Administrative Considerations Pertaining to the Use of Creative Methods of Student Assessment: A Theoretically Grounded Reflection from a Master of Biostatistics Program

Jesse D Troy[1,*], Megan L Neely[1], Gina-Maria Pomann[1], Steven C Grambow[1] & Gregory P Samsa[1]

[1]Duke University, USA

*Correspondence: Duke University, USA. E-mail: jesse.troy@duke.edu

Received: May25, 2022 Accepted: June 19, 2022 Online Published: July 13, 2022

doi:10.5430/jct.v11n5p105 URL: https://doi.org/10.5430/jct.v11n5p105

Abstract

Student evaluation is a key consideration for educational program administrators because program success depends on students' ability to demonstrate successful development of core competencies. Student evaluations must therefore be aligned with learning objectives and overall program goals. Graduate level educational programs typically incorporate course-level and program-level evaluations, e.g., a final examination in a single course vs. a qualifying examination that assesses knowledge gained from several courses. While there is often considerable attention given to the structure of these evaluations at the program level, the format is typically left to the instructor's discretion. We argue in this article that there are administrative advantages to encouraging instructors to adopt creative forms of assessment that extend beyond the typical concerns related to program structure. Specifically, we argue that advantages can be gained in terms of increasing student engagement, adding real world context to student evaluations, maintaining positive program culture, and reducing the opportunity for cheating. We present two examples of creative assessments implemented in a 2-year Master of Biostatistics program, along with a discussion of three key questions administrators should consider as they work with instructors to develop innovative assessment methods: (1) what changes to make; (2) in what order to make those changes; and (3) how to consult with instructors about making those changes.

Keywords: educational program administration, student evaluation, cheating

1. Introduction

One of the primary considerations of educational program administrators is whether coursework, curricula, and student evaluations are appropriately aligned with program goals and objectives (McNeil, 2011). For example, instructors evaluate students with respect to course-specific learning objectives that align with program goals. There are also program level student evaluations which are broader in scope and target knowledge synthesis across courses. A common example of program level student evaluation is the use of a qualifying examination in a graduate degree program (Barnett et al., 2017). Successful completion of the qualifying examination indicates that students have sufficient mastery of their coursework to move on to the next step in the educational process, which is typically a thesis, dissertation, or other capstone activity. Evaluation of graduate students on their capstone activity is typically done by a committee and the structure of this evaluation is also aligned with program objectives (Powell & Green, 2007). Less often considered at the administrative level is the format of student evaluation and whether these formats are also aligned with program objectives (Buker & Niklason, 2019).

We refer here to course level evaluations as "micro" evaluations and the program level evaluations such as qualifying examinations and thesis work as "macro" evaluations. We have written previously about reconsidering the format of macro level evaluations in light of program objectives (G. Samsa, 2021). In this article we focus on micro level evaluations. Specifically, we argue that program administrators can gain advantages by encouraging instructors to use creative assessment methods for micro evaluations. These advantages are: 1) increasing student engagement; 2) adding real world context to student evaluations; 3) maintaining positive program culture; and 4) reducing the opportunity for cheating. Importantly, we highlight that this is not an argument in favor of administrative oversight of course instructors' assessment procedures—this would inhibit academic freedom—but that our approach is part of an overall objective to support student success and encourage continued innovation throughout the curriculum.

In considering whether and how to upgrade methods of student evaluation, administrators should consider (1) what changes to make; (2) in what order to make those changes; and (3) how to consult with instructors about making those changes. In this editorial we share our experience with two non-traditional approaches to a course final examination as examples: (1) an interview; and (2) a student recorded video. We also share the results of a student survey taken after completing the interview examination. Our context is a 2-year professional master's degree program in biostatistics. While most of our students are adult learners who enter the workforce immediately after graduation, the essential elements of our argument translate to other academic environments.

