Providing an Oral Examination as an Authentic Assessment in a Large Section, Undergraduate Diversity Class

Dredge Kang, Sara Goico, Sheena Ghanbari, Kathleen C. Bennallack, Taciana Pontes, Dylan H. O'Brien, & Jace Hargis

University of California San Diego

Received 28 July 2018; Accepted 4 December 2018

This study reflects on the development and implementation of mid-term oral examinations in large-scale lecture courses at a large, public research university; specifically, this work examines the implications of oral exams for fostering student engagement and concept-based comprehension in addition to institutional and course commitments to diversity. This research traces the development of an effective method for administering oral midterms and assesses the advantages and challenges of utilizing oral examinations for student assessment by detailing student feedback and TAs' reactions to administering this examination format. Findings reveal that oral examinations provided a chance for students to develop skills through a different means of engaging material and to foster a concept-based learning approach. In a discussion of student and TA reactions, this paper reports a predominantly positive assessment by both groups while noting the challenges and disadvantages of this format.

INTRODUCTION

This study examines the implementation of oral examinations within a large undergraduate diversity classroom. Through a process of course redesign, the idea of an oral assessment was introduced as a means of better aligning with course learning outcomes for the instructor's context. Our research motivations were focused on two main inquiries:

RQI: How has the implementation of an oral examination impacted the student and TA experience in a large undergraduate diversity course?

RQ2: How do reported impacts align with the course learning outcomes?

We have outlined the structure of the course in addition to the process of piloting and integrating oral examinations.

Structuring the course. Initially, the instructor developed a closed-book word-processed written examination in lieu of blue books with the intention of reducing grading time for Teaching Assistants (TAs). TAs were responsible for grading the examinations. After encountering difficulties with classroom technology and numerous violations of academic integrity, the instructor decided to pilot an oral examination in his diversity course. The instructor believed that modifying the grading process would result in an assessment that more directly aligns with the course learning outcomes. When the instructor first piloted the oral examination as a midterm in a large diversity course in Fall 2017, students expressed anxiety towards the new format. In order to address this uneasiness, the instructor provided sample questions and practiced responses during a midterm review, thereby modeling the behavior for success. The instructor also made modifications on the structure of the oral examination when implemented a second time in Winter 2018. In its first iteration, the instructor provided ten questions and asked students to respond to one. The following quarter, the instructor provided ten questions and asked students to respond to two. The intent being that the additional opportunities and time would increase variation in scores. Between the two quarters, the instructor also re-assessed the metrics for grading the oral examination.

In the pilot, the instructor based the score solely on the student response. For the second quarter, he provided 80% of the assessment on completeness and accuracy of the student response and 20% on their confidence responding to the questions. TAs graded the confidence measure based on how much prompting students required and how quickly and certainly students responded.

During the pilot, the instructor had five TAs and the following quarter he had three TAs. One of the TAs from the pilot quarter also was a TA in the second term and took a lead role in informing the other TAs of the process. All seven of the TAs are doctoral students in Anthropology.

The pilot course in Fall 2017 was a general education Diversity and Race course that also was a diversity requirement. The class consisted of approximately 280 students and was relatively evenly distributed across grade levels, with slightly more representation from sophomores. The second course in Winter 2018 was Diversity and Health and consisted of approximately 170 students. Most students were in their third year and taking the course as a major requirement. Both courses were housed in the Department of Anthropology, and the latter course was also cross-listed with the Global Health Program.

At the beginning of each quarter, the instructor provided students with course learning outcomes, which were listed on the course syllabi and discussed in class. While these were specific to the content of each class, both syllabi expected students to complete the course with the ability to: synthesize the anthropological concept of race or health, distinguish and define key terms, and describe and assess key concepts. During Fall 2017, there was an additional objective to have students explain key concepts because the course is an undergraduate general education requirement meant to promote diversity on campus and in society at large. While this was not included as an outcome for Winter 2018, the instructor noted in the course syllabus that memorization would not suffice and active engagement with course content was an explicit expectation.

Setting. Situated in a large, public, research intensive university, this university's mission is to transform the region and create a diverse global society by educating, generating and disseminating knowledge and creative works, and engaging in public service. Our research aligns with our university philosophy and

student goals. More specifically this study directly addresses our institutional initiatives of diversity.

Justification for Study. Previously, the course midterm was structured in the form of writing a letter explaining course concepts to an international student unfamiliar with race and racism in the United States. After having consulted Educational Technology Services, the instructor was given a list of warnings regarding conducting an online Learning Management System Blackboard (BB) examination in the lecture hall. These included: lack of power outlets, limited Wi-Fi bandwidth in the lecture hall, and the possibility that BB might crash during the examination period. To avoid these issues, students were allowed to take the midterm outside the lecture hall, write and save in a word processing program and then copy and submit via BB. Thus, the instructor could not freeze the BB screen for the examination, as students would be using the word processing and might have to address questions to him via an online messenger program. Given these constraints, and the increased possibility of cheating outside the lecture hall, an honor pledge was posted at the beginning of the midterm examination.

Nevertheless, there were numerous suspected violations of academic integrity. One student admitted guilt to her College Dean and was suspended for a quarter as she had prior violations. Another student maintained his innocence and mounted a case to fight the charge, providing numerous documents and testimonies from his roommate and other students. The case was adjudicated by a panel consisting of students and faculty. The student also made in-person pleas to the instructor during office hours to withdraw the case, citing that he, an international student, would not be able to return home for the summer if he was sanctioned. Once a possible violation case is submitted, however, it cannot be withdrawn. Based on appeals, the process took over a year to adjudicate, and the student was ultimately sanctioned with a quarter suspension, which indicates a prior violation had occurred. Additionally, the instructor received student feedback that the written midterm was "too easy to cheat." The move toward an oral assessment was in part a reflection of this experience.

LITERATURE REVIEW

Assessment in higher education. Assessment is the process of gathering information from multiple and diverse sources to understand learner knowledge, skills and dispositions (Huba & Freed, 2000). Assessments can take several different formats and through integrating backwards course design, are intended to be aligned with course learning outcomes (Wiggins & McTighe, 2005). The process of creating effective assessments has been documented in the literature as a challenging process and there are texts to support educators in building assessments (Brown, Bull, & Pendlebury, 2013; Sadler, 2005).

