The aim of this section is to expand and accelerate advances in methods of teaching bioethics. Bioethics educators are invited to send submissions to T. Kushner at kushnertk@gmail.com.

Examining Ethics

Developing a Comprehensive Exam for a Bioethics Master's Program

TOBY SCHONFELD, HUGH STODDARD, and CORY ANDREW LABRECQUE

Abstract: Assessing mastery of bioethics in a graduate program requires careful attention not simply to the content knowledge and skill development of students but also to the principles of sound assessment processes. In this article, we describe the rationale, development process, and features of the comprehensive exam we created as a culminating experience of a master's program in bioethics. The exam became the students' opportunity to demonstrate the way they were able to integrate course, textual, and practical knowledge gained throughout the experience of the program. Additionally, the exam assessed students' proficiency in the field of bioethics and their ability to critically and constructively analyze bioethical issues. In this article, we offer tips to other exam creators regarding our experiences with question and answer development, scoring of the exam, and relationships between coursework and exam preparation and completion. We also include a sample rubric for others to see how we determined which student answers were satisfactory.

Keywords: ethics education; exam; assessment; rubric; minimal competency; content mastery

Assessing mastery in bioethics is no easy task. Professionals have grappled with this difficulty in medical education, in which the major licensing examination asks ethics questions in the same multiple-choice format as clinical questions.[1] Individual medical education training programs have experimented with alternative modes of assessment of bioethics, such as reflective writing.[2,3,4,5,6] However, the creation of a formal, evaluative structure for such writing projects continues to be challenging.[7,8,9]

Assessing mastery of bioethics material in a graduate program dedicated to that field of study is no less problematic. Success in individual courses may not indicate overall mastery of material, which requires a synthesis and analysis of issues that is both broader and deeper than may be apparent from the particular focuses of classes. Indeed, given that a successful education in bioethics produces changes in the knowledge, skills, and attitudes of learners, identifying any one mechanism to assess these changes is an unresolved challenge for bioethics educators.

To attempt to meet that challenge, we designed a comprehensive examination

Cambridge Quarterly of Healthcare Ethics (2014), **23**, 461–471.

doi:10.1017/S0963180114000139

as an option for students' summative project. In this article, we describe the rationale for that exam, the process of development, the logistics of the exam delivery and scoring, and lingering issues with the exam for both students and faculty.

Why Create an Exam?

Prior to the development of the examination, all students at our institution completed a thesis as their summative project for the master of arts in bioethics (MAB). In some ways, the thesis worked well: it enabled students to delve deeply into a particular issue that interested them during the course of the program, it allowed them to demonstrate their research and analysis skills, and it required them to demonstrate a sensitivity to context that reflects important attitudinal components of bioethics.[10]

Yet for some students, writing a thesis was problematic. For one thing, the thesis process takes about a year to complete, and this time frame was a challenge for students who were completing the degree as a "gap-year activity" in between other educational programs. In addition, many dual-degree students stepped out of other programs for only a year and therefore had to be able to complete all of the requirements within a 12-month period. Finally, for some students, a comprehensive exam was an attractive option because it enabled them to review the full breadth of knowledge and skills they had acquired during the program, rather than narrowing focus to a single thesis topic.

For the faculty, offering students an examination rather than a thesis presented an intriguing opportunity. Even a high-quality thesis might not provide sufficient evidence that a student had mastered any more than one particular issue in bioethics. A comprehensive exam could give insight into whether the student had met the program's content and process objectives. More quality improvement evidence—in terms of both quantity and quality of evidence—regarding the adequacy of the entire curriculum was made possible through the use of a common metric, like exam scores, rather than through individual thesis completion. Even more importantly, the creation of a comprehensive exam required deep reflection by the faculty and program leaders on the core concepts, skills, and attitudes that needed to be emphasized in each core course and the interconnections among those courses. Finally, a comparative study of individual student performance in courses with the student's overall mastery of material through a comprehensive examination led to important conversations about consistency, progress, and the framing of what constitutes "success" in the program.

