Testing for Deep Understanding
Ben Eggleston, Philosophy

To better test students' deep understanding of course material in his large lecture course, a philosophy professor develops new multiple-choice questions with answers in the form of hypothetical conversations for students to read and critically examine.

BACKGROUND

I teach an introductory ethics course with an enrollment of about 240 students, most of whom take the class in order to fulfill a Philosophy and Religion general-education requirement. Assignments include both papers and tests, and I typically have three graduate teaching assistants (GTAs) who are responsible for all of the grading.

Papers are inevitably burdensome for the GTAs to grade, and the tests can be as well, if they require the grading of constructed responses. To ease the burden on the GTAs, I began to rely on multiple-choice questions for the tests, but I was increasingly dissatisfied with the quality of the questions I used. I felt that they rewarded memorization too much and did not require students to demonstrate any deep understanding of course material. Those concerns led me to revise many of my test questions.

In doing so, I had five goals in mind:

1) Maintain a test format that could be machine graded, rather than a format that would require judgment on the part of the grader.

2) Continue to use questions that are adaptable to various textbooks or other instructional resources. The questions should not have to be completely rewritten if I switch to a different textbook that covers the same topics.

3) Use questions that measure not only memorization and basic comprehension but also deeper understanding. Usually, such understanding is tested with short-answer or essay questions; it's unusual for both goals to be achieved by the same questions.

4) Use questions that students and others will regard as respectable measuring devices. The questions should appear to students, GTAs, and other teachers as measuring deep understanding; i.e., it should be evident that they achieve the previous goal.

5) Use questions that can be adapted for use in the future to avoid the dilemma of either writing new questions from scratch each year or trying to keep each test secure after it is administered. Trying to keep tests secure is problematic on many levels; I find it far preferable to make previous years' tests available to current students, and using similar questions across years both enhances the fairness of the tests and facilitates comparisons of student achievement over time.

IMPLEMENTATION

Unlike questions that test only memorization of definitions, the new questions require students to apply deeper understandings of concepts to novel situations, as illustrated in the following two pairs of questions. The two pairs of questions are from different tests, covering different parts of the course. Within each pair, the first question is a question that I was dissatisfied with because of the ability of students to answer it correctly based on little more than having memorized a definition. The second question requires the same basic comprehension, but also requires a deeper understanding of the concepts involved.

Test One – Old question from my test on meta-ethics:

What is the main idea of cultural relativism?

> *(I)*　*Moral beliefs vary from one culture to another.*
>
> *(J)*　*Morality itself (not just moral beliefs) varies from one culture to another.*
>
> *(K)*　*No culture has one particular morality that can be identified with it.*
>
> *(L)*　*Morality is the same all over the world because all cultures share the same beliefs.*

The answer to this question is J, and a student could answer this question correctly by simply memorizing this fact from the textbook or class notes.

New question from my test on meta-ethics:

In the following dialogue, which of the following statements is incompatible with cultural relativism?

> *(A)*　*Robert: "Child labor is found in many cultures. When children are forced to work in factories as early as the age of eight, they suffer serious consequences in terms of lost educational opportunities, limited time for peer interactions, and risks of injury or death."*
>
> *(B)*　*Larry: "Some countries rely heavily on child labor, and would suffer devastating economic consequences if they were forced to give it up."*
>
> *(C)*　*Robert: "Despite these consequences, the harms to children are too great to ignore. It is wrong of those cultures to force children to work."*
>
> *(D)*　*Larry: "Perhaps those cultures could be persuaded, with carefully targeted economic incentives, to voluntarily discontinue their practices of child labor."*

Since the students had not seen this dialogue previously, nor is it likely that they had memorized a list of statements that are incompatible with cultural relativism, this question effectively requires students to have a solid grasp of cultural relativism. If a student were to read the foregoing conversation and did not notice that Robert's assertion of a culture-independent moral claim (in answer C) commits him to denying cultural relativism, then I would not regard that student as really knowing what it meant to affirm, or deny, that view. And yet, such a student might well be able to recite the definition from memory.

Test Two – Old question from my test on normative ethics:

Utilitarianism says that, in any circumstance, the right thing to do is _____.

> *(A)   whatever God commands*

> *(B)   whatever will be most immediately useful*

> *(C)   whatever will make you as happy as possible*

> *(D)   whatever will make everyone affected as happy as possible*

Again, the correct answer to this question (which is D) could be known merely on the basis of memorization.

New question from my test on normative ethics:

Which of the following statements is the strongest evidence that the person saying it is a utilitarian?

> *(A)   Ginny: "Violations of rights are very serious, from the moral point of view."*

> *(B)   Helen: "I agree. It is always immoral to violate someone's rights."*

> *(C)   Ginny: "Well, I wouldn't say 'always'. It's o.k. to violate rights whenever the good you can produce by doing so outweighs the harm you do by violating the person's rights."*

> *(D)   Kate: "I disagree with both of you. In my opinion, it's immoral to violate someone's rights unless doing so is the only way to prevent that person from violating others' rights."*

Like the previous conversation, this one would not have been shown to the students in advance, and they would not be likely to have blindly memorized anything that would point them in the direction of the correct answer (which is C). Instead, a student would have to understand the underlying idea of utilitarianism, and the commitments associated with it, in order to answer this question correctly.

