A Fresh Look at Grading and Reporting in High Schools

Sandra Herbst
Anne Davies, Ph. D.

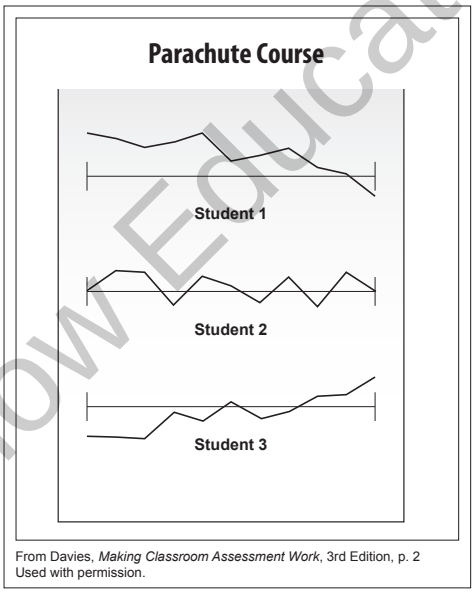
Hawker Brownlow
EDUCATION

Contents

1. A Parachute Course

In *Making Classroom Assessment Work* (2011), Davies uses an illustration first shared by Michael Burger (see figure I-1). Consider this: Three students are taking a course on how to pack a parachute. Imagine that the class average is represented by a dotted line. Student A initially scored very high, but his scores have dropped as the end of the course approaches. Student B's evaluations are erratic. Sometimes he does very well, and sometimes he doesn't. The teacher has a hard time predicting from day to day how he will do. Student C did very poorly in relation to the others in class for the first two-thirds of the course but has finally figured out how to successfully pack a parachute. Which of these students would you want to pack your parachute? Student A? Student B? Student C? Most people would choose Student C because they want the chute to open successfully; after all, a parachute course is standards based or outcomes based. The problem is that traditional thinking about grades and evidence of learning was in effect, so all numerical data was valued equally, with no professional judgment being made about which numerical data should be included and why. So, in the past, Student C did not pass the course. When his grades were tallied and averaged, they weren't high enough. Student A and Student B did pass.

Figure I-1



From Davies, *Making Classroom Assessment Work*, 3rd Edition, p. 2
Used with permission.

2. An Employment Selection Process

Consider the job selection process for a teaching position for which you are figuratively applying. At the end of the process, a person or a panel of people determined whether you were suitable for the position. Were you valued as the best candidate? What you did up to that point is the foundation of the final decision. Were your references positive, based on others' interactions and experiences with you? Were your student teaching blocks or term positions or other permanent positions successful? What were the opinions of others regarding your work? Did your college coursework reflect both a commitment to and an understanding of the content? Did the interview demonstrate your depth of knowledge of your discipline, of the developmental needs of students, of authentic ways to engage your learners? Did your professional portfolio provide evidence of your knowledge, skill,

What About...?

Tests and quizzes are the only way to make sure that our assessment is objective and fair. All this other stuff is too subjective.

Tests and quizzes are often viewed as objective, but, through their construction, elements of subjectivity result. Tests cannot measure everything that a curriculum demands that students do, create, and articulate. Instead, we need to ensure that our assessment and evaluation are reliable and fair. This builds equity, accuracy, and repeatability.

(To read more about this, see "What About 1-2?" on page 79.)

When we begin with the end in mind and explain to students what they need to know, be able to do, and be able to articulate, we set students up for success. When teachers coconstruct criteria or use samples to show a range or variety of acceptable work, they encourage students to represent what they know in different ways, while still being fair and equitable. For example, if the curriculum standard states that students should be able to "describe how human curiosity and needs have influenced science, impacting the quality of life worldwide," they could research and write about it, or they could draw a mind map and create a digital presentation with visuals that illustrate effect and change over time, or they could interview scientists to track questions that were starting points for scientific inquiry and map the impact of answers on people and environment, or they could use newspaper and magazine clippings to illustrate key ideas. The expected learning does not change, but what students can do to demonstrate their learning in relation to the standard can be different. This is a key concept for ensuring that all students show what they know using evidence of learning collected from multiple sources over time. Further, this process helps students come to know and use the language of assessment, which can then be used by students and others (e.g., educational assistants, student support teachers, and parents) to give specific, descriptive feedback during the learning, so that they can self-monitor and self-regulate.