Oral examinations. One means of assessing course learning outcomes is through conducting oral examinations, rather than traditional written examinations. While oral examinations are de rigueur in certain contexts, such as clinical exams or PhD defenses, they are less common in undergraduate education; nearly all the literature addressed herein observes that oral examinations are considered a novelty (despite their antiquity) by college teachers and students alike. Huxham and colleagues (2012) find this rarity surprising, given what they propose as its five major advantages over written exams:

- Development of [students'] oral communication skills, especially in the context of graduation and post-graduation employment (see also Buehler & Schneider, 2009; Burke-Smalley, 2014);
- Authenticity of in-person interaction (as opposed to sitting for a written test, which is unlikely to occur again in their lives post-college, again partially with a view to employment);
- Inclusivity; one study the authors cited (Waterfield & West, 2005) discussed the views of 229 British students with disabilities, who preferred oral over written examinations by a vast margin;
- Ability to focus on "deep understanding and critique" via the possibility of discourse, which is not possible on written examinations; and
- 5. Resistance to plagiarism.

Despite these clear benefits, oral examinations remain scarce and the literature presents some factors as to why this is the case (Buehler & Schneider, 2009; Huxham, Campbell, & Westwood, 2012). The most prevalent concern seems to regard the issue of reliability and objectivity; it seems that written examinations are more likely to be perceived as "objective," at least from the perspective of examiners, and this is related to concerns about possible bias in grading (Huxham, Campbell, & Westwood, 2012).

An additional concern expressed with integrating oral examinations is the time-consuming nature of this type of assessment (Burke-Smalley, 2014; Zoller & Ben Chaim, 1990). In many situations, the time constraint might make oral examinations prohibitive, because rather than all students taking the examination simultaneously during a predetermined period, each student is generally tested alone and sequentially. As was the case in our study, several examiners mentioned the presence of properly trained TAs as the best solution to this problem.

Oral examinations and diverse populations. One noteworthy consideration regarding oral examinations is that students with disabilities prefer oral examinations over written examinations (Waterfield & West, 2005). In fact, some examiners avoid oral examinations explicitly because they are perceived as being for "special populations." Conversely, some examiners were concerned that only the most extraverted, confident students would do well on an oral exam, or that students whose first language was not English or who had been previously educated in another country's system would be at a disadvantage (Roberts et al., 2000). This latter problem is well documented, in fact, Roberts et al (2000) address concerns of significant discrimination in oral examinations for the Royal College of General Practitioners. Of major concern is the concept of "institutional discourse," and that "the everyday competencies and practices... have to be presented in institutional terms through language that reifies and abstracts these practices" (p. 371). This type of discourse is distinct from personal experience discourse and professional discourse. As the authors point out, oral examinations involve a "hybrid discourse," that is, a combination of all three, and this hybridity can pose major problems of evaluation for both examiners and examinees. While Roberts et al. (2000) deals with medical doctors, institutional discourse can also be a stumbling block for undergraduate students of various national, ethnic, class, and racial backgrounds as well (Lubrano, 2010).

This study aims to build upon the research on oral examinations and fill some of the gaps in the literature regarding implementing oral examinations in undergraduate settings. Although there are some studies looking at oral examinations in undergraduate settings, this form of assessment has been predominantly examined in graduate settings. We strive to contribute towards the integration of oral examinations in diverse contexts, given our classroom setting of large undergraduate diversity courses.

METHODS

The current study is observational and not experimental. Students were not followed longitudinally, and all cross-sectional data was anonymous. The project was reviewed by an Institutional Review Board (IRB) and certified as exempt from review under 45 CFR 46.101 (b), category 1: Research conducted in established or commonly accepted educational settings, involving normal educational practices.

Oral examination pilot

Timeline

The oral examination was first piloted during Fall 2017 for the Diversity and Race course that university students take as a diversity requirement. The oral examination was administered by TAs during the time period when students usually attend weekly discussion sections. Students were divided into eight discussion sections of approximately 36 students.

Participants

The participants in the Fall 2017 term were 279 undergraduate students. The students were between the ages of 18-27 with mixed cultural identity differences. The number of students by year in college included 68 First Year; 84 Second Year; 58 Third Year; and 69 Fourth Year. While we do not have demographic data specific to this course for the general ethnicity make up of undergraduate students on this campus is comprised of: 37.6% Asian, 19.5% International, 19.1% White, 17.8% Chicano/Latino, 2.5% African American, .4% American Indian, .2% Pacific Islander, and 2.8% Undeclared.

Procedure

Pre-exam Organization. Prior to the oral examinations, the instructor provided students with written instructions regarding expectations for learning outcomes, the format of the midterm, suggested activities for studying (i.e., quizzing other students rather than repeatedly rereading lecture notes), and content that would be tested (Appendix A). One week before the oral examination, students received a list of ten short response questions that provided the question and a percentage breakdown for the grading of the question. This was the same set of questions used during the oral examination. Students practiced the oral examination, including the content and the structure, during the classroom lecture before examinations. The TAs organized an online document so that students could sign up for a time slot during section meetings. Prior to the examination, the instructor and the TAs met to discuss the organization of the testing period. The instructor was concerned about inter-rater reliability and TAs expressed concern that oral examinations left no documentation to corroborate grading. A decision was made to have two TAs present during a student's exam, as well as to have two groups of TAs grading for each section to ensure that the exams were finished within the examination hour. The TAs shared among themselves a Google Document with the schedule of students for their sections, wherein the TAs would update exam grades after each exam. This allowed the TAs to coordinate which students were being tested and note if there were major discrepancies in grading between the groups of TAs.

Proctoring the Oral Examination. On the day of the exam, one student would enter the exam room at a time and select a slip of paper containing one question from an envelope. (Incidentally, this format met the disability requirements of a student who required an individual room for exams). The student had 60 seconds to answer the question and another 30 seconds for the TAs to prompt them for any portion of the question they might have missed. Each slip of paper contained a grade breakdown and a unique number in order to keep track of the question that the student answered. The TAs had a list of all the questions with their correct responses. Grading for the oral exam was immediate. After logging the grades on Google Document, the TAs also wrote them on the slip of paper that each student took out from the envelope, giving the slip back to the student for recordkeeping.

Changes to the Procedure. During the first day of the exam, it became clear to the TAs that testing a student took longer than 90 seconds, which was causing the exams to run behind schedule. This was a concern because some students had classes directly after the examination period, and thus could not stay after their allotted time. The reason for the delay was primarily due to student nervousness. Some students required a moment to relax before answering, and many students took longer than 60 seconds to respond due to their anxiety. It also became clear that it was not necessary to have two TAs in the exam room. Grading was remarkably straightforward due to the fact that the TAs had the answers in front of them and were able to rely on cues, such as body language and clarity of responses, to determine whether an individual knew the complete answer. Based on these factors, during the first day of testing, the TAs made the decision to split into three testing rooms to speed up the testing process. During the second day of testing, TAs proctored the exams individually, and there were no issues with the exams running behind schedule.

