scored and what feedback students will be given. I provide more extensive thoughts on each of these tasks below.

1. If you intend to have students articulate their understanding in an oral exam, it is imperative to provide them with practice communicating their understanding during their learning process. This type of practice could take a variety of forms, such as recorded videos, group problem solving, discussions, or mock oral exams.

2. Next, you will need to determine how much time you are willing to dedicate to administering these exams. Although more time intensive than written exams, even short oral exams offer a substantial benefit for both students and teachers. It should be emphasized that unlike their written counterpart, following the administration of oral exams, the instructor is left with little to no grading.

3. You will then need to decide how students should prepare for the exam. This requires for you to determine if students will be provided with questions ahead of time or if they will need to think “on the fly.” Alternatively, students could be asked to reason through their response to a question on a written exam.

4. The final step is determining how the exams will be scored and what feedback students will receive. Due to the amount of time spent in the exam space, it is of critical importance for you to have a rubric from which you can grade during the exams. You also need to determine what feedback students will receive following the completion of their exam. Two options are to only provide the rubric scores students earned for each question, or to supplement these scores with justification for why they earned these scores. Providing justification for the students’ scores requires additional time, but may be worthwhile if it reduces or eliminates students asking for clarification regarding why they were awarded a given score.

Next, I will address how I navigated each of these considerations when administering oral exams in my courses.

3. Implementation

During the Fall 2020 and Winter 2021 quarters, I gave oral midterms and finals to around 60 students in a second-semester applied regression course. These exams were the sole summative assessment method used for the course. The oral exams were situated in the six dimensions of oral assessments as follows: they focused on knowledge and understanding (content), dialogue with the instructor (interaction), contextualization (authenticity), open structure (structure), authority-based evaluation (examiners), and pure orality (orality).

Throughout my course, students engaged in concept-driven discussion posts where they were encouraged to post their ideas, as well as weekly “group collaborations” where they would gather with their peers and work through a set of applications for that week’s content. For the group discussions, students were asked to post their initial thoughts to a prompt exploring the details of the week’s concepts, and then exchange ideas with other members of their group. Students were given full credit if they posted a carefully articulated idea that allowed members of their group to engage with their idea. At the end of the week I would offer feedback on both the students’ ideas and the group discussion. I would provide individual feedback to students regarding the concepts they demonstrated a robust understanding of and where they had room to grow.

Group collaborations were weekly opportunities for 3–4 students to meet in a Zoom room and reason through a set of applications of that week’s content. This avenue was intended to give students a “low consequence” space to voice their ideas, so students were given full credit for participating in the conversation in whatever way felt natural to them. In the feedback for the collaboration, I would state the concepts each student demonstrated an understanding of and where they had room for growth. Each problem from the collaboration was tied to a specific key concept for the unit, allowing me to connect these collaborations directly to the topics of the oral exams.
Each of these avenues provided students with scaffolded ways to progress through the expression of their thinking prior to the oral exam.

The goal of the oral assessments was for students to articulate their understanding of the larger concepts included in the course. As a result, I distilled the central concepts into five to six questions per “unit,” where each exam covered two units. After developing the questions, I considered how I would score student responses. I settled on using a mastery-based rubric, as it aligned with how students’ homework assignments were marked, and reiterated the course’s focus on a growth mindset (Dweck 2006). Mastery-based grading (also known as standards-based grading) “shifts the focus of grades from a weighted average of scores earned […] to a measure of mastery of individual learning targets related to the content of the course” (Owens 2015). The mastery-based rubric assigned scores 0–4 (integer values) to reflect the student’s understanding of the concepts and the degree to which they made mistakes during their explanations. These scores are roughly correlated with a letter-grade system (4 ∼ A, 3 ∼ B), but have a different emphasis. I also decided to provide students with justification of why they earned each score following their exam. This was done at the end of each day’s exams and amounted to approximately one-hour of additional work; however, I never had to address any student questions regarding their exam score.

Deciding how to administer the exams was the next task. For each exam (midterm, final) I chose to dedicate 10 minutes per student, for a total of approximately 10 hours of exams, spread out over one week. Based on the scope of the questions, I determined that during the 10 minutes it would be reasonable for students to explain their reasoning on two questions. Thus, students were given approximately 5 minutes per question. During the “extra” time after a student had provided their response, I asked clarifying or probing questions, to obtain a more complete picture of their understanding.