2. Context

Approximately 30% of graduates from our 2-year master of biostatistics (MB) program proceed directly to doctoral work after graduation and the remainder enter the workforce. In their first year all students take a statistical theory course sequence, a data analysis sequence, a programming sequence, and a practice of biostatistics sequence (Neely et al., 2022). Overall, the curriculum is designed around the "ABCs of biostatistics": namely, Analytical skills, Biological knowledge, and Communication, with the emphasis on communication being one of its distinguishing features (Troy et al., 2021, 2022).

3. Administrative Imperatives for Creative Assessment Methods

There is an extensive literature (see Durning et al., 2016 for a systematic review) describing the pedagogic deficiencies of traditional methods of student assessment, which in our context would translate into timed, closed-book, exams that focus on recall of facts and the application of those facts to relatively straightforward problems. The traditional approach has its place. For example, we use a single such question at the start of a flipped classroom session as an incentive to engage in the assigned pre-class preparation and as an entrée into the discussion of the day's content. That said, we argue that the traditional approach is not appropriate in a final examination that is intended to assess the student's ability to apply course content within a substantial context. For example, in hiring staff biostatisticians, we have found that students often leave graduate programs with sufficient technical knowledge but require additional training in leadership and communication (Pomann et al., 2020; Troy et al., 2022). The creative assessment methods we propose here also aim to improve training and retention of these critical skills. Finally, problems with traditional examinations become particularly acute when classes are offered online because of the difficulty in effectively proctoring and the associated opportunity for cheating. We believe that the responsibility of program administration regarding cheating includes:

- Treating students equitably, including the majority of students who choose not to cheat.
- Facilitate learning that encourages students to master course material, which reduces incentives to cheat.
- Maintaining a positive program culture, which is undermined when considerable measures are taken around proctoring and similar anti-cheating behaviors.

Therefore, evolving beyond traditional in-person closed-book, fact-recall-based examinations to more effective evaluations can be considered well within the scope of responsibility for educational program administrators.

4. Example Final Examinations

The examples we highlight are final examinations from two classes that use a flipped classroom format. The assessments do not require a flipped classroom, but are more consistent with this approach to teaching than a traditional didactic course. The first example is from a SAS programming course (G.P. Samsa, 2020) and the second example is based on a course in causal inference (Hernan & Robins, 2020). In both courses, the final exams test students' ability to perform functions that are essential to their professional success after graduation. The exam from the SAS programming course mimics a coding interview such as students might face when seeking a job. The exam from the causal inference course mimics an interaction between a biostatistician and a clinician scientist, which is a common scenario that students will face when working as biostatisticians.

4.1 SAS Programming Course

The final exam in the SAS course is a simulated coding interview. Grading for the SAS course is tiered as follows, with the final exam being the capstone piece that enables students to earn an A. Briefly, students can earn a B based on the successful completion of group-based programming assignments. The instructor comments on these assignments but does not assign grades. Completion of a project illustrating the ability to learn an element of SAS programming not covered in class earns a B+. These assignments are preparation for the final exam coding interview because they require students to practice explaining algorithms and program code to each other and the instructor. Students are also prepared for the final examination by:

- Previewing example interview questions and reading the grading rubric for the assignment.
- Watching a mock coding interview between two faculty members in which one faculty member plays the role of student and the other plays the role of the instructor giving the interview.
- Taking a practice coding interview with the instructor prior to the real final exam. If the students score well on the final exam they can receive an A for the course.

The coding interviews typically take 30-45 minutes, with the first few questions being consistent across all students. These initial questions provide the instructor with information about the student's level of experience and sophistication as a programmer and can assist in the choice of follow-up questions. Whenever possible, subsequent questions focus on the student's project, providing them the opportunity to demonstrate their knowledge of SAS using an example from their own work (see Appendix 1 for details).

4.2 Causal Inference Course

The grading for the causal inference course is based upon homework, a group project, and a final examination. The final examination requires students to select 6 questions out of a set of 12, and then record a 10–12-minute video with their answers. Students are instructed to explain their answers in non-technical terms, as if they were speaking with a clinician investigator. Instructors retain the option of interviewing students about unclear aspects of their answers, which in practice translates into an opportunity to improve their grade for those students whose videos do not reach the desired standard. Students are allowed to research their answers using multiple sources, including the internet, other students and instructors, but are required to correctly cite those sources.