Oral examination, version two

Timeline

The second version of the oral examination was given during Winter 2018. Since there were no discussion sections for this class, approximately 11 students who could not attend the exam times conducted the oral examination over Skype.

Participants

The participants in the Winter 2018 quarter were 167 undergraduate students and the number of student evaluations of teaching (SETs) submitted was 144. The Winter 2018 term included zero first year students; 10 second year; 81 third year; and 76 fourth year. It should be noted that this course had a very high dropout rate of approximately 20%. The average grade during the drop-add period was a D before grades were normalized. Thus, those students who remained in the class may be biased toward self-selected high achievers.

Procedure

Pre-examination Organization. Similar to the Fall 2017 oral examination, students were given the ten possible exam questions one week before the test. However, in order to help ease anxiety, an-

other practice run was given during class in which students were called to simulate the oral examination by picking a question from an envelope and attempting to answer it. When some of the students had difficulties answering a question, the instructor asked for the remainder of the class to participate. Any concerns about the concepts referenced in the set of questions were clarified by the instructor. Students were reminded that 20% of the assessment was based upon their levels of confidence in responding to the questions.

As there were no discussion sections for this class, the exams were scheduled on two different days of the week outside of regular lecture hours. The TAs created an online sign-up sheet via Google Sheets. A second sign-up sheet was circulated through Google Documents for students who could not attend the midterm at the times provided due to classes and work obligations. TAs offered those students additional times via Skype.

Proctoring the Oral Examination. Examinations were proctored as evolved in the first quarter, except in the number of questions that students answered. In contrast to the pilot, the students in this class selected two slips from the envelope and were given approximately two minutes to answer each question. If unable to answer one of the randomly chosen questions from the envelope, students had the option of "passing" the question by answering an alternative question, although they would lose some points. Students were told about this option in the in-class midterm review. Despite the fact that a few students initially struggled to answer the original questions they pulled from the envelope, only one student utilized the option of answering the alternative question.

DATA COLLECTION

In order to begin to understand the impact of shifting from written to oral examinations, a midterm student feedback survey was administered for both undergraduate courses in addition to the SET at the conclusion of both courses. Additionally, a survey to capture the TA experience with oral examinations was subsequently administered. In order to establish the legitimacy of the questions and answers, the researchers applied multiple coding strategies to the answers and triangulated the data. Open codes were established utilizing a grounded theory approach derived directly from the language used by student and TA participants (Strauss & Corbin, 1997). Open codes were reviewed and then integrated into thematic categories which are presented in the subsequent Data Analysis and Results (Saldaña, 2013). After completing the oral examination, students were asked to give midterm feedback in an in-class multiple choice evaluation using an internet poll. The questions evaluated the midterm goal, format, preparation, and student anxiety taking the exam. For Fall 2017 there was approximately an 80% student response rate and for Winter 2018 there was approximately a 70% response rate. At the end of the course the students responded to the university's student evaluation of teaching (SET). The SET included multiple choice and free response questions. Within the SET, there were two multiple choice questions that were relevant to class assignments and one free response question that asked students to give their opinion on the class assignments. Prior to students filling out their SET at the end of the course, the instructor asked students to speak to specific elements of the course that touch upon the oral examination. Students were given instructions to: "Please comment on your experience with the oral examination,

whether it improved your ability to explain concepts to others, and whether you prefer this to an in-class written midterm." For Fall 2017 there was approximately a 75% student response rate and in Winter 2018 there was approximately an 85% response rate. The TA midterm survey was administered at the end of each course and asked the TAs to evaluate their experiences administering the oral exam using open-ended free response questions. There was a 71% response rate from the TAs. Accuracy of results was enabled by a high response rate across all the surveys (Appendix B).

RESULTS

Student Experience.

Fall 2017: Diversity and Race

Two hundred and twelve of the 279 students responded to the SET that students use to assess university courses and instructors each term. The SET includes a section to evaluate "Exams, Quizzes, and Papers" that had a total of 64 responses, 53 of which evaluated the assignments in general or specifically evaluated the oral examination. Responses were eliminated if they simply listed the assignments without providing an evaluation of the assignments (i.e., "There are several short quiz questions assigned during each lecture about the previously assigned reading. There was an oral midterm and a final video project"). Eighteen of the comments were about the assignments generally and 36 specifically evaluated the oral midterm using vocabulary such as "oral exam," the "midterm," or the "exam." Responses that used the vocabulary "exams" (in the plural), were included in the midterm specific evaluations only if the response differentiated between exams and quizzes, since the oral midterm was the only exam given during the entirety of the course. Outside of this section there were three comments made about the oral midterm in the general review of the class and another one comment in the review of the instructor, which were included with the midterm specific responses. Responses could fall into more than one category.

Comments about assignments in general fell into one of two categories:

- Assignments were easy (i.e., "All the assignments were really easy"), or
- 2. Assignments were fair and representative of course material (i.e., "I thought the exams were a good way to test our knowledge").

Only one comment was not included because the subject of the statement was ambiguous ("Prepares you well"). The number of evaluations that fell into these categories is in Table 1.

Table 1. Number of responses per category from reviews that generally evaluated the assignments.

Evaluation Category	Number of Evaluations
Easy	12
Fair and representative	6
Ambiguous	1

In the midterm specific evaluations there was more diversity in responses. Categories of responses included:

- Preference [to written exams], which was broken into subcategories: general statement of liking, or specific reasons: helped explain, understand, or retain course materials.
- Ease ("Midterm and final were fairly easy do the readings and go to lecture, you'll do fine on both"),
- 3. Anxiety ("Midterm is a small interview, which can be intimidating or nerve-racking for some students"),
- Representativeness ("All of the exams/quizzes were reflective of the material we learned!"),
- Exam Procedures ("Oral midterm was not run well.
 Many students were late to other classes or waited more than 30 minutes"), and finally,
- Difficulty [than written exams] ("Oral midterm exam took more time for me to study compared to the inclass midterm").

Table 2 lists the number of responses per category. From these responses there is an overall positive student response to the oral midterms. There were only eight instances of negative feedback documented and three of these instances were nested within comments that also gave positive feedback about the structure of the oral midterm.

Table 2. Number of responses per category from reviews that specifically evaluated the midterm.

