Using Constructed Response Questions on High-Stakes Assessments in Large Classes With Limited Support

Norris Armstrong* and Sandii Constable2

1Department of Genetics, University of Georgia
2Division of Biological Sciences, University of Georgia

Abstract

Large lecture courses often rely heavily, and sometimes exclusively, on Multiple Choice (MC) exams to assess student learning. Constructed Response (CR) questions, in which students generate their own answers, can provide more insight into student learning than MC questions alone but are seldom used in large courses because they are more challenging and labor intensive to score. We describe a strategy for using CR questions on assessments even in very large lectures courses with the support of undergraduate assistants.

INTRODUCTION

Assessment has been defined as the process by which the quality of an individual’s work on a specific task is used to infer more generally what that person knows or can do. Assessment can be used by instructors to evaluate their own knowledge and skills (self-assessment), by instructors to give feedback during the learning process (formative assessment), or to evaluate an individual’s knowledge and ability after a stage of learning has been completed (summative assessment). Regardless of how it is performed, assessment is a key component of learning because it provides information on an individual’s strengths, where there is room for improvement, and can help instructors revise their teaching methods to better meet a learner’s needs (1).

To correctly infer what an individual knows and can do, it is essential that assessments align with the knowledge and skills being examined and that they accurately measure a person’s abilities. Several authors have argued that multiple forms of assessment are needed to capture different components of learning (1–6), and a variety of self- and formative assessment formats have been developed to do this (7–13). However, summative assessment in large courses often relies heavily on a single format, multiple-choice exams (14–18).

Instructors frequently use multiple-choice (MC) questions for summative assessments because this format is considered to be an objective, reliable, and efficient way to examine learning (5, 6). However, relying primarily on MC questions to assess learning has drawbacks. MC questions do not reveal the thought processes students’ use when selecting their answers and can miss misconceptions if these ideas are not among the available answer options (3, 19–23). MC questions can also overestimate knowledge if a student can correctly guess answers from the limited options available or by using criteria that have little to do with the concepts being assessed (24–26). As an example, Figure 1 shows an MC question that asks students to apply their understanding of chromosome and allele behavior to predict the genotypes of the daughter cells produced by cell division. Cell division is a fundamental process in cell biology and genetics and understanding chromosome/allele behavior is critical for comprehending how genetic traits are transferred from one generation to the next in a wide variety of contexts. We have found that students can often answer this question correctly even when they are unable to explain how cell division takes place or how this process could generate the result they selected (see Supporting File S1 for more detail).

A diploid organism is heterozygous for two genes, AaBb, that sort independently. Which of the following could be the genotype of a gamete produced by meiosis in this organism?

- a) AaBb
- b) Aa
- c) A
- d) AB
- e) All of these are gametes produced by meiosis

Figure 1. An example of a MC question examining cell division that students can often answer correctly without understanding the underlying concepts.

In contrast to MC questions in which answer(s) are selected from a limited number of options, constructed-response (CR) questions require students to generate their own answers. Self-generated answers can give greater insight into student thinking and can better reveal potential misconceptions. This format also makes it less likely that a student can guess a correct answer (27, 28). Figure 2 shows a revised version of the question shown in Figure 1 that asks students to generate their own answer illustrating how chromosome and allele behavior can lead to the production of daughter cells with different genotypes. Responses to similar questions in our classes have revealed that many students have

*Correspondence to: Norris Armstrong, 1000 Cedar St., Division of Biology, University of Georgia, Athens, Georgia, 30602; narmstro@uga.edu

Conflict of Interest and Funding Statement: None of the authors have a financial, personal, or professional conflict of interest related to this work.


Copyright: © 2024 Armstrong and Constable. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Received: 4/25/2023; Accepted: 8/15/2023; Published: 6/11/2024

Editor: Melanie Melendrez-Vallard, Anoka-Ramsey Community College

Conflict of Interest and Funding Statement: None of the authors have a financial, personal, or professional conflict of interest related to this work.

