

T E A M W



ON ASSESSMENTS

CREATES

POWERFUL

PROFESSIONAL

DEVELOPMENT

BY JAY MCTIGHE AND MARCELLA EMBERGER

Teacher collaboration is a powerful form of professional learning. One focus for collaborative efforts is designing assessments. When teachers design assessments, give each other feedback through peer reviews, evaluate student work, and plan together for improvement, they are engaged in highly effective professional development.

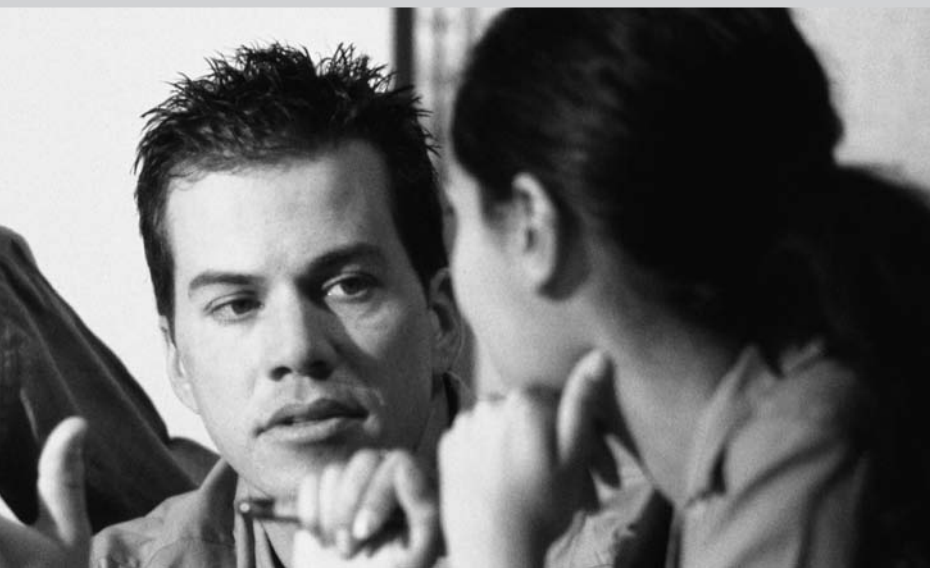
Assessments have two common purposes. One purpose is evaluation. Many teachers think assessment is summative, something done at the end of instruction to evaluate what students have learned and to give them a grade.

A second purpose of assessment is closer to the teaching-learning process. Rick Stiggins (2002) distinguishes between the two purposes as assessment *of* learning (summative/evaluative) and assessment *for* learning (ongoing, formative, and informative). Assessments *for* learning are diagnostic rather than summative. They give both teachers and students feedback to help guide

JAY MCTIGHE is an author and educational consultant. You can contact him at (410) 531-1610 or e-mail: jmctigh@aol.com.

MARCELLA (MARCY) EMBERGER is an educational consultant. You can contact her at (410) 225-7996 or e-mail: marcyemberger@earthlink.net.

O R K



their actions — revising, reteaching, focusing practice.

Ongoing assessments are a vital part of the teaching-learning cycle. Without continuous assessment, student learning is limited to a one-shot, hit-or-miss event — maybe they get it, maybe they don't. Ongoing assessments give teachers feedback so they can adjust their instruction. Ongoing assessments help students focus their efforts. The most effective teachers use assessments for *learning* in addition to evaluation.

FORM AND FUNCTION

The format of assessments should match the goals being assessed and

the reason for assessing.

How does a teacher know that students *really* understand? The evidence is there when students can *apply* what they are learning to new situations and *explain* their responses (show their work, support their reasoning, justify their answers).

Performance assessments that use real situations that reflect the world beyond the classroom are called “authentic.” These tasks are typically open-ended to allow students more choices and to encourage a variety of responses, but they still are judged against established criteria.

Because classroom, school, and district assessments are less influenced

by the factors that constrain standardized tests (large-scale implementation, limited time, etc.), teachers can use performance assessments both for both diagnosis (feedback) and evaluation. Of course, teachers can and should also use other assessments such as selected-response quizzes and tests, observations, and portfolios of student work to provide a complete picture of a student's learning.

Once teachers have recognized the value of performance assessments, they face the challenge of finding or creating tasks and scoring rubrics.