Therefore, in consideration of the expected benefits to both students and faculty from the creation and administration of a comprehensive exam, we embarked on developing and administering a comprehensive exam as an option for students completing the MAB.

Process of Development

The comprehensive examination is a common and important learning assessment tool in higher education. Exam results not only assist evaluators in measuring students' substantive knowledge of the field but also allow the student to demonstrate acquired skill in synthesis, integration, and constructive analysis—higher-level thinking skills that are essential to graduate work in most disciplines. Defining the appropriate body of knowledge and manner of analysis also orients both professors and students toward cultivating learning that is by nature *comprehensive*.

At the doctoral level, students typically tend to undertake a series of comprehensive examinations (both oral and written), which collectively serve as a prerequisite—and as evidence of competence—for the demands of dissertation writing.[11] Leonard Cassuto contends that "the comprehensive exam should therefore bear a concrete relation to the dissertation and the work—that is, the research methods and practices—that will be required to complete it."[12] In other words, the comprehensive examination must be *relevant* for the student's trajectory in the program, rather than being an obstruction to be breached.

In addition to these important considerations, the challenge of designing a comprehensive examination for a master's degree in bioethics that appeals to a diverse student body that includes postbaccalaureates and mid-career professionals further convolutes the task at hand. Moreover, if an examination is offered as an option—of equal rigor as the thesis—to culminate the program and not as a requirement for the student to proceed to thesis writing, then what "concrete relation" are we trying to forge? A master's degree program that offers an examination and a thesis as two valid options for the completion of the degree speaks to Cassuto's call for pedagogical relevance insofar as the student can choose which route best serves his or her expectations from graduate study in bioethics.

Very few universities in the United States that have bioethics degree programs at the master's level include a comprehensive examination as an integral part of the curriculum. Fewer still present it as an equally rigorous alternative to the thesis. We conducted an inventory of 30 U.S. universities that offer a master's-level degree (or a concentration) in bioethics, healthcare ethics, or medical humanities. Based on the degree requirements listed on the schools' program websites, 25 out of the 30 do not require a comprehensive examination of any kind. Only 3 out of the 30 schools require students to pass comprehensive examinations to receive a degree. The 2 remaining programs—at Emory University in Atlanta, Georgia, and Trinity International University in Deerfield, Illinois—have a comprehensive examination as an option. At Emory University, a student can take the exam or write a thesis, whereas at Trinity International, the exam is one option among many—including a thesis, integrative paper, or special project—to meet the capstone requirement. Interestingly, all 5 university programs that either require a comprehensive examination or offer it as an option are *master of arts* in bioethics programs.

Following the national benchmarking to discover how other bioethics programs used exams, we had several faculty discussions about the form and content of the exam. In order to ensure that we were assessing the appropriate learning outcomes, we attended carefully to the program objectives. Our overall learning goals for the program are as follows:

1) To enhance students' ability to demonstrate critical analysis of bioethical issues
2) To strengthen students' ability to develop robust moral arguments
3) To improve students' ability to communicate clearly in written and verbal formats

Any assessment we used needed to connect to the outcomes of the program and measure success in the achievement of those outcomes. Importantly, note that the exam is simply the final assessment of students' achievement of our goals. Assessments also occurred throughout the duration of the program to measure students' mastery of knowledge, skills, and attitudes within the boundaries of

a specific course. These in-process assessments were used for both summative and formative purposes.

Exam Features

The comprehensive examination was intended to be a culminating experience of the MAB. Therefore, the exam became the students' opportunity to demonstrate the way they were able to integrate course, textual, and practical knowledge gained throughout the experience of the program. The exam assessed students' proficiency in the field of bioethics and their ability to critically and constructively analyze bioethical issues. Accordingly, answers to the exam questions had to be thorough, structured, well deliberated, and engaging.

The exam consisted of six questions based on the core courses that all students had to take. Students were required to answer four of the six questions. The exam was divided into three sections:

- Section 1: Foundation questions, mostly from our theory course
- Section 2: Integration questions, mostly from our classic and contemporary issues courses
- Section 3: An application question, in which students responded to a case study

Students had six hours to complete the exam; there were two three-hour sittings scheduled on the same day. The exam was completed in our office, on a computer provided by the Center for Ethics. No notes, texts, or other resources could be brought into the examination room.