A chief concern I had was how student anxiety could impede the speed at which students could articulate their responses. Moreover, as we all know, not every student is quick to respond to questions; in fact, many students take their time to carefully construct a response. Because of these concerns, I chose to provide students with the exam questions and the scoring rubric the Friday before exam week. I informed students that one question from each question bank would be randomly selected for them to answer, and they would be informed of their questions when they entered the exam room. The scoring rubric qualitatively described the expectations for their ability to articulate the central concepts of each question, as well as how letter grades would be assigned based on rubric scores.

4. Reflections

I would proclaim that these oral exams went quite well, for a variety of reasons! First, immediately after their exam was finished, a large percentage of students reflected that the exam “wasn’t as bad as I thought it would be.” Second, in the mid-quarter evaluations many students stated that they appreciated the “reflection” they were required to do when studying for the exam. Additionally, I was pleased by my ability to discern students’ understanding during the exam. If a student made an incorrect statement, I was able to ask questions to probe their understanding. For example, for the student whose response is highlighted, I was able to ask what she meant when she said “just by chance.” I then followed her response with another probing question asking how a $p$-value is calculated, to see if she understood that we assume the null hypothesis is true in the calculation of a $p$-value. Both of these questions left me with a better understanding of her knowledge than a written exam could have permitted.

The flexibility of how students could explain their understanding allowed for some students who may have earned mediocre scores on a written exam to flourish. I had students who had earned C marks on their homework provide masterful explanations of concepts, something from every educator’s dreams. In addition, I appreciated my ability to applaud each student for the understanding they demonstrated during the assessment. Students who struggled during the exam could have easily stepped away distraught, but in my feedback I could remind students of the understanding they conveyed. This feedback provided me with the ability to focus student attitudes regarding assessments on demonstrating their growth as a student rather than on a single summative grade. Last, but certainly not least, after administering the exams the only grading remaining was providing students with a justification for their scores. So, after a week exams, I was able to step away from my computer for a relaxing weekend.

5. Lessons Learned

The facilitation of the exams went quite smoothly, but I did learn a few things along the way that I hope can help others interested in using this type of assessment in their classroom. My first lesson came in how I allowed students to sign-up for exam times. I used a Google Sheet for exam slots, which allowed students to edit the worksheet. However, there were students who neglected to sign-up until mid-way through the exam week, while others would change their previous times without notice for new times later in the week. For these reasons, I would recommend setting a time by which students need to sign-up or they may be denied the opportunity to take the exam, and removing edit access after that time.

Second, if you plan to administer exams online, I would suggest using an application that allows for you to have a waiting room, so that no one “walks-in” on another student’s exam. Additionally, I would recommend an application that allows for you to have an “exam room,” so you do not have to generate meetings for every student. I used my personal meeting room in Zoom, with the waiting room turned on. As an added benefit, Zoom allows for you to record your meetings in the Cloud, so you are able to revisit an exam if needed.

I used the four-point grading rubric I developed to guide how grades for the exams would be assigned. I chose for one question to be weighted more than the other, to reflect the central concepts for that portion of the course. From this weighting I then determined which scores were associated with different letter grades. For example, on the midterm exam the questions were scored such that a 4 on the first question and a 3 on the
second question (4, 3) would earn an A. Alternatively, a (3, 4) would earn a B+. This grading scheme had two downsides. First, I did not account for the possibility that students would score more than a one point difference between their questions, but many did. I then had to scramble to decide what letter grade a (1, 4) should earn. Second, there were some students for whom the discrete scale didn’t accurately capture their understanding, which resulted in some 3.5 and 2.5 scores.

Despite the discrete nature of this grading system, I did find that the distribution of scores was fairly similar to the distribution of scores from a “traditional” written exam. There were notable exceptions, where students who had earned mediocre marks on their written homework assignments were able to demonstrate a more sophisticated understanding of the concepts. These students reflected that flexibility of oral communication and their ability to connect the concepts to examples from their own discipline helped their understanding to flourish. Additionally, in the final exam, students as a whole demonstrated marked improvement in their ability to articulate their thinking. Despite the more difficult topics covered in the second half of the course, many students displayed a more advanced ability to articulate their own thinking and situate their reasoning in a context familiar to them.

Finally, although it was rare, I did have an incident where a student was so overcome with anxiety that they were unable to articulate any of their thoughts. Luckily I had thought about how I would handle this type of situation and was able to calm the student down; however, this would have been difficult with no prior planning. Thus, I would strongly encourage instructors to consider how they can support students both before and during the exam, so the exam space is less anxiety ridden.