Evaluation Category	Number of Evaluations
Preference	19
Explaining concepts	7
Understanding concepts	5
Retaining material	3
Generally	4
Ease	17
Anxiety	4
Representativeness	2
Procedure	2
Difficulty	2

There was also an oral examination specific evaluation that the instructor gave during the course. Two-hundred sixty students were present during the in-class evaluation, although not all students responded to each of the questions. Students were asked to evaluate whether the midterm met the instructor's goal for the assignment, their preference for the format of the exam, their preparation on the exam, and their anxiety preparing for the exam. As demonstrated by the responses in Figure 1, 97% of the students who responded felt that the oral exam met the goal of helping prepare students to better explain concepts about race and racism to other people. Students also overwhelmingly prefer the oral format, with 91% saying that they would not have preferred responding to the questions in a written format. The majority of the students (80%) felt that the oral format took an equal or less amount of time to prepare for than a written exam, even though over half the responders (53%) did find it more nerve-wracking to prepare for the oral exam than a written one. These results corroborate the findings from the general course

evaluation that students generally prefer the oral over the written format. It is also makes a similar finding that students do feel anxiety using an oral format, even though they overwhelmingly state that the exam effectively achieves the goal of helping them talk about race to other people.

Winter 2018: Diversity and Health

In Winter 2018, 144 of the 167 students responded to the SET. The same coding rules were used across the two quarters. The section "Exams, Quizzes, and Papers" had a total of 71 responses, 61 of which evaluated the assignments in general or specifically evaluated the oral exam. Nineteen of the comments were about the assignments generally and 42 specifically evaluated the oral midterm. Outside of this section there were two comments made about the oral midterm in the general review of the class, which are included with the midterm specific responses. A comparison of the Fall 2017 and Winter 2018 student responses are represented in Table 3.

Table 3. Comparison of Fall 2017 and Winter 2017 student response data				
F	ALL 2017	WINTER 2018		
8. Assignment	s promote learning.	8. Assignments	promote learning.	
4 (1.9%)	Strongly Disagree	2 (1.4%)	Strongly Disagree	
9 (4.4%)	Disagree	3 (2.1%)	Disagree	
23 (11.2%)	Neither A nor D	13 (9.0%)	Neither A nor D	
115 (55.8%)	Agree	66 (45.8%)	Agree	
55 (26.7%)	Strongly Agree	59 (41.0%)	Strongly Agree	
0 (0.0%)	Not Applicable	0 (0.0%)	Not Applicable	
6	No Response	1	No Response	
II. Exams are course m	representative of the aterial.	II. Exams are r	epresentative of the erial.	
I (0.5%)	Strongly Disagree	0 (0.0%)	Strongly Disagree	
I (0.5%)	Disagree	2 (1.4%)	Disagree	
9 (4.4%)	Neither A nor D	6 (4.2%):	Neither A nor D	
100 (48.8%)	Agree	48 (33.3%)	Agree	
90 (43.9%)	Strongly Agree	86 (59.7%)	Strongly Agree	
4 (2.0%)	Not Applicable	2 (1.4%)	Not Applicable	
7	No Response	0:	No Response	

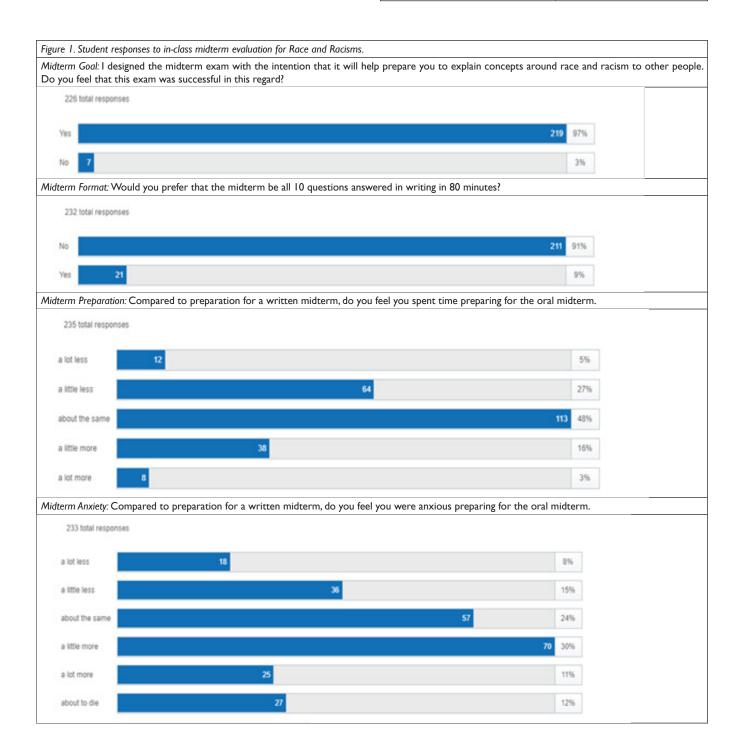
There were 119 students present in class for the midterm specific evaluation. Students were asked to evaluate whether the midterm met the instructor's goal for the assignment, their preference for the format of the exam, their preparation on the exam, and their anxiety preparing for the exam. Student responses are shown in Figure 2.As demonstrated by the graphs, I 00% of the students felt that the oral exam helped them remember key concepts. Students also overwhelmingly prefer the oral format, with 91% saying that they would not have preferred responding to the questions in a written format. The majority of the students (74%) felt that the oral format took an equal or less amount of time to prepare for than a written exam, even though close to half (47%) did find it more nerve-wracking to prepare for the oral exam than a written one. The distribution of assignment and oral midterm evaluations are provided in Table 4 and Table 5, respectively.

Table 4. Number of responses per category from reviews that evaluated the assignments generally.

Evaluation Category	Number of Evaluations
Representativeness	14
Ease	6
Difficulty	I

Table 5. Number of responses per category from reviews that specifically evaluated the midterm.

Evaluation Category	Number of Evaluations
Preference	39
Better for retaining material	13
Understanding concepts	10
Less stressful	8
Generally	8
Representativeness	7
Ease	7
Anxiety	3
Difficulty	l





TA Experience

In addition to influencing student experience. The TAs who administered the oral examination also expressed the benefits and challenges of this process. Thematic trends conveyed by the TAs included more efficient and effective grading, heightened conceptual understanding, and initial student nervousness.