*Correspondence to: Norris Armstrong, 1000 Cedar St., Division of Biology, University of Georgia, Athens, Georgia, 30602; narmstro@uga.edu

Copyright: © 2024 Armstrong and Constable. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.
incomplete or incorrect understanding of one or more different aspects of cell division including how chromosomes line up during meiosis, the difference between sister chromatids and homologous chromosomes, the importance of DNA replication in cell division, and how chromosomes and alleles will appear after replication has taken place (see Supporting File S1 for more detail).

A diploid organism is heterozygous for two genes, AaBb, that sort independently. In the space provided below, draw a simple diagram that shows how the arrangement of chromosomes during metaphase of meiosis I allows this organism to produce four different gametes. Be sure to label all alleles.

Figure 2. An example of CR question examining cell division that requires students to generate their own answer.

Another advantage of using CR questions on assessments is that this approach can encourage students to adopt study habits that promote more in-depth learning. MC assessments often assess low-level cognitive skills such as simple recall (17, 29, 30) and, even when they do measure higher order skills, MC format exams can prompt students to adopt strategies that promote low level learning. In contrast, CR questions are believed to encourage more active, higher-level learning approaches (24, 30). Indeed, we have observed that students often try to answer MC practice problems by eliminating some answer options and then selecting from the remaining choices using criteria that are unrelated to the concepts being assessed. However, when asked to generate their own answers, students were more likely to think through and apply relevant concepts in order to develop a response. As a result, including CR and other question types on exams may promote more effective learning than using MC-based assessments alone.

Despite the potential benefits of using CR questions, large classes generally do not use this format because the responses generated by students can be more labor intensive and difficult to score. Below, we describe how we have been able to use and score CR questions on summative assessments in very large enrollment courses with careful planning and the help of undergraduate assistants.

QUESTION DESIGN

We have found that successfully using CR questions on assessments in large courses depends heavily on the quality of the questions being used and by the length of the answers these questions prompt. Questions that are unclear or complicated tend to generate unclear responses that require more time, effort, and interpretation to score. Similarly, questions that prompt lengthy responses tend to be more difficult to score (31) and can be made even more challenging when students write longer answers than needed.

Instructors can help ensure that the exam questions they write are straightforward and clear with a couple of simple steps. One is to prepare a clear, detailed scoring rubric that includes examples of possible answers before the assessment is administered (31–33). Questions for which it is difficult to write brief, clear answers on the scoring rubric are almost certain to generate messy student responses. Similarly, questions that have multiple acceptable answers should generally be avoided as these can require more time and effort to interpret and score. A second useful step to ensure question clarity is to have another instructor or a trusted graduate or undergraduate student read over the assessment. A fresh set of eyes can identify questions that might seem clear to the instructor but could be confusing to students or that may have alternative answers that the instructor did not anticipate.

Another approach to discourage excessively long answers to questions is to provide instructions at the start of an assessment, or in the questions themselves, letting students know how long their answers should be and/or where they should be placed. For example, instructions could indicate that an answer should not extend beyond a given length or should fit within a space provided with each question. Long and messy answers can also be avoided by assessing different aspects of complex concepts separately. Assessing individual parts of a complex concept generates answers that are easier to score and also make it easier to identify where students have difficulty (31–33). Breaking questions that assess complex concepts into multiple parts also allows instructors to award partial credit when students understand part, but not all, of concept. Having students draw or annotate simple figures or diagrams rather than using words may also generate answers that are easier to score (34). Most CR questions on our assessments can be answered using a few words, 1–2 sentences, or simple figures (see Supporting File S2).

It could be argued that CR questions that require very brief answers are no better than MC questions. However, we have found that answers consisting of as little as a single word, number, or a simple figure can be very effective at assessing students’ understanding of concepts and can reveal different types of errors and misconceptions they may hold. For example, the question shown in Figure 3 asks students to apply their understanding of DNA/RNA structure and replication to predict the results of a chemical reaction central to both cellular processes and biotechnology. Though this question requires only a brief answer, students have submitted almost a dozen different responses to this type question that enable us to determine if they understand how primers and template DNA function, how DNA polymerase acts, and the importance of DNA directionality during replication. CR questions have also enabled us to ask a broader range of questions than is possible using MC questions alone and has given us deeper insight into student learning and understanding.

A short RNA primer pairs with a template strand of DNA as shown below. In the answer box, write out the sequence and directionality of the nucleotides that would be added by DNA polymerase enzyme.