Teachers use three strategies to collaborate to develop performance tasks and assessments:

1. Collaboratively design tasks and assessments based on desired learning results.
2. Have peers review tasks and assessments for feedback on designs.
3. Conduct a group evaluation of student work elicited by the tasks.

STRATEGY 1:

Collaborative design

Step 1: Form the group

Anyone can encourage a group to form to work collaboratively to design performance tasks and assessments, give feedback, and produce a portfolio

Teachers use three strategies to collaborate to develop performance tasks and assignments.

of usable performance assessments. The optimal group size is three to five people teaching the same grade level or

subject area. They do not need to be from the same school or even the same district. Some teams that cross subject areas may collaborate to develop multidisciplinary performance tasks.

Step 2: Meet as a team

A facilitator helps participants address:

1. The goals or content standards being assessed.
2. The task students will perform to

Assessing understanding

A primary goal of teaching is to help students understand the important ideas and processes identified in content standards. Classroom, school, and district assessments should provide evidence of student understanding.

While it makes sense to familiarize students with the format of state standardized tests, fixating on the format is counterproductive in the long run. The best way to raise test scores over time is to:

1. Teach the key ideas and processes outlined in content standards in meaningful and engaging ways (this assumes the test is aligned with standards);
2. Use local performance-based assessments (more rigorous than one-shot, standardized tests) to find out whether students understand the content;
3. Raise the standards and quality of local assignments and assessments using the processes outlined here; and
4. Use the results of ongoing, authentic assessments and other evidence to plan improvements, rather than waiting for the once-a-year standardized test score report.

demonstrate their understanding and proficiency.

3. The criteria by which the student's performance will be judged.

Step 3: Decide which standards to measure

Each team, guided by the facilitator, decides which goals or content standards can appropriately be assessed. Not every goal requires a performance assessment. Performance assessments are needed when the goals are procedural (involve skills or processes, such as problem solving) or call for students to understand concepts and principles. For example, if the standard expects students to be able to identify state capitals or to know chemical symbols, multiple-choice or fill-in-the-blank formats provide appropriate evidence of learning.

Step 4: Create a task

Teachers develop an authentic situation through which students will demonstrate their knowledge and skills. The team can brainstorm tasks using the G.R.A.S.P.S. framework (Wiggins & McTighe, 1998):

Goal: What is the purpose, challenge, or problem (to persuade, to

inform, to entertain, to sell)?

Role: What real-world role will the student assume (editorial writer, museum director, artist, business owner)?

Audience: For whom is the student working (newspaper reader, museum visitor, viewer, client/customer)?

Situation: What is the situation or context (a controversial community issue that must be resolved)?

Product/Performance: What will students make or do to accomplish the goal (a letter to the editor, display, mural, business proposal)?

Standards: How will the product or performance be judged as successful?

Step 5: Develop evaluative criteria

The team develops criteria that teachers and students will use to appraise students' work on the performance tasks. For most complex performance tasks, designers should use three types of criteria:

1. Criteria to assess the *degree of understanding or proficiency* (accuracy, thoroughness, thoughtfulness, efficiency).

2. Criteria to assess *work quality* (well-crafted, mechanically correct, skilled, neat, creative).

3. Criteria related to *impact or result* (Was the letter to the editor persuasive? Was the museum display informative? Did the scientific investigation actually test the hypothesis? Was the role play convincing?).

These criteria are the basis for developing a scoring rubric. The performance scale — for example, one to four — includes descriptions of the level of understanding, proficiency, work quality, and impact.

**STRATEGY 2:
A peer review process**

We rarely review and critique units and assessments teachers have designed. Structured peer reviews, guided by design standards, can help teachers improve designs.

Peer review teams can be homogeneous — based on content areas or grade levels — or heterogeneous. Both have advantages. In general, homogeneous groups provide more specific feedback about content-oriented criteria, such as whether a task matches content requirements (task validity) and is authentic (related to life outside the classroom).

Heterogeneous groups can provide information about whether the task is clear, potentially engaging to students, and easily implemented. Administrators and teachers who have not helped design the task are useful members of the review team.

Groups of three to five members work well.

For peer review to be successful, team members must have a high level of trust so they feel safe when giving and receiving feedback. Creating trust takes time and is built in part through practicing the skills of providing descriptive, non-evaluative feedback.