The issue of time was frequently cited by TAs as a benefit of integrating an oral examination. As one TA noted, "Selfishly, for me as a reader, this greatly reduced the amount of time I had to spend on grading." The TAs also noted being able to record their answers on the same day. Two TAs acknowledged that part of why this process was so efficient was that the instructor had shared a clear answer key that enabled them to provide quick and consistent grades. Having this resource increased TA confidence, as one TA states, "we were provided with detailed answer sheets and so there was no concern about knowing the right answers." The TAs expressed that the grading was not only simpler, but also more clear. Another TA explains, "there was less room for ambiguity because I could simply ask the students to elaborate on their answers." Another stated, "those constant discussions with students after written examinations about what they were trying to say, simply isn't a concern in the oral examination because you resolve what the student means in the moment." The fact that students had opportunities to elaborate on their answers meant that they could not claim a lapse of memory afterwards.

TAs also noted that the oral examination improved conceptual understanding. "I think it was a really useful exercise for students, and also spoke to the concept-learning focused pedagogy of the Professor." A different TA adds, "I think the benefit of an oral midterm is found in its ability to encourage creative thinking and synthesis in a conversational format." TA feedback reveals that the oral examination emphasizes students' ability to comprehend and explain course concepts. "For example, there were situations in which it was clear that students had tried to simply 'memorize' the right answers, but I could tell that they did not fully grasp the material (indicated either by their inability to elaborate further or by the fact that they were getting certain basic terms/concepts mixed up)."

While TA responses on administering the oral examination reflected increased efficiency and simplification of the grading process, they also consistently noted that students were initially nervous about the oral examination. For most students this lessened once they took the examination. "Once they sat down and took the exam, they were able to respond fluidly and without difficulty. I believe that those waiting in line felt relieved as others exited and expressed that it had gone well. I gave the students their grades immediately after the exam, and since they all did well, it helped boost the confidence of those waiting." While all seven of the TAs commented on initial nervousness, the one TA, who administered oral examinations in both of the classes noted

that anxiety did contribute to poor performance among a few students.

Her responses included clear strategies that she employed from the first iteration to the second iteration of the oral examination to address students' nervousness. First, she chose to set up the graduate lounge as the office in which the test would be conducted by her and another TA, believing that a more casual setting might alleviate anxiety. Some of the students commented that the room felt "nice" to them, however, there was no attempt in measuring the differing levels of anxiety based upon this criterion. During the second oral examination she had a traditional office room with a desk, but chose not to sit behind the desk, instead sitting to the side of each student being tested. The TA's intent in this was to diminish the hierarchical structure expressed in the placement of the furnishings. In both occasions, the TA also watched her own body language to avoid any expressions of rigidity. During the examination, the TA allowed students to have brief moments to breathe deeply and restructure their thoughts, when needed. These were some of the techniques thought necessary by the TA to mitigate anxiety.

Course Learning Outcomes

To address our research sub-question, student responses were analyzed in respect to the course learning outcomes. Beginning with the student data gathered from the Fall 2017 Diversity and Race course with the following course learning outcomes:

- "distinguish and define key terms such as ethnicity/ race/nationality, ethnocentrism/racism/nationalism, and implicit bias/prejudice/discrimination.
- describe and assess differences in racialization and racial systems among various racial minorities in the US and other global contexts.
- synthesize an anthropological concept of race that attends to biological, socio-cultural, and historical perspectives.
- 4. explain key concepts around race and racism such as the social construction of race, systemic racism, and racial privilege."

The SET did not prompt students if each of these outcomes were achieved, yet their general short answer addressed these aims. The first three learning outcomes were sparingly addressed in student commentary, but there was a great deal of student feedback that directly addressed the final course learning outcome.

The final learning outcome revolved around students' ability to explain course content and the oral examination was created specifically to assess this outcome. There were multiple comments from the SET that reinforced this statement. One student noted, "I really appreciate his efforts in cultivating us students into functional human beings of society. His push to make sure we are able to explain things to others by not just taking a normal written exam really is a good one. I appreciate that because he knows that his class structure is not a traditional one, he makes it not as stressful by making it not hard to get a good grade as long as the students put in the effort." Another student comments directly on the effectiveness of this assessment, "the oral exam was great and surprisingly easy to do I felt I had a better understanding of the information because I was able to explain it to someone else." While there was not unanimous agreement on enjoying the process of oral examination, student responses reveal that the oral examination was a direct measurement of the final course learning outcome. "The oral exam did help increase my ability to explain things to other people. I would not prefer this to a written exam just because of anxiety issues but I guess this could help me in the long run with getting over that," states a student in the Fall 2017 Diversity and Race course.

Unlike the predetermined SET questions, the midterm feed-back did have a question that directly stated the goal of the midterm and asked students if this was achieved. As noted earlier, 97% of students stated that the oral examination was successful in helping them explain concepts around race and racism to other people.

The course learning outcomes for the Diversity and Health course in Winter 2018 are to be able to:

- synthesize an anthropological concept of health that attends to socio-cultural, biological, and historical perspectives;
- **2. distinguish** and define key concepts such as culture, ethnomedicine, health inequality, structural violence, biopolitics, and syndemics; and
- describe, assess, and apply how key concepts relate to various health issues in global contexts.

Like the SET for the previous quarter, the SET for the Winter 2018 course did not include any questions that directly ask students if course learning outcomes are met, but several student responses directly reference the oral examination and general retention of course concepts.

As one student noted, "The oral midterm helped me solidify the concepts learned in the class because we were forced to talk about the concepts and put me in the position to essentially teach the TA about the concepts." While the students did not pinpoint one learning outcome in particular' they speak to the retention of the collective learning outcomes: "to this day, remember the questions, and the answers to the questions. It truly made me learn the concepts and not short-term memorize them." Another student specified, "The oral midterm definitely helped me to retain the key concepts we covered in class. It definitely helped me retain and understand the key term of structural violence."

The students also compared the experience to a traditional midterm and presented examples of how it was more conducive to learning. "I liked the oral midterm because it was a lot less stressful than traditional midterm exams and having to do an oral midterm actually made me learn the material better because there's more effort required of you to actually know the material and examples." Another student echoed, "it [the oral examination] helped myself retain the information much better. Instead of studying to memorize key terms, I had to fully understand the concept, which helped me remember and fully conceptualize the topics."

The instructor also gathered midterm feedback in the Diversity and Health course and one of the questions directly inquired if the oral examination helped students remember key concepts. One hundred percent of students stated that it did achieve this goal and the responses from this survey are illustrated in Figure 2. Overall, the responses show a similar set of responses from the pilot, including preference over a written midterm, similar studying time, heightened anxiety, and improved material retention.