5’ – GCTTAATCGGACGAG – 3’
3’ – UAGCC 5’
3’ CGAAT 5’

Figure 3. An example CR question with a brief answer that can reveal multiple types of errors.
USING UNDERGRADUATE ASSISTANTS TO HELP EVALUATE CR QUESTION RESPONSES

Scoring even brief responses to CR questions in large courses is a labor-intensive process that instructors would find challenging to do by themselves. One potential solution to this problem is to enlist the help of undergraduate students in the evaluation process. Using undergraduates to support classroom activities and learning is well-established. Undergraduate students are often asked to work together on class activities, can work as classroom assistants to help instructors support the learning of other students, and schools often offer ad-hoc tutoring programs that are staffed largely by undergraduates (35–37). Methods have been developed that ask students to evaluate work completed by their peers (38) and are popular enough that commercial software has been created to facilitate this technique (39, 40). In some schools, undergraduates can even serve as teaching assistants who independently teach, manage, and assess students in laboratory and recitation sections (41–45). We have found that Undergraduate Assistants (UAs) can do an excellent job helping to evaluate responses to CR questions on assessments in large courses.

If undergraduates help with assessment, recruiting qualified candidates is important to ensure that assessment questions are scored reliably and accurately. There is no single trait that is critical for a student to be a good UA, but we have found that students who have previously taken and done well in the same course they will support or have completed more advanced courses in the same discipline will have a good understanding of the concepts being taught and will be familiar with the types of CR questions typically being asked. Students with prior work experience or have held positions of responsibility, such as leadership positions with campus groups, tend to be hard working, dependable, can follow instructions well, and are usually good at working independently.

The number of UAs needed will depend on the size of the course they will assist, the number of assessments that need to be evaluated, and the number and type of CR questions on each assessment. We recruit enough UAs to ensure that scoring can be completed within a week of administering an assessment and our goal is to require less than 10 hours of time from each UA for each assessment. We also recruit at least one more UA than the minimum needed so that we can accommodate the UAs’ own schedules and still score the assessments in a timely manner. For example, our courses enroll about 600 students a semester with four to five midterm exams each with 10–15 CR questions that can be answered with a single word/number, 1–2 sentences, or a simple figure. To score these assessments, we need a minimum of 3 and typically recruit 4–5 UAs. Courses with different enrollments and assessments may need a different number of UAs.

SCORING ASSESSMENTS

Scoring CR questions, even with UA support, can be time consuming but is manageable if well organized. Below are some of the steps that we have found to be most important.

- Students are asked to enter only their initials and student ID on their assessments. This reduces the likelihood that individual assessments can be identified during scoring.
- The UAs score the assessments in or near the instructor’s office so that the instructor is readily available to answer any questions and can address problems as soon as they arise. The assessments should also be securely stored in the instructor’s office reducing concerns about privacy and academic integrity.
- The same person scores all responses to a given question. This speeds up and improves the consistency of scoring. Exceptions can be made for questions that are straightforward to score.
- The instructor creates a detailed rubric that includes correct answers, common incorrect answers, and point values that the UAs will use to score the assessments (8, 46) (see Supporting File S3).
- All UAs use same notation system for scoring (e.g., awarding points for correct answers rather than deducting for incorrect). This facilitates tallying point totals for the assessment after scoring has been completed and makes it easier for students to interpret their results when the assessments are returned.
- The UAs are discouraged from leaving written comments as these can greatly slow down the scoring process. Instead, the instructor can provide students with the scoring rubric and/or an assessment key explaining how the assessment was scored and addressing common errors.
- The UAs are asked score answers that are either clearly correct or incorrect first and to set other answers aside. The UA can then sort ambiguous answers into groups based on the type(s) of errors they display and discuss with the instructor how each answer group should be scored, increasing the speed and consistency of scoring.
- Establish a system for identifying which questions and assessments have been scored and which have not. This is especially important if the UAs work at different times or for very large courses when it may not be possible to fully score a question in one sitting. Labeling partially scored assessments with sticky notes has worked well for us.
- We print adhesive labels with each students’ name and student ID using the Microsoft Word Mail Merge function and affix these to assessments with the same ID after scoring has been completed. We then enter students’ scores for each assessment question into a spreadsheet. Scoring individual questions allows us to better evaluate different parts of the assessment and to identify concepts that students find more challenging.