We notified the students of the questions for sections 1 and 2 well in advance of the exam. In fact, a full list of potential questions from which examiners would choose was made available to students on the first day of the fall semester. Thus, exam preparation focused on reviewing and integrating knowledge, rather than guessing what would be asked on the exam. We also informed students that any topic included among the published questions was expected knowledge from the program and might be tested, even if the students did not master those topics in the particular core course that they had taken. Therefore, students needed to rely on self-directed learning in their exam preparation by self-identifying topics about which their understanding was insufficient. Likewise, students were expected to make use of extracurricular materials as needed. We provided a program-level reading list for students that should have been accessed in addition to the material included in the syllabi of the core courses.

Question and Answer Development

Particular questions were created by *expert panels* of faculty members. The comprehensive exam was intended to measure students' achievement across the entire program, rather than in a single course or content domain. To accomplish this, the faculty stakeholders needed to come to a consensus on the content that would be examined as well as on the actual formulation of each question. Because students were given the questions in advance, the question bank needed to be truly comprehensive, rather than containing only a sampling of topics. The use of a faculty panel for this task was consistent with a basic tenet of assessment that examiners must have a theory of how examinees will represent knowledge.[13,14] Negotiation among the members of the expert panel was a long and engaging process that unified the exam in terms of the expected student cognition and that ameliorated differences among examiners based on professional interests or personal biases. The use of expert panels to define exam

content is recommended practice among testing professionals.[15,16]

In order to ensure that questions were specific and clear, the expert panel created individual questions and then *composed answers to those questions to define what they expected students should be capable of answering*. The answer-writing process is a step that we suspect is often skipped pedagogically, as faculty members may feel that they "know" the right answer and therefore do not need to write it out. For the faculty, composing answers revealed fundamental ambiguities in some questions, gave us pause to reflect on what and how we were teaching, and compelled us to think self-critically about what we expected students to demonstrate and whether the question could elicit the expected response. In the vast majority of instances, we then revised individual questions to be more precise in terms of asking for the material we judged to be essential. For example, in the initial drafting phase, one exam question was written as follows: "If we can enhance someone (make them smarter, more agile, more attractive, etc.), then we should do so. Describe this position fully and then offer a compelling objection." After the participation of the expert panel, the question was revised to read: "In bioethics, what do we mean by 'human enhancement?' (2) Some have argued that if we *could* enhance someone, then we *should* do so. Describe this position fully by providing at least three pertinent arguments in favor of enhancement (2). Then, offer a compelling objection to this claim by providing at least three pertinent arguments against enhancement (2)."

The answer-writing process was guided by the simultaneous creation of a scoring *rubric*.[17,18] A scoring rubric has three main components: a set of criteria on which students are being assessed, an ordinal scale on which students are scored for each criterion, and descriptors for each level of the scale that describe the features that characterize that score level for that criterion.[19]

Our purpose for creating and using a scoring rubric was threefold:

1) To ensure a consistency of expectations for answers among the faculty graders and to remove construct-irrelevant variance
2) To ensure interrater reliability in scoring particular answers
3) To facilitate the communication of faculty expectations to students, both preexam and postexam

The interpersonal negotiation necessary to create the scoring rubric, especially in writing the descriptors for various levels of the criteria, led to an increased sense of purpose among the faculty for the entire program. Because they were creating a comprehensive exam in which students were required to integrate knowledge from multiple courses in order to achieve a passing score, developing the rubric forced faculty stakeholders to expose their beliefs and assumptions about the knowledge and attitudes that students were achieving in our program. A key component of this refocusing was to remove construct-irrelevant variance from the scoring rubric.[20] For example, nonnative English speakers would not receive lower scores based on grammatical proficiency, as long as the appropriate levels of knowledge and attitudes were demonstrated with clarity and coherency.

