January 11, 2014

Grading by Response Category: A simple method for providing students with meaningful feedback on exams in large courses

Cassandra Paul, *San Jose State University*
Wendell H. Potter, *University of California, Davis*
Brenda Weiss

Grading by Response Category:
A simple method for providing students with meaningful feedback on exams in large courses

Cassandra Paul, San Jose State University, San Jose, CA
Wendell Potter, University of California-Davis, Davis, CA
Brenda Weiss, Davis, CA

As instructors, we want our students to develop a deep understanding of course material, and feedback is essential in their sense-making process. Providing effective individualized feedback to students in large courses is especially difficult. While researchers suggest,[1] and many instructors of large courses are,[2,3] incorporating interactive techniques that allow peer feedback, studies have shown that it's important for students to also have direct feedback from the instructor.[4] Since the requirement for individualized feedback is difficult to meet during class time in large courses, providing effective feedback on exams and quizzes takes on added importance. Some instructors choose to assess their students using open-ended written exam items that require students to demonstrate their understanding of physics by solving a problem and/or explaining a concept. Because grading these items can be time consuming, the challenge is to develop an approach to grading and provision of feedback that is both efficient and effective.

This paper describes Grading by Response Category (GRC), an effective approach to evaluating assessments that provides feedback to students, improving the learning process for both students and teachers by encouraging students to reflect on their thinking and giving instructors information on student difficulties. GRC is a method of grading quizzes and exams utilizing good feedback practices that is especially suited for large courses because it can be done within a time period similar to traditional grading. As we describe the GRC process, we examine the benefits of the GRC method in the context of Carol Evans's recent review article and resulting "Principles of Effective Assessment and Feedback"[5] (Table I). When we describe a feature of GRC that aligns with Evans's findings, we indicate the principle it exemplifies by its number listed in Table I. Finally, we provide an example of the GRC feedback provided to students and discuss student reception of the GRC process.

The GRC method was developed at the University of California, Davis (UC Davis) for use in 300+ student active-engagement introductory physics courses.[2] It has been in use for over 15 years, and is currently being used by dozens of instructors at UC Davis and other institutions. GRC is used to assess exam items that require students to reveal their thinking, often by drawing diagrams or writing explanations, as well as those requiring calculations. Once student quizzes or exams are collected, an instructor follows the steps outlined below to implement the GRC method.

STEP 1: Examine a sample of student work

The instructor views 20-30 student responses for an exam item to get an idea of common student responses. Instructors often find it helpful to place sticky notes on exams illustrating responses that demonstrate particular misconceptions, errors in analysis, lack of understanding, and mistakes as they sort solutions into different piles.

Table I. This table includes a partial list of the "Principles of Effective Assessment and Feedback" from Evans 2013. Evans organized the principles into "themes" as indicated by the letters a-e. For the sake of brevity we chose not to define those themes, and instead we label each principle in the theme with a Roman numeral that refers to the order in which it is listed in the Evans paper.

a-ii	Methods of assessment and feedback are constructively aligned with learning objectives.
a-vi	Feedback informs the process of learning, encourages reflection, and focuses on the self-regulation level.
b-i	Guidance is explicit in relation to requirements of the assessment and what quality is. It demystifies the assessment process through use of exemplars, modeling aspects of good practice, clarifying assessment criteria, and giving clear signals about good practices.
d-ii	Feedback enables development of self-assessment skills. Feedback is not so detailed and specific that it scaffolds the learning so completely that the students do not think for themselves. Feedback is accessible to the learner, for example, provided in conjunction with the learner having sufficient knowledge of how to be able to use feedback effectively.
e-i	Feedback is appropriate to the purpose of the assessment task and level of student understanding.
e-ii	Feedback focuses on the specific features of the task. It focuses on the what, how, and why of a problem rather than simply indicating to students whether work is correct or not.
e-iii	Feedback focuses on performance.
e-iv	Feedback provides suggestions on how to improve rather than focuses on the personal attributes of the learner.
e-v	Feedback identifies actions including strategies to enable the student to improve.
e-vi	Feedback involves an equitable dialogue between student and tutor to clarify meanings, expectations, misconceptions, and future actions.
e-vii	Feedback encourages positive motivational beliefs, self-esteem, and trust.