Students are prepared for the final examination by:

- The in-class assignments, which focus on the core constructs addressed by the 12 questions in the final examination.
- Group projects, a characteristic of which is that, although students are welcome to divide up the work, including the in-class presentations which follow, instructors can ask questions of any student about any portion of the content, thus encouraging everyone to be familiar with all of the content.
- Exercises which practice explaining core constructs in plain language with examples from the instructor modeling the desired behavior in a similar context.
- The ability to send preliminary versions of the videos to the instructors for comment.

5. Key Question 1: What Changes to Make?

From the perspective of pedagogy, interview-based examinations have the advantage of clarity of communication. This clarity is bidirectional: the student can clarify what they believe the instructor is asking, and the instructor can probe to clarify the student's answers. Interview-based examinations also provide the opportunity for tailoring. For example, if a knowledge deficit is identified, the instructor can probe to better understand the nature and scope of that deficit, which in turn can support remediation. The instructor can also choose to tailor the level of difficulty of follow-up questions to the student's initial answers. This can, for example, help distinguish between A+ and A, a distinction which will only be relevant to a subset of students.

From the perspective of pedagogy, video-based examinations have similar advantages to interview-based examinations. Namely, students have the opportunity to provide more detail, providing instructors with more information to support their assessment of performance. Because questions are known in advance, students can give well-researched and thoughtful answers. Further, the ability to perform such background research helps make the format of the assessment consistent with how students will use their skills after graduation. By reviewing their videos before submission, the student will have access to the same information as the instructor, which in some circumstances will suggest that the videos should be further edited before their final submission. An additional advantage of both types of examination within our program is their emphasis on effective communication, a key programmatic focus and point of differentiation, which is enhanced by placing that communication within a realistic context (i.e., an interview, a lay-friendly explanation) (Gregory P. Samsa, 2018; Troy et al., 2022).

The primary disadvantages of both types of examination pertains to consistency and fairness. In this regard, videos have the advantage of creating a permanent record of student responses to questions of their own choosing, thus allowing others to review the assigned grade in case of challenge. By design, interviews are not intended to be identical, although consistency can be achieved by asking a standardized set of questions to all students. Although interviews could be recorded for documentation purposes in case of challenge, we have not chosen to do so, in large part in response to student preferences. Of note, we have given interview-based final examinations in multiple classes and never had the resulting grade challenged, one explanation being that the interview process brings deficits to the attention of not only the instructor but the student as well. Another explanation is that, in contrast to some employment interviews, our simulated coding interviews are designed to be low stress (although, it must be admitted, are not always perceived as such). Techniques used to reduce stress include encouragement, prompting in case a question proves difficult, and simply moving on to another question if students become overly flustered. In practice, we have found these techniques especially helpful for students with different learning styles to produce their best work. The interview procedure also facilitates easy adaptations for students who require accommodations for learning disabilities. For example, a student could research a question and return with a partially written answer in addition to their oral exam. Additionally, exam time is not limited for this evaluation approach.

6. Key Question 2: What Order to Make the Changes?

Both of our use cases pertain to final examinations within the context of flipped classrooms, the design of which can prove daunting to instructors. However, the assessments we describe can be used in a traditional classroom format as well, so that transitioning to a flipped classroom is not a prerequisite. In addition, the transition away from traditional assessment methods need not be absolute. Indeed, both courses where the non-traditional format for the final examination was applied also used traditional assessments such as homework, in-class exercises, and projects prior to the final examination. Similarly, the transition need not be absolute at the level of the examination. For example, perhaps a single examination question could require the students to make a video, with the remainder of the examination being more traditional. Indeed, this was the approach we took to the most recent iteration of our qualifying examination and were sufficiently pleased with the results (G. Samsa, 2021).