Interrater reliability is often borderline when using rubrics[21,22] and especially when scoring responses to open-ended questions. High-quality scoring rubrics are very difficult to develop[23] and should be created through an iterative process.[24] Nonetheless, a key part of any rubric's educational value lies in the process of creating it, which is why we would advise against borrowing rubrics from

others. We provide an example of our rubric in the Appendix merely to demonstrate its utility and form as a way of inspiring the creation of rubrics in bioethics. In the attempt to achieve adequate interrater reliability for a rubric, developers, scorers, and examiners must discuss and negotiate the nuances of the criteria, scale, and descriptors used. When such a discussion of program goals occurs among those faculty members who teach courses, advise students, create the exam, and score the papers, then the whole process contributes to the quality of the students' overall educational experience in the program.[25]

Well-constructed scoring rubrics not only improve scoring reliability but also facilitate communication with students about program goals and expectations.[26] In some instances, a rubric may even be presented to students as notification of the standards to which they will be held; we give a sample rubric to students to demonstrate how their answers will be assessed. As a post facto device, a completed rubric will provide a low-performing student with specific detail about deficiencies or areas for remedial attention.[27]

Scoring of the Exam

Because the questions for sections 1 and 2 were available in advance, students' responses were expected to reflect careful consideration and integration of the issues described. Students' answers were expected to be complete, while focusing on the answer to the particular question asked. Answers were expected to be concise, coherent, well organized, and clear.

The overall exam was graded pass/fail. Each question was scored according to a *minimal competency standard*. That is, students were expected to achieve the level of minimal competence for graduate work. The goal was to assess whether or not the student had acquired a sufficient breadth and depth of knowledge of bioethics. With this philosophy in mind, we established that students must pass all four questions in order to pass the exam.

Each student exam was graded by two faculty members who were assigned that semester by the director of graduate studies. The primary faculty member assigned a score to each question, using the rubric for the answer as a scoring guide. The secondary faculty grader provided a global pass/fail assessment, and then both graders came together to make a final judgment about the acceptability of the answers. In the unlikely event that the two graders could not come to an agreement, the director of graduate studies reviewed the answer and provided assistance. If the director of graduate studies was one of the assigned graders, then disputes were to be resolved by the director or associate director of the Center for Ethics.

Students who failed the exam could retake it up to a maximum of three total times. The first time a student failed the exam, the director of graduate studies contacted the student to report the failure and to provide information related to rewriting the exam. The student was then reexamined only on the area(s) in which the student failed the first exam. This reexamination occurred one month after the failure was reported to the student, and the reexam consisted of (a) new question(s) in said area(s).

In the event of a second failure, the student would take a new version of the full examination the following semester. In this case, there was no retake option; the student had to pass all questions in this final attempt. After three exam failures, the student would be dismissed from the MAB program.

Additional Factors during Development

A number of other factors were essential to the exam development process. Space precludes a detailed discussion of these issues; however, each factor in the creation process is worth noting:

1) *Establishing a minimum competency standard for the assessment.* Standard setting for any educational measure is complicated and fraught with pitfalls.[28] To define a performance standard for our narrowly tailored master's degree program, we used the Structure of the Observed Learning Outcome (SOLO) taxonomy as our theoretical framework.[29] The SOLO taxonomy defines level of performance according to quantity of information presented and level of integration of that information. In our case, because this was a comprehensive exam and students were expected to combine what they had learned across multiple courses, the SOLO taxonomy was particularly useful. We employed the direct consensus method,[30] with modifications for a constructed response exam format. Detailed information about the critical step of establishing a minimum performance standard will be presented in a forthcoming manuscript.
2) *(Lack of) utility of numbers.* As is evident from the sample rubric included in the Appendix, we originally designed the assessment so that a numerical value accompanied each discursive category. However, in the actual administration of the exam, we found the numbers to be more distracting for graders than helpful—examiners spent more time discussing the mathematical relationships of the numbers than they did whether or not the student's answer represented a satisfactory mastery of the material. This experience could have been predicted based on the literature about rubrics that emphasizes treating ratings as categorical rather than quantitative.[31,32] In future iterations, we intend to remove the numbers from the grading rubric.
3) *Equivalency of summative assessments.* Because our assessments are designed to measure achievement of the core competencies of our program, it is essential that both the exam and the thesis are equally rigorous. In addition to combatting student perception that one path was easier or less time-consuming than the other, we also endeavored to ensure that the time and intensity of the actual preparation for both options (draft writing of the thesis and studying for the exam) required equivalent efforts on the part of the student. Whether or not we were successful in achieving the goal of preparatory similarity will be reevaluated once the comprehensive exam program matures.
4) *Curricular revisions.* Even after having reaffirmed our commitment to the core competencies prior to creating the exam, it became clear after the first administration of the examination that curricular modifications were likely to be necessary in order to better prepare students for the examination. Rather than immediately making content modifications, faculty members have begun discussing how course-specific assessments do (or do not) engage students in the kind of discourse required for a satisfactory examination response. These conversations are ongoing.