DISCUSSION

In conducting this preliminary study of oral examinations, we agree that this assessment had benefits for both students and TAs, but despite the largely positive findings, heightened testing anxiety seems to be a key feature of oral examinations (Sparfeldt, Sparfeldt, Rost, Baumeister, & Christ, 2013). SETs from both courses described here show that students were initially more anxious about the oral examinations because they were an unknown format. Students also reported spending about the same amount of time studying for the exam compared to a written one. Huxham et al. (2012) found that students were more nervous in oral examinations but that they performed better compared to written ones. They suggest the anxiety could be attributed to a number of factors including expectations that oral tasks require greater understanding or having to socially perform in a professional setting. In our experience, while there was greater anxiety around the oral exam; with preparatory modeling including study suggestions and a review day of practicing the questions before the exam, students felt that the format was generally easier and preferable to written examinations. As we did not experimentally randomize participation in oral versus written examinations, we cannot comment on any potential differences in this area.

The instructor expected that an oral examination would force students to process the information in novel ways. Some students noted in their SETs that the format improved their understanding of concepts and retention of material and this reinforces existing studies. Boe (1996) found that oral exams increased critical thinking and analysis rather than simple retrieval and Badger (2010) showed how oral examinations improved reflective thinking among future teachers. Dicks et al. (2012) showed that oral examinations improved creativity and interaction between faculty and students. Calumet College of St. Joseph has developed a system of oral examinations for core courses to improve oral communication and critical thinking skills (Crist & Robinson, 2015). Critical thinking and reflection are important components of diversity education and align with the instructor's classroom goals as teaching diversity is not simply about knowledge acquisition but opportunities for affective experiences that lead to changes in attitude and worldview. The positive response to the oral examinations was even higher when students were taking an upper-level required course for the major. This may simply reveal a preference for a shorter, though perhaps more intensive and focused assessment.

Much of the oral examination literature focuses on the extra time needed to administer the exams (with the exception of Buehler & Schneider, 2009). However, in our experience, this format was highly praised by TAs for reducing grading times, a primary objective of the instructor in reducing the load for graders and TAs, who typically run two discussion sections throughout the quarter. During the Fall 2017 pilot, there were difficulties in coordination and timing. Students were given time to relax before responding to their question, creating a backlog. However, a second implementation of the exam with longer time windows for students to sign-up demonstrates that the format is feasible to administer by TAs using Google Documents. TAs also preferred this format as it reduced judgments about scoring, as clear rubrics were provided and assessed in a way transparent to students. Although our first iteration initially had two TAs to check for inter-rater reliability, the TAs felt that it was not necessary. Students also appreciated leaving the session with their grade

rather than waiting for grading to occur after turning in a written exam. TAs also noted that student anxiety in the oral format may reduce their performance and that the format may be more difficult for non-native English speakers, particularly those better at written English. This was not a population group in Waterfield and West's 2006 paper on inclusive assessment. We would want to explore this question further at an institution like ours, where approximately 20% of students are international.

It is also possible the learning outcomes from the syllabus primed (Ratcliff & McKoon, 1988) students for the experience they would have. For example, in the Fall 2017 course there was a learning outcome of explaining course concepts and the students stressed liking the oral exam for that reason. In Winter 2018, the learning objective was to understand, and students focused on how it helped them apply and retain course concepts. On the midterm in-class evaluations, the midterm outcomes are different (i.e. first question in Figure 1 and 2). Student responses on SETs reflect this when looking at the reasons that people emphasize for why they preferred the oral midterm. This may also show that students were clear about course learning outcomes and that the oral examination can be effective for steering students towards different learning outcomes.

In the second implementation, anxiety was markedly lower than in the first, even though the examination had effectively doubled. However, as noted above, the class was particularly motivated. This may also reflect a maturation effect as students' comfort with speaking out loud in an academic setting improves over time. Similarly, while the exam increased in difficulty between the two quarters, students were less likely to comment on that, suggesting that it was proportional to their expectations in an upper level course. Attempts to create more variation in grade outcomes by increasing the number of questions, incorporating a confidence score, and allowing students to pass a question did not lead to greater variation. In Fall 2017, 97% of students received an A compared to 95% in Winter 2018.

As part of reflective teaching practices and continuous quality improvement, additional attempts to increase grade variance will be introduced in the future. In a future oral examination, the instructor may ask students to define two key terms and make connections to how they are linked via various course examples. This would require conceptual work that can not necessarily be practiced and memorized in advance.

In summary, addressing our first question, we find that student and TA experiences of an oral examination were extremely positive. SETs from both courses show that students were initially more anxious about the oral examinations because they were an unknown format. This is in line with the literature on oral examinations. Students also reported spending about the same amount of time studying for the exam compared to a written one. However, with modeling, students felt that the format was generally easier and preferable to a written examination. Some students also note that the format improves their application of concepts and retention.

Our sub-question focuses on examination alignment with learning outcomes. Student responses on SETs regarding the alignment with course outcomes are also extremely positive. For example, 97% of students in the Diversity and Race course stated that the oral examination was successful in helping them explain concepts around race and racism to other people, an explicitly stated outcome of the first course examined here.

Thus, at this exploratory level using a combination of SET and qualitative evaluation data, we also find that oral examinations are appropriate assessments to help prepare students for actual discussions in everyday life. Students generally agree that the oral examination facilitates this ability. As survey results were not linked to the oral exam scores, we cannot establish that there was greater learning by using the oral exam format, only the reporting of this by students on surveys and in conversations with TAs and the Professor. Additionally, in order to establish an actual increase in learning, we would have to conduct a randomized control trial comparing oral exam methodologies to a standard written exam or other exam modes. However, without an independent measure over time, we would only know the difference in performance on the exams rather than actual learning. Nevertheless, the positive response from students suggests the oral exam format is perceived as worthwhile.

As this oral examination format has been popular with TAs, the format has spread to other courses. Since the implementation of these two instances of the oral examination in the Fall and Winter of academic year 2017-2018, two other instructors have used this approach in Spring 2018. A former TA of the instructor who implemented the first version of the oral midterm reproduced it as an instructor for her own course on Diversity and Race. Another of the instructor's former TAs proposed the format for a linguistic anthropology course taught by a different instructor. Although novel, this instructor believed it made it clear whether the students knew the material, when they could speak about the concepts in this assessment format.

LIMITATIONS

This study is retrospective and based on cases that may be of interest to instructors with large lecture classes. The report is based on continuous improvement of assessment instruments (such as SETs) and reflection by a new assistant professor teaching large lecture classes for the first time, his TAs (who are graduate students in a PhD program), and student self-reports in SETs. Student preferences via SETs are not objective measures for the validation and success of an assessment (Uttl et al., 2017) and there are limitations in the data analysis as coding of qualitative data was conducted by one individual. Additionally, the sample is limited to students of this instructor and the two cases offered here also vary significantly. The student populations and their motivations compare a lower-level diversity requirement and an upper-level major requirement. As the data on oral assessments are limited and there is minimal literature on this assessment form, results must be viewed with caution.