In that spirit, a natural place to start would be to add the option of an interview follow-up to all examinations, regardless of format. This allows instructors to query students about responses which are problematic in terms of content -- for example, responses which suggest that the student might have misunderstood the question, responses which utilized sufficiently non-standard notation as to make it difficult to understand what was intended, etc. This also serves as a reminder to students who are tempted to cheat that simply finding the "correct" answer might be insufficient, that they might be required to explain and defend that answer, which is an incentive to understand their answers and thus learn the material. From a perspective of maintaining positive culture within a program, we believe that this approach is preferable to considerable efforts around proctoring, which suggest that students are not to be trusted and negatively impact those many students who choose not to cheat.

7. Key Question 3: How to Discuss Changes with Instructors?

Potential concerns with these changes include unfamiliarity with non-traditional assessment methods, worries over consistency and fairness, and the possibility of extra work. Unfamiliarity can be mitigated by providing show-and-tell from early adopters. Worries about consistency and fairness can be addressed by circulating an evaluation rubric to students, and by sharing feedback from students who have already taken the exam. For example, we surveyed the students in our SAS course after completing the coding interview and found the results overall to be very positive (see Appendix 2). Sharing such survey results with instructors and students can help alleviate concerns about the non-traditional format.

Worries about extra work can be addressed by decoupling the thought of a non-traditional final examination (which might be a modestly sized effort) from the thought of converting to a flipped classroom (which is a very significant effort). For example, when considering a video-based exam, grading time is unlikely to be different from the time to grade short-answer questions or student narratives, although admittedly longer than the time required to grade a multiple-choice examination. However, the literature is consistent about the deficiencies of this latter modality and these citations can be shared with instructors (Durning et al., 2016).

An additional barrier to progress is that some instructors implicitly adopt an assessment philosophy which could be termed "complete recall". For example, by developing a test which covers detailed material in a limited time, the hope is that the student will study exhaustively to master the entire content of the course backward and forward. This approach uses the examination as an incentive to memorize all course materials covered. This approach is pedagogically unsound because rote memorization does not assess students' ability to synthesize course content to develop a thoughtful and appropriate answer to a question.

Finally, it is likely that students will have many of the same concerns as instructors, especially regarding fairness and additional work required. We believe these concerns can be addressed by adequately preparing students for the new examination format as described above in Section 4.

8. Conclusion

In this article we have described an approach to creative student evaluation that assesses knowledge in a real-world context, supports positive program culture, and reduces the opportunity for cheating. In addition, this approach serves as a first step toward a more expansive approach to creative assessment throughout the curriculum. For example, if micro assessments such as those described here are successfully adopted by instructors, this facilitates greater creativity in macro assessments such as qualifying examinations, primarily because students will have been immersed in a program culture that emphasizes wholistic understanding over fact-based recall. These benefits may also serve to expand the inclusivity of the program. For example, engaging students in live discussion gives instructors the opportunity to help students navigate language barriers, learning disabilities, or other barriers to expression that might give a false impression on a traditional exam that the student lacks understanding of key concepts.

There are limitations of our approach that should be noted. Based on our survey of students in the SAS programming course, the non-traditional examination methods we applied were well-received by our students (see Appendix 2). However, this does not argue for the efficacy of our approach. To assess the efficacy of this innovative approach we plan to evaluate the performance of those same students using targeted questions on our qualifying examination, given after completion of their first-year courses. This will assess long-term retention of knowledge and depth of understanding. We will report on these findings in a subsequent publication.

In summary, we believe it is crucial for educational program administrators to actively pursue the use of both micro and macro level creative assessments of student knowledge. We have successfully implemented two such examples in a professional graduate degree program and have highlighted key questions that program administrators should consider when adopting similar methods of assessment.

References

Barnett, J. V., Harris, R. A., & Mulvany, M. J. (2017). A comparison of best practices for doctoral training in Europe and North America. *FEBS Open Bio, 7*(10), 1444-1452. https://doi.org/10.1002/2211-5463.12305

Buker, M., & Niklason, G. (2019). Curriculum Evaluation & Improvement Model. *The Journal of Health Administration Education*.