DOI: 10.1119/1.4897587

–Why we do it: The GRC method allows an instructor to focus on the features of the task, and not assign points based on a binary system of correct or incorrect [principle e-ii]. It's easy to assign points based on how many steps the students completed correctly, but this doesn't help the students figure out why they were wrong, or what they need to think about so they don't make the same errors. Sampling student responses sheds light on student thinking and allows us to categorize errors based on line of reasoning rather than percent correct.

STEP 2: Assign categories

As the common errors become evident, instructors define a category for each error or set of similar errors. Note that a solution cannot be placed in two categories simultaneously. Each category should be exclusive. Many instructors find it useful to make a stack of student solutions for each category.

It is important that these categories are based on what the instructor thinks is important for the students to reflect on; therefore, a category is defined by the most significant error made. For example, if two student responses to a question on motion show a similar violation of conservation of energy, their solutions would be placed in the same category, even if one included an additional minor mathematical error. In contrast, a student response with a correct solution except for a minor mathematical error would not be placed in the same category as the student who solved it completely correctly. This is not to say that categories can consist of only one error. In situations where more than one conceptual error can be identified, a category can represent multiple errors. For example, a category could represent a violation of conservation of energy *and* an additional misconception regarding potential energy.

–Why we do it: Effective feedback aligns scoring with learning goals [principle a-ii]. In the example described above, the instructor is more concerned that students understand fundamental ideas about energy than whether students can carry out algebraic steps without error. The students who make only math errors are forced to consider them. However, students who make major conceptual errors in addition to math errors are given feedback that emphasizes the conceptual issues. This allows students to concentrate on correcting their most grievous errors as determined by the instructor, rather than focusing on the minor details of the procedure [principle d-ii].

STEP 3: Score each student solution with a symbol

Each category is given a unique symbol (usually a letter to facilitate grade entry). Only this symbol is written on the student's exam. Each exam problem or question is assigned its own symbol. Thus, a student exam with three items would be marked with three symbols; e.g., XLP. Categories and corresponding symbols are written down so there is a record of what each symbol means.

–Why we do it: Exams are often handed back at the end of class when students are rushing out the door. When students are given a numerical score on their exam, many will have an emotional response. A subset will be satisfied with their score and never look at it again. Another subset will be disappointed and also not want to look at their exam again. Denying students the instant knowledge of their numerical score allows them to consider their performance when they have the resources (time and categories) to make sense of what they did wrong. Evans cites several studies finding that student emotional states affect their ability to reflect on the feedback they receive.[6] This method allows students to select the time for reflection, which can make the experience more positive [principle e-vii].

STEP 4: Assign numerical values to each category

Once all of the student solutions have been assigned symbols, the written list of categories and symbols is reviewed, and each category is weighed against the learning outcomes for the lessons the exam covered and the effectiveness of the instructional design in enabling the students to achieve those goals. The numerical scores that are assigned to each category are not written on the student exam papers.

–Why we do it: The reasoning behind this is similar to that outlined in Step 2. We rank the categories from best to worst based on our individual learning goals and award grade points accordingly [principle a-ii]. This feature also affords us flexibility in the grading process. After seeing all the solutions, instructors sometimes figure out that there is a reason students are making an odd error, such as poor item wording or a typo, or they realize that there is valuable reasoning in a particular type of error repeated by many students. The GRC method saves the grader from choosing between regrading the whole set or giving some students a lower score than they deserve. This feature is also especially useful to instructors who have graders do the actual sorting into categories, but want to reserve for themselves the final say on the students' scores.

STEP 5: Describe categories for student consumption

The complete solution that meets expectations for full credit is written out for the first category. Each category description thereafter includes a short description of the error and may also include some form of instructions on how to improve. For example, questions designed to get students to confront their preconceptions or instructions of ways they can check their work can be included. The point value for each category is clearly indicated next to each category description. Some instructors indicate the number or percentage of students in each category.

–Why we do it: Categories that are well described give students feedback specific to their error and so are arguably more useful than solutions alone. If the category descriptions include examples of metacognitive questions, the GRC method is aligned with Evans's principle of effective feedback concerning self-regulation [principle a-vi]. Well-written categories will include not only information about the error [principle b-i], but also include instructions on how to improve [principle e-v] and examples of questions that can be asked to facilitate thoughtful reflection on the problem-solving process. In our experience, this is the most difficult part of the process for first-time implementers of GRC to master.

