

MULTIPLE-CHOICE QUESTIONS

Roy L. Bishop

Department of Physics, Acadia University, Wolfville, Nova Scotia, B0P 1X0, Canada

Although few people enjoy tests and examinations, these aspects of education appear to be essential to the maintenance of standards, and they provide a vital learning environment for both students and instructor. Opinions regarding “objective,” or “multiple-choice,” questions range from very positive, which is my own qualified opinion, to very negative. In the latter instance, some claim that such questions should *never* be used, and refer derisively to multiple-choice questions as “multiple-guess questions.”

I shall first address the disadvantages of multiple-choice questions. I begin with a disadvantage that is not well appreciated:

1. It is difficult to make *good* multiple choice questions. The difficulties include the following:
 - (a) One must try to avoid standardized questions and, especially, trivial questions.
 - (b) It is not easy to arrive at unambiguous, concise wording, including all the necessary caveats.
 - (c) Even in questions that are otherwise good, numerical answers that are incorrect can lead students astray.
2. Multi-step problems are difficult to present in a multiple-choice format.
3. Assuming the usual 5 possible answers, random guessing will guarantee, on average, a grade of 20 per cent. (However this grade is too low to be of great concern. Also, a marking scheme that eliminates this possibility would result in some students receiving negative grades! Further, guessing based on elimination of incorrect answers or on a “feeling” for what is correct, deserves some credit.)

What are the advantages of multiple-choice questions?

1. Good multiple-choice questions confront a student with several plausible and often subtle choices. Such situations are more like real-life than those proposed in the usual, essay-type questions where just enough information has been provided to allow the answers to be deduced. The necessity of sorting through plausible sounding alternatives probes a student’s understanding in a special way. A good multiple-choice question goes beyond mere regurgitation. It can stimulate critical thought and/or require an awareness of the real world.

2. Multiple-choice questions make efficient use of testing time. Nearly all of a student's time is spent reading and thinking, rather than writing, and writing, ...and writing. Not only is this more efficient in probing a student's knowledge, but also more material can be examined in the time available.
3. There are some important practical advantages for the instructor:
 - (a) Marking is easy and quick, a tremendous advantage when dealing with large classes. The total time saved is not as great as might first appear, however, for it takes much *more* time to make up good multiple-choice questions than essay-type questions. Nevertheless, making up questions is an interesting, creative endeavor compared to the drudgery of marking hundreds of essay-type answers!
 - (b) The grade for each student is quantitative and unambiguous (provided the questions are good ones).
 - (c) The instructor does not have to decipher students' writing! (Some, however, will rightfully bemoan the loss of a chance for students to practice written expression.)
4. Analysis of incorrect answers is relatively easy and can be quantified. Such analysis yields insights into the reasoning processes of students and suggests ways that teaching can be improved. For example, in the paper: "Constructing Objective Tests" by Aubrecht and Aubrecht (*American Journal of Physics*, 51, pp. 613, July 1983), two indices of question quality are defined:

"Difficulty Level": The average grade on any one question should be about 65 per cent. An average of 100 per cent or 20 per cent (in the case of a question with 5 possible answers) indicates a question useless for testing purposes.

"Discrimination Index (DI)": Pick the top 27 per cent (=N) of papers ("highs") and the bottom 27 per cent (=N) ("lows"). For a given question, the number of highs that get it correct = H, and the number of lows that get it correct = L. DI is defined as $(H-L)/N$ and ranges from +1 (perfect discrimination) to -1 (poor question!). DI should be at least 0.3.

I conclude by presenting 6 examples of multiple choice questions at the introductory astronomy level. The first 3 are poor, and the second 3 are good:

P1 Space debris is of most interest to astronomers for which one of the following reasons?

1. It contains much gold and silver.
2. It all came from the Moon.
3. It all came from comets.
4. It tells us about the early Solar System.
5. It tells us about the last ice age on Earth.

Comment: The question itself is not good because the term “space debris” is not defined (fragments of satellites in near-Earth orbit? Interplanetary dust and meteoroids?). Also, answers (1), (2), and (5) should be obviously incorrect to even a relatively poor student, so a guess at the remaining answers would mean a 50% chance of choosing the correct answer.

P2 During which season of the year is Earth closest to the Sun?

- (1) winter (2) spring (3) summer (4) fall (5) Earth is always the same distance from the Sun.

Comment: The instructor who composed this question has ignored both the opposite seasonal pattern in Earth’s other hemisphere and the absence of mid-latitude seasons in the tropics.

P3 The surface temperature of Venus’ atmosphere is most accurately described as:

1. as hot as the surface of the Sun.
2. twice as hot as Earth’s.
3. about as hot as Earth’s.
4. half as hot as Earth’s.
5. less than a tenth as hot as Earth’s.

