LEARNING POINT

What is collaborative scoring? Why can it be so valuable?

One of the most valuable professional learning activities teachers can engage in is the formal use of performance standards for common student assessments to gauge the performance level of students. This value can be found in several aspects of their scoring.

What is Collaborative Scoring?

Collaborative scoring occurs when each piece of student work is scored by one or more teachers working independently. Collaborative scoring is a means to assuring that student work is judged reliably by teachers who are trained in the use of high-quality assessment prompts, scoring rubrics, and the collaborative process to be used.

If a group of teacher-scorers are used in scoring, pairing of teacher-scorers may be randomized, so that teacher-scorer 1 may score the work of student 27 with teacher-scorer 6, and then score the work of student 38 with teacher-scorer 4, and so on. Thus, in the course of scoring the work of a group of students, teachers may be anonymously scoring student work with most, if not all, other teacher-scorers.

Why is this important? In more formal scoring sessions, the agreement of each teacher-scorer with other scorers is tracked. This "inter-scorer reliability" is a measure of whether each teacher-scorer is scoring student work in agreement with other teacher-scorers. Teacher-scorers who don't agree with the others sufficiently are given additional training and then subsequently monitored to assure that they score reliably .

What are the benefits of collaborative scoring?

There are several potentially beneficial aspects of collaborative scoring:

1. Teachers are exposed to high-quality assessment prompts that are known to produce a range of student responses sufficient for collaborative scoring.
2. Teachers develop familiarity with the scoring rubrics or scoring criteria that will be used to judge the work of students. Explicit scoring standards might be useful to teachers as they consider what constitutes high-quality student work. In addition, teachers are engaged in an extended discussion of how to apply the scoring rubric or criteria to judge student work.
3. Teachers examine actual samples of student work associated with each performance level of the scoring rubrics or scoring criteria. The samples of student work help teachers not only to more reliably judge the work of students, but also to internalize "success criteria" they can use as they provide instruction to students in the future. The student samples can then be used to conduct assessments in their classrooms. This can prompt teachers to set higher standards for both classroom instruction and classroom assessment.
4. The training and scoring process can provide the opportunity for these teachers to "re-set," if needed, their internal definitions of quality. This happens when teachers observe both intuitive and explicit evidence that their internal standards for what constitutes high-, moderate-, or low-quality student work are not in agreement with their colleagues.
5. Following the scoring process, teachers can discuss the implications for classroom instruction and thus deepen their understanding of the relationships

---

FIGURE 1
Types of scoring rubrics or criteria

Holistic rubrics – This type of scoring generally results in a single overall score being applied to a piece of student work. This rating might be numeric (e.g., 4, 3, 2 or 1) or use performance labels such as "advanced," "proficient," etc.

Analytic rubrics – In this type of scoring, multiple dimensions that describe different aspects of student work are used in scoring. These dimensions are to be rated and recorded separately, although a total score across all dimensions might also be computed.

Checklists – In this type of scoring schema, the scorer answers "yes" or "no" for several aspects of student writing (for example, whether key aspects of student writing are present in a persuasive essay that students were asked to write). The total number of "yesses" might serve as a total score, although the response to each item on the checklist might also be reported.

between the nature of student responses and the nature of instruction.

Preparing for collaborative scoring

There are several steps in the development and use of the assessment, as well as the training of teachers to be scorers, that bring value to the time and effort spent on collaborative scoring:

Assessment development and use

1. In order to have student work to score, assessment prompts must be developed and field-tested to elicit student work. Do the assessment questions or prompts produce student responses worthy of the time and effort given to scoring?

2. The criteria for judging student work, often expressed in terms of scoring rubrics or criteria, also need to be created and field-tested. There are several types (see Figure 1), but key is for all student work to be judged with common criteria across students and teacher scorers. Do the rubrics or other criteria capture the most important aspects of student responses to the prompts?

3. The assessment developer and advisors need to select samples of student work to illustrate the range and/or types of student performances obtained in field testing. Are there good examples of student work at every score level? Are there clearly stated reasons why each piece of student work was given the score point assigned to it?

4. The assessment needs to be administered in a standardized manner, so that all students are asked to respond to the same task in the same manner. This does not, of course, preclude tasks that permit students to respond in a creative, one-of-a-kind manner.

FIGURE 2
New generation of collaborative scoring software

Michigan educators have contributed to a research and development project coordinated by the Michigan Assessment Consortium (MAC) and software developer MZD. The Michigan Collaborative Scoring System (MI-CSS), powered by Oscar Classroom—an online platform for arts teachers to collaboratively score student responses to performance assessments—is now available for voluntary use by arts educators statewide. Although MI-CSS was developed for the arts, it could benefit all disciplines, including arts, sciences, social studies, mathematics, English language arts, health, and world languages.

Learn more about the software and how to get involved at https://mzdevinc.com/classroom.

Teacher scoring process

5. Rather than rely on each teacher scorer to use their own definition of high-quality work, teachers are shown what the assessment developer and advisors believed to constitute high, moderate, and lower levels of response quality. This helps teacher-scorers reach a common understanding of what doing well on the assessment looks like.

6. Teacher-scorers may be trained to understand and reliably judge the varying qualities of student work. In such training, teacher-scorers are exposed first to the different levels, based on previously created rubrics or criteria (#2 above).

7. Teacher-scorers are shown examples of student work (#3 above) at each performance level and given the reasons why each piece of student work was judged to be at that score level. Then teacher-scorers practice on other pre-scored (but unidentified) student work.

8. Once they seem to be able to judge student work in agreement with the pre-scorers, teacher-scorers can begin

to score student work. Periodically, pre-scored work (not identified as such) is given to scorers to make sure that their scoring remains in agreement with the scores given by experts.

What are the challenges of collaborative scoring?

The two largest challenges to the collaborative scoring of student work are 1) the time that it takes to do such scoring and 2) the management of the scoring process (paper-flow in a document-based scoring system or the expense of traditional computerized scoring systems). However, there are management strategies for each of these challenges.

First, rather than consider time spent on scoring student work as "assessment time," innovative school sites conceive of it as "professional learning time," for which schools already budget time. As noted, there are real, tangible benefits available in collaborative scoring, so this is not just a "sleight-of-hand" switch. Collaborative scoring is real professional learning!

Second, newer online software makes managing the scoring process easy now, and far less expensive. Rather than rely on mainstream test vendors and their expensive software, microcomputer-based software such as OSCAR Classroom (see Figure 2) is available at far less cost. It can be used to upload student work, distribute it to teacher-scorers, determine whether scorers are in agreement with one another, and, if not, resolve any differences in scoring. The software also produces student and classroom reports and keeps track of the scoring reliability of each teacher-scorer.

TO LEARN MORE
....

Learning Point: Performance Assessment: What is it and why is it useful?
(Michigan Assessment Consortium, 2017.)
bit.ly/Perf-Assessment

Learning Point: Performance assessments in the visual arts classroom.
(Michigan Assessment Consortium, 2017.)
bit.ly/Perf-Assessment-VA

Performance Assessment Research & Development
(includes information on MI-CSS)
www.michiganassessmentconsortium.org/research-development/
performance-assessment

The Michigan Assessment Consortium's Assessment Learning Network (ALN) is a professional learning community consisting of members from MI's professional education organizations; the goal of the ALN is to increase the assessment literacy of all of Michigan's professional educators.