FURTHER WORK AND CONCLUSION

Additional studies are necessary to make clear judgments about the efficacy and efficiency of the oral examination format. Only one study, Huxham et al. (2012), has used randomization between oral and written exam conditions to assess differences in student performance. Future studies could use randomized control designs powered to detect significant differences, i.e. randomizing half of the students or discussion sections of a large lecture to an oral or written exam condition. Regarding efficacy, oral examinations could test for differences in student understanding and retention at shorter and longer intervals (i.e., in a week, month, or six-month follow-up), processing of conceptual

information, self-efficacy in explaining concepts to others, and actual utilization or application of the material.

Future work can also examine differences in exam modalities for different student populations. There may be benefits for various groups of students, such as students with various disabilities. Future studies could also examine if novelty of the format makes a difference compared to standardized assessment practice. Older students may be more ingrained in traditional exam styles and thus more appreciative of a different format. As we observed a decrease in anxiety from between the two quarters an oral examination was administered, questions of difference in student maturation (i.e., practice speaking in courses) or commitment to the class content could also be explored as potential mediators of examination outcomes.

Finally, future inquiries should also closely measure the time commitments involved in preparing students for, administering, and grading both types of exams to determine how much time savings or costs, if any, are present. The types of scaffolding and preparation for students should be explored more thoroughly to reduce student anxiety. Time studies could identify potential benefits for large courses where TA grading needs to be minimized, for example, to increase the focus on graduate student research or where union contracts restrict TA hours. However, the time constraints of other oral examination formats, such as short interviews in smaller class settings, could also be explored more thoroughly to weigh the costs and benefits of oral assessments.

REFERENCES

- Badger, J. (2010) Assessing reflective thinking: pre-service teachers' and professors' perceptions of an oral examination, Assessment in Education: Principles, Policy & Practice, 17(1), 77-89.
- Boe, B. L. (1996). Diversity in assessment procedures. Paper presented at the American Education Association of Colleges for Teacher Education Conference, IL.
- Brown, G. A., Bull, J., & Pendlebury, M. (2013). Assessing student learning in higher education. Routledge.
- Buehler, M. J., & Schneider, L. U. (2009). Speak up! Oral examinations and political science. *Journal of Political Science Education*, 5(4), 315-331.
- Burke-Smalley, L. A. (2014). Using oral exams to assess communication skills in business courses. Business and Professional Communication Quarterly, 77(3), 266-280.
- Crist, J. F., & Robinson, K. (2015). Stewards of the word: Employing oral examinations in required theology courses. *Journal of Catholic Higher Education*, 34(2), 235-256.
- Dicks, A. P., Lautens, M., Koroluk, K. J., & Skonieczny, S. (2012). Undergraduate oral examinations in a university organic chemistry curriculum. *Journal of Chemical Education*, 89(12), 1506-1510.
- Huba, M. E., & Freed, J. E. (2000). Learner-centered assessment on college campuses: Shifting the focus from teaching to learning. Allyn & Bacon, Needham Heights, MA.
- Huxham, M., Campbell, F., & Westwood, J. (2012). Oral versus written assessments: A test of student performance and attitudes. Assessment & Evaluation in Higher Education, 37(1), 125-136.
- Lubrano, A. (2004). Limbo: Blue-collar roots, white-collar dreams. John Wiley & Sons.
- Perdigones, A., García, J. L., Valiño, V., & Raposo, C. (2009). Assessing heterogeneous student bodies using a methodology that

- encourages the acquisition of skills valued by employers. Assessment & Evaluation in Higher Education, 34(4), 389-400.
- Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, *95*(3), 385-408.
- Roberts, C., Sarangi, S., Southgate, L., Wakeford, R., & Wass, V. (2000). Oral examinations—equal opportunities, ethnicity, and fairness in the MRCGP. *BMJ: British Medical Journal*, 320(7231), 370.
- Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. Assessment & Evaluation in Higher Education, 30(2), 175-194.
- Saldaña, J. (2015). The coding manual for qualitative researchers. Sage.
- Sparfeldt, J. R., Rost, D. H., Baumeister, U. M., & Christ, O. (2013). Test anxiety in written and oral examinations. *Learning and Individual Differences*, 24, 198-203.
- Strauss, A., & Corbin, J. M. (1997). Grounded theory in practice. Sage. Waterfield, J., & West, B. (Eds.). (2005). Inclusive Assessment in Higher Education: A resource for change. University of Plymouth.
- Wiggins, G., & McTighe, J. (2005) *Understanding by design* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development ASCD.

APPENDIX A. ORAL EXAMINATION INSTRUCTIONS

Fall 2017 Oral Examination Questions with Rubrics and Answers:

Learning Outcomes:

The goal of this course is to increase student commitment to valuing diversity, equity, and inclusion in line with the Principles of Community. Students will gain an understanding of situations, practices, and policies based on racial analysis.

- Students will be able to synthesize an anthropological concept of race that attends to biological, socio-cultural, and historical perspectives.
- 2. Students will be able to distinguish and define key terms such as ethnicity / race / nationality, ethnocentrism / racism / nationalism, and implicit bias / prejudice / discrimination.
- 3. Students will be able to describe and assess differences in racialization and racial systems among various racial minorities in the US and other global contexts.
- 4. Students will be able to explain key concepts around race and racism such as the social construction of race, systemic racism, and racial privilege.

Midterm: The midterm tests students on their ability to explain concepts as covered in the first half of the course. The midterm will be comprised of approximately 10 short response questions, provided in advance. Students must be prepared to answer all questions. During the midterm, the student will randomly select one of the questions and answer it orally. TAs administering the exam will provide prompts to help students in completing their responses, but the response will be limited in time (approximately 90 seconds). We will go over the exam questions and practice responses in class.

Instructions: The midterm will be administered during discussion section. You will sign up for a time range and come prepared to orally answer the following questions with your TA. You will be in the room by yourself with other students waiting outside, so you will not be observed by other students nor receive their help.

You will have approximately 60 seconds to respond to the question posed to you. After that, the TA will prompt you and allow an additional 30 seconds to complete your response. Based on your response, you will be provided a grade for the midterm. A checklist of items that need to be in the response is included with corresponding point values, so you will know how the response will be graded. Questions vary from year to year.

Practice repeating your answers with peers so that you can respond effectively and efficiently.