Durning, S. J., Dong, T., Ratcliffe, T., Schuwirth, L., Artino, A. R., Boulet, J. R., & Eva, K. (2016). Comparing open-book and closed-book examinations: A systematic review. In *Academic Medicine*. https://doi.org/10.1097/ACM.0000000000000977

Hernan, M., & Robins, J. (2020). *Causal Inference: What If*. Chapman & Hall/CRC. Retrieved from https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/

McNeil, R. C. (2011). A program evaluation model: Using Bloom's Taxonomy to identify outcome indicators in outcomes- based program evaluations. *MPAEA Journal of Adult Education*, *40*(2), 24-29.

Neely, M. L., Troy, J. D., Gschwind, G., Pomann, G.-M., Grambow, S. C., & Samsa, G. P. (2022). Preorientation Curriculum: An Approach for Preparing Students with Heterogenous Backgrounds for Training in a Master of Biostatistics Program. *Journal of Curriculum and Teaching*, *In Press*.

Pomann, G.-M., Boulware, L. E., Cayetano, S. M., Desai, M., Enders, F. T., Gallis, J. A., Gelfond, J., Grambow, S. C., Hanlon, A. L., Hendrix, A., Kulkarni, P., Lapidus, J., Lee, H.-J., Mahnken, J. D., McKeel, J. P., Moen, R., Oster, R. A., Peskoe, S., Samsa, G., … Thomas, S. M. (2020). Methods for training collaborative biostatisticians. *Journal of Clinical and Translational Science, 5*(1), e26. https://doi.org/10.1017/cts.2020.518

Powell, S., & Green, H. (Eds.). (2007). *The Doctorate Worldwide*. McGraw Hill.

Samsa, G. (2021). Evolution of a Qualifying Examination from a Timed Closed-Book Format to an Open-Book

Collaborative Take-Home Format: A Case Study and Commentary. *Journal of Curriculum and Teaching*, *10*(1), 47-55. https://doi.org/10.5430/jct.v10n1p47

Samsa, G.P. (2020). Using Coding Interviews as an Organizational and Evaluative Framework for a Graduate Course in Programming. *Journal of Curriculum and Teaching*, *9*(3), 107-140.

Samsa, Gregory P. (2018). A Day in the Professional Life of a Collaborative Biostatistician Deconstructed: Implications for Curriculum Design. *Journal of Curriculum and Teaching*, *7*(1), 20-31 https://doi.org/10.5430/jct.v7n1p20

Troy, J. D., Granek, J., Samsa, G. P., Pomann, G.-M., Updike, S., Grambow, S. C., & Neely, M. L. (2022). A Course in Biology and Communication Skills for Master of Biostatistics Students. *Journal of Curriculum and Teaching*, *11*(4), 120-138. https://doi.org/10.5430/jct.v11n4p120

Troy, J. D., Neely, M. L., Grambow, S. C., & Samsa, G. P. (2021). The Biomedical Research Pyramid: A Model for the Practice of Biostatistics. *Journal of Curriculum and Teaching*, *10*(1), 10-17 https://doi.org/10.5430/jct.v10n1p10

Appendix 1: Coding Interview Questions from the SAS Programming Course Final Examination

The instructor will start by asking each student the 3 questions listed in Section 9.1. The instructor will select additional questions from Sections 9.2 and 9.3 based on the topic of each student's programming project that they completed prior to taking the final exam.

1. Initial Questions to Ask All Students

- Describe the structure of base SAS. How does this structure differ from that of R?
- What sorts of tasks would be more efficient to program in R than in SAS? Why?
- What do the terms literate programming and reproducible programming mean to you? Illustrate some techniques for performing literate and reproducible programming using SAS.

2. Questions About the Structure of SAS

- Briefly describe PROC SQL. How does the structure of PROC SQL differ from the structure of base SAS? What are the advantages of SQL?
- Briefly describe PROC IML. How does the structure of PROC IML differ from the structure of base SAS? What are the advantages of IML?