One method for building trust is practicing peer review sessions using

sample assessment tasks and rubrics. The skills of giving and receiving feedback need to be modeled and practiced before initiating the process with teachers' own designs.

Peer review is more successful when:

1. Feedback is specific, descriptive, and guided by the criteria in design standards. For example, instead of saying, "We liked your performance task," a group member might say, "The task is authentic because it asks students to apply their knowledge in a 'real world' way."

2. Feedback is not personalized. The reviewers provide feedback to help improve the task and rubric and do not praise or criticize the designers.

3. The designer listens to the feedback and asks clarifying questions. Designers should not try to explain or defend their work. After the peer review, designers can decide whether to incorporate the feedback.

4. Meetings stay on schedule. Participants must guard against tangential discussions or sidebar conversations.

Excerpted from *Powerful Designs for Professional Learning*, edited by Lois Brown Easton (Oxford, OH: NSDC, 2004). Available through the NSDC Online Bookstore, <http://store.nsd.org>.

**STRATEGY 3:
Anchor evaluation in student work**

When teachers use common performance assessment tasks and rubrics, they collect data in the form of student products and performances that can be used to determine how well students understand what they are learning. Focusing on student work increases teachers' ownership of student achievement since the work is a result of their own curriculum, assessment, and teaching.

Step 1: Reconvene teams

Reconvene the teams that designed the performance assessments and rubrics after teachers have had a

chance to use them in the classroom. Each teacher should bring five to eight randomly selected samples of student work resulting from the assessments, with enough copies of each sample for every team member. If the assessment required a performance, it should be ready to view on a videotape or listen to on an audiotape. The sample student work should not have a visible score.

anchors give teachers and students clear targets that help guide their work.

The teams examine the student work to be able to describe, rather than score or grade it, so that those who created the performance assessments and rubrics can make adjustments that are likely to improve the results.

Step 2: Describe the student work on the performance task

Working with one performance assessment and resulting student work — one sample at a time — each team describes what is in students’ work. A recorder makes notes on chart paper so the group can use comments later. The group asks itself to:

Describe:

- What knowledge and skills are assessed?
- What kinds of thinking are required (recall, interpretation, evaluation)?
- Are these the results I (we) expected? Why or why not?
- In what areas did the student(s) perform best?
- What weaknesses are evident?
- What misconceptions are revealed?
- Are there any surprises?
- What anomalies exist?
- Is there evidence of improvement or decline? If so, what caused the changes?

Evaluate:

- By what criteria am I (are we) evaluating student work?
- Are these the most important criteria?

Facilitator’s checklist

To help ensure the design process is successful, the facilitator should:

1. Use computers when designing tasks and rubrics to make editing and distributing them easier. Meet in or near a media center or computer lab with Internet access.
2. Provide teachers with relevant resources to support their design work. For example, have content standards documents and curriculum frameworks on hand, and provide sample tasks and rubrics to serve as models.
3. Help teachers use Internet resources related to assessment. Teachers are masters at adapting ideas and can build on others’ ideas rather than starting with a blank slate.
4. Schedule multiple opportunities for the group to meet for informal sharing and feedback sessions throughout the design process. A formal peer review session toward the end of a design workshop should not be the only opportunity for feedback. A gallery walk offers a practical and energizing way to share and get feedback during any part of the process. For a gallery walk, design teams post their draft performance tasks and rubrics on a wall and participants view the works in progress, offering feedback and suggestions anonymously with sticky notes posted to the charts.

- How good is “good enough” (what is the performance standard)?
- **Interpret:**
- What does this work reveal about student learning and performance?
- What patterns are evident?
- What questions does this work raise?
- Is this work consistent with other achievement data?
- Are there different explanations for these results?

Identify improvement actions:

- What teacher action(s) are needed to improve learning and performance?
- What student action(s) are needed to improve learning and performance?
- What parent action(s) will support improved learning and performance?

After about 15 minutes of describing, evaluating, and interpreting the work, the group is ready to anchor the work to the scoring levels on the rubric.

Step 3: Anchor the work

The next step for the reconvened teams is anchoring. Anchoring means selecting examples of student work to represent each of the score points on an evaluation scale. These examples illustrate the quality or proficiency expected at each level based on established criteria. Anchors help teachers understand and apply the criteria and standards consistently when they evaluate student products or performances. Anchors give teachers and students clear targets that help guide their work and help students understand and apply the criteria when they are evaluating themselves or doing peer evaluations.