Conclusion

The comprehensive examination was designed to assess students' proficiency in the field of bioethics and to make a global assessment of their ability to critically and constructively analyze bioethical issues. For our students, the exam option fulfills the need for a high-quality summative assessment that students who are unable to devote the time to writing a thesis can complete. For students for whom writing a thesis or preparing another large project is not a realistic possibility or desire, the comprehensive exam allows them to demonstrate their knowledge and skills in a manner that is equal in rigor to writing a thesis. Because the comprehensive exam should be perceived as equivalent to writing a thesis, careful attention to the design of the exam and the accuracy of scoring the work is particularly important. Although we will continue to revise pieces of the exam, we remain committed to offering an option for students that enables them to demonstrate the integration of knowledge, skills, and attitudes they have developed over the course of their study in our master of arts in bioethics program.

Notes

1. Federation of State Medical Boards and National Board of Medical Examiners. *United States Medical Licensing Examination*; 2013; available at http://www.usmle.org/ (last accessed 10 Dec 2013).
2. Aronson L, Niehaus B, DeVries CD, Siegel JR, O'Sullivan PS. Do writing and storytelling skill influence assessment of reflective ability in medical students' written reflections? *Academic Medicine* 2010 Oct;85(10 Suppl): S29–32.
3. Boyd A. The impact of a reflective writing workshop on GP trainee attitudes to writing eportfolio log entries. *Education for Primary Care* 2012 Dec;23(6):430–3.
4. Chretien K, Goldman E, Faselis C. The reflective writing class blog: Using technology to promote reflection and professional development. *Journal of General Internal Medicine* 2008 Dec;23(12):2066–70.
5. Wear D, Zarconi J, Garden R, Jones T. Reflection in/and writing: Pedagogy and practice in medical education. *Academic Medicine* 2012 May;87(5):603–9.
6. Wald HS, Reis SP. Beyond the margins: Reflective writing and development of reflective capacity in medical education. *Journal of General Internal Medicine* 2010 Jul;25(7): 746–9.
7. Eckles RE, Meslin EM, Gaffney M, Helft PR. Medical ethics education: Where are we? Where should we be going? A review. *Academic Medicine* 2005;80(12):1143–52.
8. Goldie J, Schwartz L, McConnachie A, Morrison J. The impact of three years' ethics teaching, in an integrated medical curriculum, on students' proposed behaviour on meeting ethical dilemmas. *Medical Education* 2002 May;36(5):489–97.
9. Terndrup C. A student's perspective on medical ethics education. *Journal of Religion and Health* 2013;52(4):1073–8.
10. Core Competency Task Force. *Core Competencies for Healthcare Ethics Consultation*. Report No. 2. Glenview, IL: American Society for Bioethics and Humanities; 2011.
11. Nolan R. How graduate students of adult and continuing education are assessed at the doctoral level. *Journal of Continuing Higher Education* 2002 Sept 1;50(3):38–43.
12. Cassuto L. The comprehensive exam: Make it relevant. *The Chronicle of Higher Education* 2012 Mar 4; 58(27):A33.
13. Lane S, Stone CA. Performance assessment. In: Brennan RL, ed. *Educational Measurement*. 4th ed. Westport, CT: American Council on Education; 2006.
14. Pellegrino JW, Chudowsky N, Glaser R. *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academies Press; 2001.
15. See note 13, Lane, Stone 2006.
16. Ewing M, Packman S, Hamen C, Thurber AC. Representing targets of measurement within evidence-centered design. *Applied Measurement in Education* 2010;23(4):325–41.
17. Popham WJ. What's wrong—and what's right—with rubrics. *Educational Leadership* 1997;55:72–5.
18. Andrade HG. Teaching with rubrics: The good, the bad, and the ugly. *College Teaching* 2005;53(1):27–31.
19. Moskal BM. Scoring rubrics: What, when and how? *Practical Assessment, Research & Evaluation*