SOT2816 • 9781760012816 • © 2014 Hawker Brownlow Education

Figure 1-13

Possible Observations to Show Evidence of Learning

***Digestion and Nutrition*—A Biology Unit of Study**

- Dissection skills during real or virtual dissections
- Collaborating with others to achieve group goals and responsibilities
- Attention and focus during demonstrations
- Confidence to carry out investigations
- Science lab skills
- Attention to personal safety, the safety of others, and the environment
- Appropriate selection and safe use of scientific equipment
- Synthesis of information and the evaluation of the quality of sources
- Interaction with and examination of data to identify its relevance, reliability, and adequacy

Collecting Products

Teachers collect various kinds of evidence to show what students can do in relation to the standards or outcomes for the course or grade level. These could include projects, assignments, notebooks, and tests. There are many ways for students to represent their learning, and if we as teachers prescribe the evidence of learning, we may unnecessarily limit students' options to show what they know. (See figures 1-14, 1-15, and 1-16 for different possibilities that represent learning.)

Figure 1-14

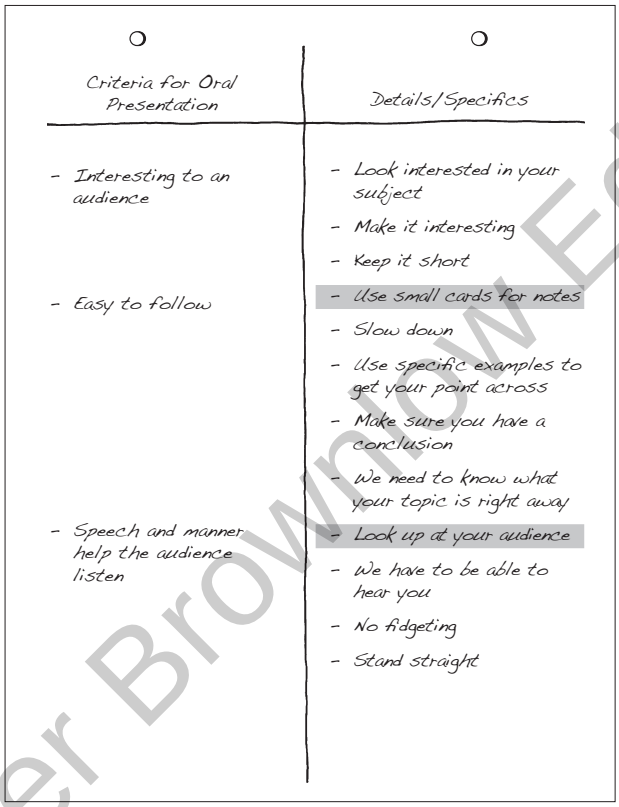
Some Possible Ways to Represent Learning

- Draw a diagram.
- Make a timeline.
- Make a poster.
- Write a story.
- Build a website.
- Do an oral presentation.
- Create an animated movie.
- Create or add to a Wikipedia entry.
- Build a model.
- Create a puzzle.
- Conduct a survey.
- Make a podcast.
- Produce a YouTube video.
- Make a recording.
- Do a report.
- Write a song.
- Interview people from other countries.
- Write a newspaper article.
- Create a song.
- Create a piece of visual art.
- Build a diorama.
- Write a play.
- Start and monitor an online discussion forum.

Figure 2-9

Highlighting Criteria

When students have criteria for a specific assignment, goal setting can be straightforward. We ask our students to highlight any criterion that they did not meet. This criterion becomes their short-term goal (see below).



Adapted from Gregory, Cameron, and Davies, *Self-Assessment and Goal Setting*, p. 52.