STEP 6: Record student categories and numerical scores

The symbols for each student exam are entered into a spreadsheet or database, which associates the symbols with grade point values assigned by the instructor based on her/his criteria. The instructor has a record of the students' numerical scores, but these numerical scores (grades) are not immediately posted online for student viewing.

–Why we do it: Evans indicates that effective feedback involves a dialogue to clarify expectations, misconceptions, and future actions [principle e-vi]. The first two of these are addressed in the previous step, but determining the future actions of the student and instructor requires reflection by the instructor as well as the student. The GRC method allows an instructor to have a detailed quantitative report of all student abilities on every problem of every exam given throughout the semester. With GRC, the grade spreadsheet can easily reveal the number of students who make the same error on any given problem, which gives the instructor information regarding classwide difficulties. Using traditional grading methods an instructor might be able to determine the average score for each problem but only if each numerical grade is entered separately for each problem, and this tells nothing of the types of errors students are making and if these errors appear in later problems. Information on common errors can help the instructor address areas where students are struggling, modify teaching activities, and improve future feedback accordingly.

STEP 7: Give students access to category descriptions with numerical scores

The instructor posts the categories with their associated grade values online and returns the students' exams with only the symbols on the top, so that the students are required to actually look at the category definitions in order to calculate their grade. The exams are returned on the same day the categories are posted.

–Why we do it: A detailed report of the categories gives students timely access to feedback and demystifies the grading process. Every point deduction is accounted for and clearly explained in context of the instructor's learning goals [principle b-i].

STEP 8: Provide an avenue for regrades

After the exams are returned, if a student feels that her response was incorrectly categorized, she can request a regrade. The student fills out a form stating the category she was placed in and what category she believes she should be placed in, or why she believes that the particular category should receive more points. She must also solve the problem correctly. The student staples her exam to the form and turns in both to the instructor. The instructor reviews the student's work and determines if and how the student's grade should change.

–Why we do it: By providing an avenue to contest a grade, we involve students in the assessment dialogue, which has several benefits [principle e-vi]. The regrade process relieves some of the agony of assigning a category in cases where it is difficult to follow student reasoning. If instructors know the students can request regrades through a pretty painless process if the students understand the material, they are less likely to spend an inordinate amount of time trying to decipher student work. This puts the "burden of proof" on the students to show their understanding instead of the instructor to fill in the gaps. It's been suggested that shifting this burden to the students can encourage students to explain their reasoning in future exam situations.[7] In the authors' experience students are less likely to challenge a grade with this method, but whether this is because students understand why they received the score that they did or because effort is required of them to prove their understanding has yet to be investigated.

Example

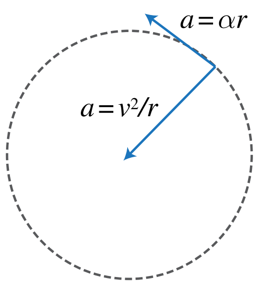
We show below an example of a recent exam item along with the categories that were developed from the student responses. The categories, the grade points assigned to that category, and the percentage of students assigned to that category were posted online for the students in a form similar to that seen here. Students calculated their grades using this information along with the categories marked on their exams. The numbers in parentheses are the grade points assigned to each response category.

Exam question:

There are two different equations listed on the equation sheet for acceleration under "linear/angular relation": $a = \alpha r$ and $a = v^2/r$. <u>Draw a picture of these two vectors and explain how they are related to each other.</u> Are they the same? If so, why? Are they different? If so, how?

Category definitions and grade point values:

- **Q (4.0):** Complete response: The two pieces of the acceleration vector are perpendicular to each other. $a = v^2/r$ is the magnitude of the centripetal acceleration and is pointed toward the center of the circle an object is rotating about. $a = \alpha r$ is the magnitude of the tangential acceleration and is tangent to the curve that the object is traveling in. (27%)



- **M (3.5):** Same as Q but mixed up the vectors. Make sure you check that your answers make sense in relation to the variables in the relationships. (6%)
- **L (3.2):** Ideas were mostly complete and the same as in Q, but the pictures were incorrect or unclear. Practice communicating your ideas before the next quiz. (Must discuss the tangential and centripetal acceleration to be in this category.) (10%)
- **C (2.5):** Some correct information was given regarding these quantities, but drawings were incorrect or unclear. The definitions of the vectors (as indicated in Q) were not given, and the fact that the vectors are perpendicular to each other was not indicated. Consider what would have made your solution more complete. What made your diagrams incorrect? (6%)
- **P (2.0):** Argued that one of the equations represented an angular acceleration and one represented a linear one. The "*a*" in each equation represents a linear acceleration. Consider what α represents. (14%)
- **S (1.0):** Argued that the acceleration equations were the same, but you just use them in different instances. If this were the case, $\alpha r = v^2/r$. Does your answer agree with what you know about physics? (22%)
- **N (0.5):** Some thoughts shown but did not successfully complete the question. (10%)
- **Z (0.0):** Blank or essentially blank (4%)