RETURNING ASSESSMENTS

We recommend scanning and storing all assessments electronically after which the original assessment (or scanned copy) can be returned to the students. The instructor’s copy serves as a record of the original submission if students request that their assessment be rescored and to help encourage academic integrity (47).

CLOSING REMARKS

We have used the process described above to include CR questions on assessments even in very large courses without the need for graduate teaching assistants. The UAs did an excellent job helping us to evaluate student responses and provide a unique perspective how students interpret CR questions. Discussing the questions and student responses with the UAs
improved how we prepare and score our assessments and enabled us to use a wider variety of question types. Using CR questions has also helped us better determine where students have difficulty and to identify the types of errors they tend to make. We have been able to use this information to change how we teach different topics to better help students understand and master the concepts being examined.

A concern about CR questions is that scoring these questions is more subjective and potentially less accurate than the scoring of MC questions. This has led some to some debate as to whether CR questions should be used for summative assessments (48), especially when scoring is performed by individuals who are not content experts and are more likely to make mistakes (42, 49). However, research has shown that student work can be evaluated accurately by non-experts, including other students, if the assessments are well designed, have defined scoring rubrics, and if the evaluators are provided with appropriate guidance and support (28, 33, 50–54). We have found that UAs can score CR questions efficiently and effectively when provided with guidance and a detailed rubric. When problems do occasionally arise we have been able to easily address them by discussing the questions and student responses with the UAs during the scoring process. By providing students with a detailed scoring rubric that they can compare against their own answers and allowing them to request that the instructor rescore a question, any errors that are made can be identified and corrected. Regrade requests that we have received on our assessments indicate that errors occur infrequently and at similar rates for questions scored by UAs and by the instructor.

Additional concerns about having undergraduates evaluate the work of other students include privacy and the risk of potential bias during scoring (50, 54–57). FERPA regulations allow peer evaluation of student work (57), but some students may not be comfortable if peers who are not part of the course are able to view their assignments. Bias during assignment scoring can take many forms, can affect student scores both positively and negatively, and even the perception of possible bias during scoring can damage students’ trust that their work is being assessed fairly and accurately. Both privacy and bias concerns can be minimized through some of the steps described above. Having students only provide their initials and student identification numbers on assessments reduces the likelihood that assessments could be associated with specific individuals until after scoring has been completed. We further reduce the possibility of bias during scoring by having each UA score only part of each assessment, by providing students with a key and/or scoring rubric when we return their assessments, and by allowing students to request that a question be reevaluated by the instructor if they believe that their response has been assessed incorrectly.

Many of the steps to facilitate the scoring of CR questions we describe above can now be managed electronically. For example, the software GradeScope (58) enables instructors to create and adjust a scoring rubric while evaluating student responses, can sort students’ responses to individual questions into related groups, assign scores to responses in sorted groups en masse, securely return scored assessments to the students electronically, and can facilitate student regrade requests. The software can even presort very short answers (single words or numbers) into groups automatically. As a result, this software can enable instructors to use CR questions on assessments in moderately sized courses even without the aid of Undergraduate Assistants or for very large courses with undergraduate support.

It is important to note that, even with the help of software and Undergraduate Assistants, the use of CR questions in large courses still requires a good deal of effort. Receiving and incorporating feedback to early assessment drafts, preparing scoring rubrics, organizing and scoring CR questions, and responding to student regrade requests requires a considerable amount of time. However, careful planning along with and the help of UAs and/or software can significantly reduce the instructor’s workload and the increased insight into learning provided by students’ responses makes the extra work involved in using CR questions on assessments well worth the effort.

SUPPORTING MATERIALS

- S1. Using CR Questions – CR Question Benefit to Instruction
- S2. Using CR Questions – CR Question Answer Limits
- S3. Using CR Questions – Example Scoring Rubric

ACKNOWLEDGMENTS

We would like to acknowledge the UGA Biology Division and the UGA Office for Instruction who provided support for our efforts to develop new assessment formats for our large courses and the undergraduate assistants who have helped us implement and refine this approach over several semesters.
REFERENCES