2000;7(3); available at http://pareonline.net/getvn.asp?v=7&n=3 (last accessed 28 Apr 2014).
20. See note 13, Lane, Stone 2006.
21. Sadler DR. Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education* 2009;34(2):159–79.
22. Tierney R, Simon M. What's still wrong with rubrics: Focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research & Evaluation* 2004;9(2); available at http://pareonline.net/getvn.asp?v=9&n=2 (last accessed 28 Apr 2014).
23. See note 21, Sadler 2009.
24. See note 13, Lane, Stone 2006.
25. See note 22, Tierney, Simon 2004.
26. Reddy YM, Andrade H. A review of rubric use in higher education. *Assessment & Evaluation in Higher Education* 2010 July;35(4):435–48.
27. See note 18, Andrade 2005.
28. Hambleton RK, Pitoniak MJ. Setting performance standards. In: Brennan RL, ed. *Educational Measurement*. 4th ed. Westport, CT: American Council on Education; 2006.
29. Biggs JB, Collis KF. *Evaluating the Quality of Learning*. New York: Academic Press; 1982.
30. See note 28, Hambleton, Pitoniak 2006.
31. See note 22, Tierney, Simon 2004.
32. Mertler CA. Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation* 2001;7(25):1–10.

APPENDIX

Example Question and Grading Rubric

Examinations will be graded *by each question*; questions will be scored according to the following scoring rubric:

- Missing information or incorrect information: 0 points
- Partial credit: 1 point
- Satisfactory answer: 2 points
- Exceptional answer: 3 points

Please note that this requires separating each question into its constitutive parts and assessing each part separately; the parts are then summed to determine the overall score for the answer. *A satisfactory answer is one in which the student scores at or above the total of the satisfactory points possible.* For example, for a question that has three parts, students must score at least a six out of nine in order to pass the question. (So a student might excel in one part of the question but be deficient in another part and still pass the question).

General Descriptions of Score Levels

Level 0: Missing or Incorrect Information

Answers in this category have the following characteristics:

- The student fails to address every part of the question.
- The student fails to address items that are central to the theory or concept.
- The student makes claims that are incorrect.

Level 1: Partial Credit

Answers in this category have the following characteristics:

- The student's answer is confusing, disorganized, or poorly written but has some truth to it.
- The student's answer is correct but incomplete (i.e., it describes part of a theory correctly).
- Nuances and key details are missing or glossed over.

Level 2: A Satisfactory Answer

Answers in this category have the following characteristics:

- The student answers every part of the question correctly.
- The student's answer is complete, coherent, and well written.
- The student demonstrates competency for a master's-level graduate student in bioethics.

Level 3: An Exceptional Answer

Answers in this category have the following characteristics:

- The answer demonstrates sophisticated, integrative thought and is developed clearly throughout.
- The student identifies, discusses, and extends conclusions, implications, and consequences, while considering context, assumptions, data, and evidence; the student qualifies his or her own assertions with balance.
- The student's language clearly and effectively communicates ideas; it may at times be nuanced and eloquent.
- The student fully integrates perspectives from a variety of sources; any analogies are used effectively.

Sample Question

> Defend the obligation to return research results to participants utilizing the principles from the Belmont Report (6). Then offer one counterargument for this position (2). Which do you support and why (2)?