In summary, we give students time to learn and time to get it right by helping them understand the learning destination, by coconstructing criteria around process or products, and by involving them in analyzing samples of student work. Then, as they create evidence of learning, teachers engage students in self- and peer assessment and goal setting in relation to the criteria that they have helped to construct. As the learning proceeds, teachers ask students to collect specific evidence of learning in relation to the

SOT2816 • 9781760012816 • © 2014 Hawker Brownlow Education

As mentioned earlier, informed professional judgment results from teaching to the standards or learning outcomes based on a common reporting scale (often decided by the jurisdiction); it is also based on thoughtfully considering samples of student work and collections of evidence, scoring common assessments, and analyzing external test data with colleagues. While a teacher's written and verbal comments may speak to the amount of *growth* students have made in their learning, the evaluation must reflect their progress in relation to the standards for the subject area or course and the grade level at which they are working.

While teachers do not have to base their evaluation decision on the same body of evidence of learning for each student, they must base their evaluation on a reliable and valid collection of evidence of learning. And this evaluation must be equitable—that is, all students, regardless of how they learn, show their learning, or how much they struggle (or not), must have the same opportunities to show proof of learning. A helpful definition of the term *informed professional judgment* is: the professional determination, after a review of evidence of learning present (not absent), of what has been learned and achieved.

We need to stress that adding the scores and averaging them misrepresents the learning that has been accomplished. To evaluate well, we should look at *all* the evidence: observations, products, and conversations. *Triangulation of evidence* is essential because it puts single pieces of evidence into context. Just as a judge in a court of law must examine all the evidence in light of the legal statutes, teachers must look at all the evidence in light of the description of learning based on the standards or learning outcomes. They must consider the entire range of information (all the quantitative and qualitative data): the evidence students have collected, the self-assessments they have made, their observations, criteria-based assessments attached to projects or assignments, and the evidence that teachers have collected, including performance grids, rubric scores, and grades from projects and tests. As teachers examine all the evidence, they are seeking to make the most informed and defensible final professional judgment possible.

It is at this point that many of the "hot issues" currently being debated about reporting cease to matter. For example, whether the evidence of learning was produced in the midst of learning time (formative) or at the end of learning time (summative) isn't an issue. The timing is just information for the teacher. The Assessment Reform Group (2006, p. 10) states, "For summative purposes, common criteria need to be applied and achievement

builds a test of five different questions that have served him well in the past. By selecting questions that they view to be important, these two teachers have built in subjectivity.

Instead of asking ourselves whether our measures are objective, we need to be asking ourselves whether our measures are reliable and valid. Reliability refers to repeatability. Can the student show what he knows in different situations and at different times? Validity refers to the match of the evidence of learning to what is to be assessed—what is to be learned. And to illustrate, let us go back to the example of an oral presentation. We cannot evaluate whether a student can orally communicate ideas to an audience by asking them to complete a paper-and-pencil test. From a classroom assessment perspective, this evidence of learning is not valid, given what was to be learned.

Teacher professional judgment is more reliable and valid than external tests when teachers have been involved in examining student work, coconstructing criteria, scoring the work, and checking for inter-rater reliability (ARG, 2006; Burger et al., 2009).

Inter-rater reliability is defined as the result of learning to make an informed professional judgment. Educators engage in a process of inter-rater reliability when they meet, create quality criteria, and build a scoring rubric for student work. The student work could include, for example, a performance task, a product, observations of application, or a body of evidence. Each educator examines and then scores the student work using a scoring rubric.

At this point, all the scores are examined for consistency among all the educators. The percentage to which they agree is used to determine inter-rater reliability. The higher the percentage of agreement among all the educators' ratings (i.e., the more the scores are similar), the higher the inter-rater reliability will be. This process helps teachers refine and improve their professional judgment.

How else might you respond to this "What About"?

What About 1-3?

> *We need to make sure that our assessment and evaluations are fair, and that means that we need to use all the same assignments, items, tests, and tasks to determine a grade or score.*

SOT2816 • 9781760012816 • © 2014 Hawker Brownlow Education