3. Data Management and Programming

- Describe the difference between list input and column input. When might you prefer column input?
- Suppose that an investigator has generated a cleaned, well-structured csv file with data for you to analyze. How would you move the information from this file into a SAS dataset?
- Describe various ways you can create a SAS dataset. (For example, one way is from input data, with DATA/INPUT, as per the previous question.)
- How could you use the FORMAT statement to embellish your output? For example, how could you print a variable with a dollar sign and 2 decimal places? (If you don't remember the precise syntax pseudocode would suffice.)
- Suppose that sex=0 for females and 1 for males. How could you print female/male instead of 0/1?
- How could you permanently assign female/male to 0/1?
- What is the general logic for performing a character-to-numeric conversion? Write SAS code if you can, pseudocode if you must.
- Within the DATA statement, what does the RETAIN statement do? Why is it needed?
- Describe a counter-based algorithm. What role does the RETAIN statement play within counter-based algorithms?
- Within a DATA statement, what does an ARRAY statement do? Give an example of where an ARRAY statement might be used. How would you test the code for an array statement?

- Describe how a PUT statement can assist in program development and debugging. What is the difference between a PUT statement and a PUT function?
- Within a SAS procedure, what does the BY statement do?
- How can you identify records 2-5 within a SAS dataset?
- What is filtering? Give an example of using filtering within a SAS procedure. Given an example of filtering within a DATA statement.
- Within a DATA statement which includes a BY statement, what do the FIRST.id and LAST.id variables do?
- How can you identify which variables within a dataset have unique ids?
- How could you create FIRST.id and LAST.id from scratch? Write SAS code if you can, pseudocode if you must.
- A dataset has multiple records for each patient, including pain scores. Create a new variable containing the mean pain score for each patient.
- A dataset has multiple records for each patient, including pain scores. Count the number of records, per patient, with PAIN<4. Appropriately account for missing values of PAIN.
- SAS has a tool for creating samples with replacement. The output of that tool is a list of ids and the number of copies of each id. So, id=1 might have copies=2, id=2 might have copies=1, etc. Transform the output of the tool into an actual dataset.
- Describe how to restructure a long-thin dataset into a short-fat dataset. Write SAS code if you can, pseudocode if you must.
- Describe how to restructure a long-thin dataset into a short-fat dataset. Write SAS code if you can, pseudocode if you must.
- Describe how to stack two SAS datasets vertically. Suppose that one dataset contains the same information but with different variable names -- for example, AGE in one dataset and AGE_YEARS in the other. How can you fix the problem?
- Describe how to join two SAS datasets horizontally. How could you filter in the records which are in both parent datasets?
- Assume that SAS doesn't have a procedure which calculates the mean absolute deviation, that is, the sum of |(Y-Ym)| divided by the sample size. Describe how to perform this calculation using a long-thin data structure. Describe how to perform this calculation using a short-fat data structure.
- Explain how SAS processes dates.
- You have multiple records for each patient, denoted by VISIT (VISIT=0,1,2…). Each visit has a calendar date, in character format, named DATE_CHAR. Follow-up time is the amount of time between the current visit date and the date of visit 0. Calculate follow-up time.

9.3 Statistical Analysis Programming

- Many of the examples of statistical analysis programming involve some form of randomization. Describe a general algorithm for randomization, and then illustrate how you can implement this algorithm using SAS.
- Describe the difference between exact and approximate randomization, and illustrate how to do both using SAS.
- Explain the logic behind bootstrapping and describe how to implement bootstrapping using SAS.
- Describe how to validate a simple linear regression model using SAS.
- Explain how power calculations can be performed using SAS.