WYSIWYG: What you see is what you get. The questions includes the grading rubrics, which the TAs will also prompt you with. **NOTA BENE:** TAs are grading the midterm based on the material that was covered in lectures and the readings. Whatever you learn through google or other sources may not count toward the midterm grade if it was not covered in the course.

I Why do anthropologists say that race is not primarily biological but socio-cultural? What does it mean for race to be "arbitrary" or "socially constructed?"

40% explain why race is "arbitrary" or "socially constructed" 20% give a biological example 20% give a socio-cultural example 20% give a historical example

2 If race is "arbitrary," then why does it still matter? Isn't it just about grouping different kinds of people into categories?

40% explain why race is "arbitrary" or "socially constructed" 20% explain the relationship between classification and consequences 20% give a socio-cultural example 20% give a historical or biological example

3 What does clinal variation mean? Can you give me two examples of clinal variation and at least one reason why that variation might exist?

40% definition of clinal variation 20% cline example I 20% cline example 2 20% hypothesized rationale for example 4 What are some problems with defining human races based on genetics? Can you give me two examples of genetic variation and at least one reason why that variation might exist?

40% critique of defining race based on genetic "populations"

20% genetic example I

20% genetic example 2

20% hypothesized rationale for example

5 How is it possible that some anthropologists claim that biological races do not exist but that how one is socially categorized into racial groups can produce biological differences?

40% explain developmental plasticity

20% what was Boas' foundational contribution about developmental plasticity

20% give a biological human variation example

20% give a social inequality example

6 The US government has a particular way of defining race and ethnicity that does not necessarily fit American folk beliefs about race and ethnicity. Describe some of these categories and their differences in relation to other systems of classification and how they have changed over time.

40% name 4 of the 5 official, mandatory racial classifications in the US Census

20% what "racial" group that is prominent in this state is considered an "ethnic" category on the US Census

20% give an example of an ethnic group that has changed or contested its racial category over time

20% give an example of an ethnic group that the system tracks differently from the US Census

7 Define these key terms: Race / Ethnicity / Nationality and how they differ in analytic and everyday definitions. Also, explain the difference between Racial Formation / Racialization.

20% Race

20% Ethnicity

20% Nationality

20% What is the difference between analytic and everyday definitions of these terms?

20% What is the difference between racial formation and racialization?

8 Define these key terms: Implicit Bias / Prejudice / Discrimination and explain what it means if someone says "we are all racist."

20% Implicit Bias

20% Prejudice

20% Discrimination

40% Explain why the statement "we are all racist" makes sense.

9 What is systemic, structural, or institutional about racism?

40% explain how racism is systemic, structural, or institutional

20% give an example from housing, education, or health

20% give an example from wealth, employment, or criminal justice

20% give an example dependent on one of the two prior examples or that demonstrates historical or cultural difference in racism

10 How is race biologized through social inequality and health disparities?

40% explain developmental plasticity

20% give an example of a biological difference that is modified by the environment

20% what is a mechanism described for differential racial health outcomes

20% give an example of a health condition that is negatively affected by ascribed race

Winter 2018 Oral Examination Questions with Rubrics and Answers:

The midterm is worth 20 points.

There are 10 questions that you must be able to answer. You will randomly pick 2 questions and answer them orally for a TA. It is expected that you will be able to respond to each question within 120 seconds. Each question will be worth 10 points (8 points based on content effectiveness and 2 points based on response efficiency). You are allowed one question pass, in which your response will be worth a maximum of 7 points).

- I. Most public health scholars and practitioners use the terms "health disparity" and "health inequality" interchangeably. What is a health disparity? [2] As a critical medical anthropologist, how does Singer (2013) differentiate between a "health disparity" and a "health inequality." Please define how Singer defines the two terms [4 define each @ 2 points] and give an example that contrasts the two [2 to describe a contrast between a disparity and inequality].
- 2. Critical medical anthropological perspectives are concerned about the overdetermination of "unreflexive depictions of [local] cultural practices as causal factors" (Craddock 2004:3). This often means that victims are blamed for inadequate outcomes of health measures. Describe what this critique is about [4] and relate it to social determinants of health or concepts in biocultural health, social epidemiology, or cultural inconsonance by providing two examples [4, 2 points each].
- 3. What is an epidemiological transition? Define the term [2], describe Omran's original proposal [4], and how Zuckerman's (2014) anthropological conceptualization of the concept has broadened our understanding of epidemiological transitions [2].
- 4. From an anthropological perspective, the first epidemiological transition is triggered by the development of agriculture (Neolithic period). Provide at least three examples [1-3 points per example] of ecological changes related to agriculture that negatively affected the health of human populations.
- 5. Anthropologists are ardent defenders of the need to consider culture. Douglas (2004) dismisses the idea of "traditional culture." Provide one definition of "culture" [4] and explain why this is not a contradiction with what Douglas describes as her critique [4] of "traditional culture."
- 6. Biocultural perspectives in medical anthropology, social epidemiology, and social medicine critically evaluate and decenter the dominance of biomedical reductionism by addressing interrelated sociocultural and environmental issues. How does Farmer (2004) define "structural violence." [4] What is Farmer's two-pronged strategy to ameliorate the health effects of structural violence? [4]
- 7. How does Farmer define "structural violence?" [2] What is the difference between behavioral risk and social vulnerability [2]. What level do each of these concepts stress [2]? What does a focus on social vulnerability in global health achieve? [2]
- 8. Gender and sexuality are often ignored in global health programs. What is important about the transition from reproductive to sexual health? [2] How does the mandate of sexual health expand from reproductive health? [4] Describe an additional impact of feminism or HIV/AIDS on how we conceptualize health and social justice. [2]
- 9. Define "syndemic." [3] How is it different that co- or multi-morbidity? [1] Provide an example of a syndemic focused on infectious diseases [2] and another focused on non-communicable diseases [2].
- 10. Singer (2013) cites Farmer as writing that health outcomes and profitable markets are at odds. Why is this the case? How does the "market" fail to protect human health? [3] Explain the conflict and provide examples from Farmer (2013) regarding diseases of poverty in low and middle income countries (LMICs). What are the neglected tropical diseases (NTDs)? [2] What is the 90-10 gap? [1] What is an example of a privileged lifestyle drug that is not essential but profitable? [1] What is an example of a low cost NTD drug or treatment that is not profitable? [1]

Pass Question: (minus 30% from one response) Draw a diagram that distinguishes vertical, horizontal, and diagonal global health interventions. Your diagram must identify at least four vertical programs [4], two horizontal initiatives [2], and one diagonal relationship [1].