6. Equity

Oral exams provide a valuable experience for both students and instructors, but are not devoid from issues of equity. In the age of online instruction, an oral exam over Zoom assumes that students have sufficient internet access to prevent any hiccups occurring during the exam. I had more than one student choose to take their exam with their cell phone, for fear of their internet cutting out. Furthermore, like classroom spaces, there is an implicit power structure in oral exams, whether online or in person. Although it is likely that every student feels a power differential with their professor, this differential potentially differs by the identity of the student and professor. I believe that my students potentially felt less intimidated during their exam because I am a woman, not a man. However, I also believe that students from underrepresented groups potentially felt a greater power differential in their exam because I am White. Because these power dynamics have the ability to dramatically impact student performance in an oral exam, educators ought to consider what actions they can take to make every student feel comfortable sharing their thinking.

As with other types of assessments, oral exams will favor some learning styles over others. However, for students with dyslexia (Waterfield and West 2006) and non-native English speakers (Ramella 2019), oral exams provide a more equitable space for students to share their statistical thinking. Additionally, the roles confidence and identity have on students’ success in STEM disciplines is well documented (Dweck 2006; Holmegaard, Madsen, and Ulriksen 2014). Yet, speech is a key way in which students form their pedagogical identities (Barnett 2007). Thus, oral assessments have the possibility to cultivate students’ statistical identities and increase persistence within the discipline.

7. Conclusion

I hope this article has piqued your interest in using oral exams in your classroom. Oral exams excel not only in their ability to differentiate student understanding, but also in the freedom of expression they permit. Although they are potentially more time intensive than a written exam, the wealth of information they provide instructors is unparalleled. Furthermore, I believe oral exams have the ability to increase student self-efficacy and sense of belonging. I plan to continue my use of oral exams into the future and I hope you’ll join me!

ORCID

Allison S. Theobold © http://orcid.org/0000-0002-2635-6895

References

ASA GAISE College Group. (2016), Guidelines for Assessment and Instruction in Statistics Education College Report 2016, Alexandria, VA: American Statistical Association.

ASA Undergraduate Guidelines Workgroup. (2014). 2014 Curriculum Guidelines for Undergraduate Programs in Statistical Science, Alexandria, VA: American Statistical Association.

Barnett, R. (2007), A Will to Learn: Being a Student in an Age of Uncertainty, Maidenhead, UK: McGraw-Hill/Open University Press.

Dweck, C. S. (2006), Mindset: The New Psychology of Success, New York: Random House.

Goodman, A. L. (2020), "Can Group Oral Exams and Team Assignments Help Create a Supportive Student Community in a Biochemistry Course for Nonmajors?" Journal of Chemical Education, 97:3441–3445.

Holmegaard, H. T., Madsen, L. M., and Ulriksen, L. (2014), "To Choose or Not to Choose Science: Constructions of Desirable Identities Among Young People Considering a Stem Higher Education Programme," International Journal of Science Education, 36, 186–215.

Huxham, M., Campbell, F., and Westwood, J. (2010), "Oral Versus Written Assessments: A Test of Student Performance and Attitudes," Assessment & Evaluation in Higher Education, 35, 1–12.

Iannone, P., Czichowsky, C., and Ruf, J. (2020), "The Impact of High Stakes Oral Performance Assessment on Students’ Approaches to Learning: A Case Study," Educational Studies in Mathematics, 103, 313–337.

Iannone, P., and Simpson, A. (2015), "Students’ Views of Oral Performance Assessment in Mathematics: Straddling the ‘Assessment of’ and ‘Assessment for’ Learning Divide," Assessment & Evaluation in Higher Education, 40, 971–987.

Joughin, G. (1998), "Dimensions of Oral Assessment," Assessment & Evaluation in Higher Education, 23, 367–378.

Owens, K. (2015), “A Beginner’s Guide to Standards Based Grading,” AMS Blogs, AMS (American Mathematical Society).

Ramella, D. (2019), "Oral Exams: A Deeply Neglected Tool for Formative Assessment in Chemistry," in Active Learning in General Chemistry: Specific Interventions, eds. M. Blaser, T. Clark, L. Lamont, and J. Stewart, Vol. 1340, Chapter 6, pp. 79–89. Washington, DC: American Chemical Society.

Waterfield, J., and West, B. (2006), Inclusive Assessment in Higher Education: A Resource for Change. Plymouth, United Kingdom: University of Plymouth.

Wiggins, G. P., and McTighe, J. (2005), Understanding by Design, Alexandria, VA: Association for Supervision and Curriculum Development.