Comment: This is an almost trivial question. Answer (1) is ridiculous, and a knowledge of the order of the planets from the Sun (which is taught at elementary school) would rule out answers (3), (4), and (5).

G1 Suppose you are in Williamstown. If, toward the northwest, you see a first quarter Moon near the horizon, what month is it?

- (1) August (2) March (3) December (4) June (5) September

Comment: This question is not trivial, and it does not involve regurgitation of standard material. The student must know about: (i) the phases of the Moon and the associated Sun-Earth-Moon geometry; (ii) the inclination of the ecliptic; (iii) that the Moon’s orbit lies (approximately) near the ecliptic; (iv) the possible range of compass directions of the intersection of the ecliptic with the horizon at mid-northern latitudes; and (v) how the Sun’s position on the ecliptic is related to the time of year. Furthermore, the student must be able to correlate these various things and fit them to the question. In brief: A low Moon in the northwest means that the summer solstice point (most northerly point) of the ecliptic must be in that vicinity. The first quarter phase means that the Sun must be about 90 degrees further westward along the ecliptic ...near the spring equinox. Therefore the answer is option (2).

G2 In which stage of the life cycle of a single, isolated, massive star does it have the least mass?

1. a dilute cloud of collapsing gas and dust.
2. a T-Tauri type star.
3. a main sequence star.
4. a red giant star.
5. a black hole.

Comment: The wording of the question rules out significant accretion of material by the star from a dense region of the interstellar medium after it has formed. The answer (5) is obvious provided the student knows about energy/mass conservation and has an appreciation for a star: namely, that it continually loses photons, protons, electrons, helium nuclei, *etc.*, to the dark, cold void surrounding it. Vague misconceptions about black holes (5) can lead the less able student to avoid this answer.

G3 Most locations in space (*i.e.*, locations selected at random anywhere in the universe) receive an amount of light

1. much greater than daylight on Earth.
2. comparable to daylight on Earth.
3. comparable to moonlight on Earth.
4. comparable to night-time starlight on Earth.
5. much less than night-time starlight on Earth.

Comment: Although intentionally not obvious at first reading, this question is designed to see if the student has an appreciation of the large-scale structure of the universe: that it consists basically of dim galaxies scattered here and there in the darkness of intergalactic space, and that Earth is located well inside of a galaxy. If this appreciation exists, then the locations specified in the question will clearly be in intergalactic space, and so the answer can only be (5).

In conclusion, I should add that only the best students can handle questions like the last three. A test composed entirely of questions of this level would be a disaster (pun intended) for a majority of the students in an introductory, university-level astronomy course. Tests must have a mixture of simple factual questions, questions that require one step (or two) in reasoning, and a few questions that require more sophisticated patterns of reasoning.

Discussion

M. Zeilik: *Do you do an item analysis on your tests, and if so, what kind of reliability index do you get?*

R. Bishop: I determine the success rate for each question; that is, the fraction of the total number of students who answer correctly. A success rate near 6/10 or 7/10

indicates a good question: neither too difficult nor too easy. Questions below 5/10 are examined for clarity and level of difficulty. If they still appear reasonable, then they indicate topics that need extra emphasis.

I also examine the frequency distribution of the incorrect answers selected for each question. This reveals the sort of errors or misconceptions that students have (and will also reveal inadequate questions — fortunately this does not occur too frequently!).

In addition, for questions that are disasters (success rates of 2/10 or 3/10), I make particular note of the answers given by a few of the best students in the class. Their answers usually indicate that the question was not totally unreasonable, but sometimes their answers will reveal flaws in the question that had missed my attention.

Barrie W. Jones:

1. *At the Open University (UK) we carry out item analysis on every multiple-choice question and find now (with time and care spent on the question setting) a good correlation between a student's overall performance and his/her performance on individual questions.*
2. *We also correlate performance on multiple-choice questions with performance on conventional tutor-marked assignments. Again (with care and time spent on the question setting) we find a good correlation between the two performance indicators.*

William Waller: *I would like to point out the virtues of true/false questions. Like multiple-choice questions, they have the potential of being unambiguous, analyzable gauges of student understanding. The restriction on possible answers, however, helps to focus on what is important in the course. To my surprise, I have found that the true/false format can efficiently probe student understanding — even for astronomy majors in their senior year of study.*

Lon Clay Hill, Jr.: *How confident are you of the actual validity of test items when judged solely by intrinsic internal consistency tests (I have some data that suggest that regard for important naked-eye phenomena is negatively correlated with success with the more academic foci of the classroom.)*

R. Bishop: I am not certain what you mean by “intrinsic internal consistency tests.” I try to examine *all* aspects of the course — theoretical and observational. Particular attention is given to questions that good students cannot answer “correctly,” and to questions that *few* students answer correctly. Only after I have examined the answers of many students to a particular question can I rate that question as being “good” or “bad.”