Resource: See Miller FA, Christensen R, Giacomini M, Robert JS. Duty to disclose what? Querying the putative obligation to return research results to participants. *Journal of Medical Ethics* 2008;34(3): 210–13.

Answer Key (10 Points Required to Pass)

A. Respect for persons

a. *Example of missing or incorrect information (0):* Student fails to list or accurately describe the principle.

b. *Example of partial credit (1):* Student limits the discussion to respect for autonomy; student fails to connect obligation to informed consent.

c. *Example of a satisfactory answer (2):* Answer includes the following:
   i. Participants are ends in themselves, not means to an end.
   ii. Student describes disclosure of results as the final part of the informed consent process.
   iii. Disclosure of results provides for full consideration of the risks/benefits of this and similar trials, especially when the results are real-time (consider interim reports).

d. *Example of an exceptional answer (3):* Student describes the deontological ground of this principle; differentiates this principle from simply respect for autonomy; describes the nature and origins of professional obligations.

B. Beneficence

a. *Example of missing or incorrect information (0):* Student fails to list or accurately describe the principle (e.g., talks about "benefiting" patients and not about risk-benefit analysis).

b. *Example of partial credit (1):* Student describes the possible benefits of disclosure but not the risks of disclosure (or vice versa); student fails to consider the influence of the results on participant welfare.

c. *Example of a satisfactory answer (2):* Answer includes the following:
   i. Risk-benefit assessment may be altered by knowing whether or not the participant will get the (aggregate or individual) results.
   ii. Student describes the possible risks *and* benefits of disclosure and how the practice ensures a positive benefit-risk ratio.
   iii. Student notes that the results may influence patient health or healthcare; that is, the results may either suggest

(dis)continuation of the trial, in the case of interim results, or provide information about healthcare more generally, in the case of posttrial results.

d. *Example of an exceptional answer (3):* Student provides a nuanced description of retroactive risk-benefit assessment; includes a description of how to disclose and minimize risks or maximize the benefits of both the study results and the disclosure itself; distinguishes between aggregate and individual results from the perspective of risk-benefit analysis.

C. Justice

a. *Example of missing or incorrect information (0):* Student fails to list or accurately describe the principle.

b. *Example of partial credit (1):* Student describes the relevance of fair participant selection but not of fair sharing of benefits and burdens (or vice versa).

c. *Example of a satisfactory answer (2):* Answer includes the following:

i. Participants are research partners who deserve to know the results.

ii. Student links disclosure to a fair sharing of the benefits and burdens of research.

iii. Participants ought to have direct and immediate access to information.

iv. Student describes how the failure to return results may impede recruitment and retention and may create an environment of distrust.

d. *Example of an exceptional answer (3):* Student discusses the limits of collaboration with research subjects; describes in detail how this relates to historic or other sources of distrust in research; formulates a mechanism to ensure all participants have access to trial results.

D. Counterargument (must identify one)

a. *Example of missing or incorrect information (0):* Student fails to cogently describe a counterargument or incorrectly describes a counterargument.

b. *Example of partial credit (1):* Student names a counterargument but fails to defend it completely or misses important points.

c. *Example of a satisfactory answer (2):* Answer includes the following:

i. Disclosure may create more harm than benefit.

ii. The duty itself is unclear (proactive or passive).

iii. Should aggregate data or individual data be disclosed?

iv. We often don't know what the results mean.

d. *Example of an exceptional answer (3):* Student offers a cogent rebuttal for any counterargument; offers more than one counterargument.

E. **Justifies own position**

a. *Example of missing or incorrect information (0):* Student fails to offer a position.

b. *Example of partial credit (1):* Student uncritically endorses the proposal or counterargument.

c. *Example of a satisfactory answer (2):* Answer includes the following:

i. Student identifies a position and defends that position.

ii. If the student endorses obligation, then he or she offers a convincing rebuttal for the counterargument.

d. *Example of an exceptional answer (3):* Student offers a clear and convincing argument *not* included in the original obligation or in the counterargument proposed.