While quantitative research is pending, student response to GRC is positive overall. When asked about the GRC method, most students will identify at least one reason they like it more than grading techniques used in their other science courses. The only complaint students have about GRC is that they don't immediately have access to their numerical score.

– *"I liked how you put the letters instead of numbers on the test so that we could get more feedback without all the red marks on the physical tests." [principles e-ii, e-vii]*

– *"I think in general the system is nice because not everyone really wants to show off their grades straight up to everyone." [principle e-vii]*

– *"I felt that it … helps the students get a better grade than they … would have if you used a more traditional method. It also allows the teacher to get a better idea of where the students are having trouble; this can help you help us." [principle e-vi]*

– *"From the [categories], I was able to understand where I had made mistakes and it helped guide me in the right direction to solve the problem. The grading system made me also look closer at the exam as opposed to looking at my grade or percentage and hiding it in my binder." [principle a-vi, d-ii, e-iii, e-iv]*

We believe that assessment is the best way instructors can convey expectations to students, and that students continue to learn after the assessment is given. With the Grading by Response Category method, we not only provide our students more feedback for improvement, but that feedback is matched and weighted to appropriately communicate the relative value of the skills and concepts they learn in our class.

For resources and further instruction on the Grading by Response Category method, please visit our website: http://www.sjsu.edu/people/cassandra.paul/gradingbycategory/.

References

1. David Nicol, "From monologue to dialogue: Improving written feedback processes in mass higher education," *Assess. Eval. High. Educ.* **35** (5), 501–517 (2010).
2. Wendell Potter, David Webb, Cassandra Paul, Emily West, Mark Bowen, Brenda Weiss, Lawrence Coleman, and Charles De Leone, "Sixteen years of collaborative learning through active sense-making in physics (CLASP) at UC Davis," *Am. J. Phys.* **82**, 153–163 (Feb. 2014).
3. Catherine H. Crouch and Eric Mazur, "Peer Instruction: Ten years of experience and results," *Am. J. Phys.* **69**, 970–977 (Sept. 2001).
4. Carol Evans, "Making sense of assessment feedback in higher education," *Rev. Educ. Res.* **83**, 70–120 (2013).
5. Ref. 4., pp. 80–83.
6. Ref. 4, p. 95.
7. C. Henderson, E. Yerushalmi, V. Kuo, P. Heller, and K. Heller, "Grading student problem solutions: The challenge of sending a consistent message," *Am. J. Phys.* **72**, 164–169 (Feb. 2004).

Cassandra Paul *completed her PhD in physics at UC Davis in the area of Physics Education Research (PER). She is currently an assistant professor at San Jose State University in California and a member of both the Department of Physics & Astronomy and the Science Education Program. She is currently taking part in reform efforts at SJSU by teaching using the CLASP curriculum, an interactive model-based curriculum. Her research interests include the areas of assessment and instructor professional development. She is the co-creator of the Real-time Instructor Observing Tool (RIOT), and has recently received NSF funding to develop the Student Participation Observational Protocol (SPOT) to assist STEM faculty in making research-based changes to their teaching practice.*
cassandra.paul@sjsu.edu

Wendell Potter, *now officially retired from the Physics Department at UC Davis, began his faculty career there in 1970 as an experimental solid-state physicist, but switched fields mid-career to science/physics education. He led the development of the creation of the reformed introductory physics course for life science majors at UC Davis, the CLASP curriculum, which is now taken by >1700 students per year. He continues to collaborate with both science education faculty and high school teachers in the development of interactive engagement and model-based curricula and with university physics faculty in the ongoing evolution of the CLASP curriculum.*

Brenda Weiss *received a PhD in experimental physics from the University of California, Davis. She worked with Wendell Potter's UC Davis team in developing the CLASP curriculum, graded using response categories. She studies how students learn, homeschooling her children, and participating in the UCD PER group.*