Appendix 2: Student Survey Administered After SAS Coding Interview

Response to questions 1, 3,4,6, and 7 was obtained using a Likert scale as follows: Not at All, Slightly, Moderately, Very, Extremely. Results are summarized as the proportion responding in the two categories. For example, 18/21 students (86%) responded Very Well or Extremely Well to the question “How well did the coding interview allow you to demonstrate your understanding of SAS?”. Responses are summarized similarly for questions 2 and 5 but the possible responses are different. Possible responses to Question 2 were Extremely Unfair, Somewhat Unfair, Neither

Fair nor Unfair, Somewhat Fair, Extremely Fair. Possible responses for Questions 5 were Extremely Uncomfortable, Somewhat Uncomfortable, Neither Comfortable or Uncomfortable, Somewhat comfortable, Extremely Comfortable.

Question	**n/N (%) Responding In the Top 2 Categories**
1. How well did the coding interview allow you to demonstrate your understanding of SAS?	18/21 (86)
2. As a method of assessment, how fair did you find the coding interview to be?	21/21 (100)
3. How well did the preparatory materials (e.g., sample interview questions, practice interview with the other faculty member) assist you in preparing for the coding interview?	20/21 (95)
4. How confident were you before the coding interview?	8/21 (38)
5. How comfortable were you during the coding interview? (For example, did the instructor help put you at ease?)	21/21 (100)
6. If you encountered difficulty during the interview, how well did the instructor help you get back on track?	21/21 (100)
7. How confident were you after the coding interview?	18/21 (86)

Appendix 3: Final Examination for the Causal Inference Course

Instructions

The final examination will have two steps:

1. You submit a video explaining key concepts pertaining to causal inference (a list of questions follows)
2. You meet with the instructor to clarify and improve your explanation

If the video is satisfactory, then step 2 will not be required.

IMPORTANTLY, the intended audience for this video should be a non-statistical investigator. Thus, for example, an explanation such as "bias means that Pr[Y=1|A=1] - Pr[Y=1|A=0] is different from …" will not be acceptable. Instead, you should use plain English and either a minimal or non-existent amount of notation. We aren't asking you to simply copy formulas and explanations from the book. Most of the material in the book isn't accessible to non-statistician investigators without significant translation, mostly because mathematical notation is not a natural language for this audience.

For example, if you are asked to explain confounding, imagine that the scientific question you are studying involves confounding, and that you pause your conversation with the investigator to provide an operational definition of confounding in terms they can understand. Similarly, if you are asked to explain DAGs, imagine that you are about to create a DAG to discuss with the investigator but that the investigator is unfamiliar with DAGs, and so you pause your conversation to provide an operational definition of what is a DAG and why it can be helpful.

Your video can be as short or as long as you like, but please do keep in mind that, in an actual collaborative interaction, you don't have time to launch into an elaborate discourse. Consider trying to answer each question in about 2 minutes. Feel free to use diagrams or other visual aids in your video if you wish. If you are uncertain of the desired length and detail, please feel free to try recording an answer to one of the questions and submit it to the instructor for their comments.

You are welcome to use outside sources (e.g., other students, instructors, the book, the internet) to clarify your understanding of the content, but the words should be your own. If you are uncertain what this means, it is your responsibility to clarify this with your instructors before submitting the final copy of your videos.

Choose 6 of the following topics to explain to investigators:

1. What is the difference between an RCT and an observational study?
2. What is bias?

3. In an observational study, what must be assumed in order to generate valid inferences about the treatment effect?
4. What is effect modification? What is interaction? How are they the same? How are they different?
5. Aim 1 of a study is to evaluate the efficacy of treatment. Aim 2 is to evaluate whether this efficacy differs by sex. How can the investigator conceptualize these aims in terms of main effects (average causal effects) and interactions (effect modification)?
6. What is a DAG? How will the DAG we're about to draw be useful?
7. When drawing a DAG, why might we be interested in including an unmeasured variable?
8. What is confounding? How do we deal with it?
9. What is selection bias? How do we deal with it?
10. The investigator has heard about propensity scoring. What is it, and when might it be used?
11. We've concluded that a single variable requires being conditioned upon. What does it mean to condition on a variable? How might you do so? (You can go into a bit of technical detail here.)
12. We've concluded that multiple variables require being conditioned upon. How might our analytical approach differ from conditioning on a single variable?