The revised questions, then, measure deep understanding by requiring students to detect statements in hypothetical conversations that meet certain criteria. Because the statements that the students are being asked to detect are related to the main ideas (e.g., cultural relativism and utilitarianism) in somewhat rigidly defined ways (e.g., incompatibility and inductive support), the questions are not arbitrary and can be seen to have one clearly best answer. The advantage of the conversational format for presenting the statements is two-fold. First, they provide for better "camouflaging," so that the student has to grasp the content of what is being said rather than merely being confronted with a phrase or expression that he or she would be more likely to remember from the book or class notes. Second, they better test the kind of understanding that will serve students well outside the classroom. In normal conversations, ethical views are rarely affirmed, denied, or supported in the canonical terms of textbook definitions; instead, views are expressed more indirectly. The revised questions come closer to simulating this mode of discussing ethical issues.

OBSERVATIONS

The new questions appear to achieve the five goals discussed in the Background section. First, they do not depart from the multiple-choice format, thereby continuing to allow for mechanical grading; and second, they are perfectly adaptable to various textbooks or other instructional resources.

Third, as explained in the Implementation section, these questions are designed to measure not only memorization and basic comprehension, but also deeper understanding. This is their most significant feature.

My fourth goal was to design questions that students and others will regard as respectable measuring devices. Student reaction has been largely positive; anecdotal evidence suggests that most students appear to see the point of these more complex questions. A question is occasionally criticized by a student as being a "trick question," but I think that, in a class of 240 students, it might be impossible to ask challenging questions and avoid hearing that phrase. In addition to being favorably received by students, these questions have met with positive reactions from GTAs and other teachers.

Finally, I wanted to design questions that could be easily adapted for use in the future so that each test would not have to be kept secure after it is administered on pain of having to write new questions from scratch every year. The revised questions achieve this goal, in that small rephasings of the remarks attributed to the speakers in the conversation can give the question a sufficiently different feel from one year to the next. It is also feasible to write whole new conversations. Once you have these "conversation" questions in mind, you'd be surprised how often you hear conversations around campus or around town that give you ideas for new questions. The revised questions are actually better than the old ones in this respect, since questions asking for a simple definition can be rephrased only to a limited extent before they no longer accurately define the concept in question.

Moreover, as I mentioned above in connection with this last goal, I prefer to make previous years' tests available to current students. Even if that weren't necessary as a means of counter-acting the unfairness of unequal access to cold-test files, I would gladly make these "conversation" questions available, since they provide yet another opportunity for students to learn to apply course concepts to hypothetical but realistic out-of-class contexts.

REFLECTIONS

In reflecting on this project, two lines of thought occurred to me. One has to do with the thought processes involved in answering the new, conversation-based questions, and the other has to do with gathering useful data.

In regard to the thought processes involved in answering the new questions, there seem to be good reasons for thinking that the new questions test deep understanding better than the old ones do. As I mentioned in the Implementation section, the old questions essentially tested recall ability, whereas recall ability would be of very limited use in answering the new questions. Moreover, when I think about what long-term skills I would like for students to take away from my introductory ethics course, I think of the ability to discern—in a casual conversation, a policy-making meeting, or any realistic context—what ethical commitments are involved in certain opinions or positions. As I noted above, people rarely explicitly affirm or deny specific ethical views; instead, they more often say things that allow such affirmations of denials to be inferred (with greater and lesser degrees of confidence). So it's much more important, to me, for my students to be able to make such inferences than for them to just be able to recite some definitions.

Therefore, I am trying to develop in my students a certain kind of alertness. Just as a music-appreciation class should help students notice properties of music that they might otherwise miss, I think an introductory ethics class should help students notice properties of ethical statements and conversations—as these are found in ordinary life, every day—that they might otherwise miss. After an introductory ethics class, a student who hears other people making ethical statements, or who makes them herself (as everyone does), should be struck differently by those statements, and the commitments they entail, than she would have been if she had not taken the class. It is precisely this alertness, or heightened sensitivity, that the revised questions test. Just as a musically informed person hears more in a certain piece of music than a musically uninformed person does, I want my students to hear things in ethical statements that others hearing the same conversation might be oblivious to.

This project suggests a number of further questions to explore empirically. First, although I know the new questions are answered correctly less frequently than the old ones were, which is not surprising, it would be interesting to see whether virtually all of the students who can correctly answer the conversation-based questions can also correctly answer the simpler corresponding questions. While I spoke of the new questions as revisions of the old ones, the new ones are different enough from the old ones that both could be asked on a test without redundancy. If this turns out to be the case, this would be evidence in favor of my strong impression that the new

questions really do test deeper understanding, and not just an understanding of something different.

Second, it would also be worthwhile to have some individuals not in control of the students' grades, such as the staff of CTE, interview a representative sample of students about the new questions. Their candid feedback would shed valuable light on how they perceive the questions, what thought processes they go through in answering such questions, what they infer about my priorities from my use of the questions, and how the questions influence or otherwise affect their intellectual development.

I hope, in the future, to use conversation-based questions more and to pursue these empirical investigations.