There are two models for anchoring the scoring system for performance assessments.

Model 1 uses established scoring criteria on the rubric, and each team evaluates student responses, products, or performances according to the pre-set criteria. Next, the group sorts student work by score. The group then selects responses, products, or performances for each score point that

Design standards

Design standards define the qualities of effective curriculum and assessment. Design standards are a reference point during design to be sure the assessment meets the standards, to help teachers review and refine drafts, and which can be used by independent reviewers (such as a curriculum committee) before assessments are distributed to other teachers.

The Maryland Assessment Consortium developed these design standards:

To what extent does the performance assessment task:

1. Assess student performance on the identified content standard(s) and benchmarks?
2. Establish a meaningful context based on issues, problems, themes, or student interests?
3. Require the student to apply thinking skills or processes rather than merely recall factual information?
4. Establish criteria linked to the standards/benchmarks for evaluating student products and performances?
5. Contain activities likely to engage students?
6. Provide clear, unambiguous directions to students?
7. Contain accurate and credible information?
8. Use interrelated activities to achieve its purpose?
9. Allow for easy use in the classroom?
10. Provide feedback to teachers and students about identified goals or content standards?
11. Integrate subject areas?
12. Provide opportunities for students to reflect on and self-evaluate their performance?
13. Allow students to revise?
14. Allow for a choice of products or performances?
15. Use technology appropriately?

illustrate the criteria for that score. With only five to eight samples, the group may not find an example for each score. Use Model 1 when a performance task and the scoring rubric(s) have been validated through field testing, reviews, and revision.

Model 2 uses student responses, products, or performances to identify or refine the scoring criteria. The group sorts student responses into three (high, medium, low) or four (excellent, good, fair, poor) levels

based on general quality. The group reviews each set and determines the distinguishing characteristics of the responses. They then develop criteria for each level and select several responses to illustrate those criteria. Use Model 2 when a task has been used for the first time and no rubric exists or the rubric is a draft.

The reconvened teams evaluate the student responses, products, or performances. The team member submitting the performance assessment

and rubric for anchoring does not share the scores the samples received.

At the same time, teams should examine the performance assessment task itself, particularly the directions given to students, to see if the directions lead students to produce the desired outcome. Task directions that are vague or misleading may cause students to prepare a response that fits neither the intention of the task nor the criteria on the rubric.

Tips for successful anchoring

1. Use anchoring to refine performance standards or create them if a rubric has not been designed. When

Ultimately, students benefit when educators work together on assessments.

educators choose examples of student work that illustrate the various levels in a rubric, they can easily answer the question, “How

good is good enough?” Anchors also help scorers judge work more consistently and help students assess their own work more accurately. With tangible illustrations of what quality work looks like, teachers and students can understand the specific qualities

of effective work and get beyond general statements, such as “well-organized” or “persuasive.”

2. Select several examples for each level. A single example suggests that there is just one best answer or pathway rather than several approaches to an authentic task (diverse excellence). Using several anchors provides a richer set of examples to guide teachers and students.

3. Collect and publish the anchor examples at the grade, school, or district level to promote more consistent evaluations and to help teachers explain scores and grades to parents and students. Many teachers report that grading quibbles virtually disappear when clear rubrics and anchors are available.

CONCLUSION

Collaborative designs and peer reviews honor and enhance teachers’ professionalism, expertise, and collegial learning. Working in teams to evaluate student work against established criteria, identify models of excellence (anchoring), and plan

needed improvements promotes a results-oriented culture of quality.

By designing performance assessments, educators enhance their understanding of content standards and of the evidence needed to show that students really understand the important ideas and processes contained in those standards. Teachers discover that the connection between curriculum and assessment becomes clearer, teaching is more sharply focused, and evaluation is more consistent.

Ultimately, students benefit by having defined learning goals, opportunities to demonstrate their understanding in more authentic ways, and advance knowledge of the evaluation criteria so they have greater purpose in their learning.

REFERENCES

Stiggins, R. (2002, June). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765.

Wiggins, G. & McTighe, J. (1998). *The understanding by design handbook*. Alexandria, VA: ASCD. ■