Susan A. Chavez

Soundoff!

If at First You Don't Succeed . . . Test, Test Again (Not!)

THE FEINWAY STORY

The Feinway Violin Company, a hypothetical business, handcrafts fine, affordable violins in a twelve-step assembly process. Different craftspeople are responsible for each of the twelve steps. Feinway recently decided that 100 percent of the violins that the company produces must attain a measurable, high quality. The company understands that simply asking each craftsperson to do an excellent job is not enough. To give the craftspeople a clearer picture of where problems might occur and specifically what parameters need to be addressed, Feinway implemented a quality-control program. The quality-control program tests every violin to a defined set of standards at steps 3, 5, 8, and 10 in the process. All test data are recorded and archived for each individual violin.

After testing the violins, Feinway analyzes the test data to determine possible causes of failures, estimates the number of violins associated with each type of failure, and prioritizes the fixes that will be implemented. This information is communicated to the craftspeople. The managers and craftspeople create a plan to implement improvements. Violins are then retested after remediation. Retesting does not occur until after test results are analyzed and improvements are implemented.

How and when do we start fixing it?

The management at Feinway discourages finger-pointing and blaming. Feinway assumes that the vast majority of its workers strive for excellence and that teamwork is most productive.

The Feinway method: Plan, implement, test, analyze, and remediate. Sound familiar? (Hint: think statewide assessment.) Planning established the mathematics standards, implementation includes standards-based teaching, testing is . . . yup, you guessed it: benchmark testing. I hope that Feinway's supportive culture does not sound like it comes from another planet. If you wonder where analysis and remediation fit in, you are not alone. My main objective in this article is to encourage discussion on the concepts of analyzing test results and implementing remediation.

We are doing all this testing. Clearly stuff is out there to fix; how and when do we start fixing it?

MOVING FORWARD

For this discussion, the background on benchmark testing is that it is with us for at least the immediate future and that it takes a lot of our time and energy, so we need to take advantage of what it yields. I am not arguing the pros or cons of high-stakes tests. I am arguing that testing and retesting (and still more retesting) something that is broken does not generally fix it. Thoughtful remediation is needed.

To productively move forward toward data-driven decision making, we must make the following three assumptions:

- Benchmark testing should be more than a triage-like tool for classifying schools. (Triage separates incoming wounded into three classes: not needing immediate care, critically wounded, and hopeless.)
- Benchmark testing can be used as a tool to get a clearer picture of where and how our education process needs improvement.
- Some things need fixing, but we will not spend time assigning blame.

The views expressed in "Soundoff!" reflect the views of the author and not necessarily those of the Editorial Panel of the *Mathematics Teacher* or the National Council of Teachers of Mathematics. Readers are encouraged to respond to this "Soundoff!" by sending typed and double-spaced letters to the *Mathematics Teacher* for possible publication in "Reader Reflections." "Soundoffs!" from readers are welcomed.

Susan Chavez teaches mathematics at Clackamas High School, Clackamas, OR 97015. She enjoys applying problem solving to all aspects of mathematics education.

ROOT CAUSES OF FAILURES: IDENTIFYING AND QUANTIFYING

I believe that retesting without remediation is generally not the best plan. Retesting without remediation assumes that students are capable of passing the test but that they fail for unknown reasons. Even if this belief is correct, we want to ensure that every student has a fair chance each time that he or she takes a benchmark test. We need to identify the root causes of failures.

Identifying root causes

Root-cause analysis is a technique commonly used for failure analysis in industry. **Table 1** is a sample root-cause table. Causes of failures are grouped into three broad categories associated with the system, the student, or the teaching. Examples of root causes are listed for each category—individual schools may have different root causes or additional ones to add.

To make sure that all who are involved understand their roles, each district, administration, and faculty should discuss and publish its root-cause list, even if the list is not perfect or complete. We all have a common goal of providing high-quality education, and we all need to understand the road maps to get there.

Quantifying root cause

Quantifying root cause is difficult. Many root causes need to be studied at a district level. It probably will not happen tomorrow. We should not let that stop us. We must work with the school's administration to estimate the percent of failures for the various root causes. Again, as part of the road map, it is important that these estimates, even if they are just guesses, are communicated among all those involved.

Guesses and estimates are only a stopgap. Teachers should ask their districts to work on analyzing the data. Seriously. District motivation should come from the requirement that schools and districts use test-result data to revise their school improvement plans (SIP) and consolidated district improvement plan (CDIP).

Giving these tests time and time again without meaningful analysis and remediation is a disservice to our students (not to mention ourselves).

REMEDIATION: DEFINING AND PRIORITIZING

After identifying root causes and correlating them with percentage failure estimates, we need to define the remediation, estimate its cost, and prioritize.

Root causes under "the system" can be among the least expensive ones to fix. For example, my school, in suburban Portland, Oregon, formerly had a system that required students to save their scored mathematics work samples and give them to an adviser. (A "work sample" is a situational problem that requires students to demonstrate conceptual understanding. Students must fully develop their solutions and clearly communicate all mathematical processes.) Advisers, who met with the students twice a month, recorded the students' scores in a cumbersome database. At one point in the semester, I noted that 50 percent of passing scores on a work sample that I gave my students had not been entered into the system. Students lost their work samples, and advisers entered the scores weeks—and sometimes months—after students submitted them. Although high school students

We need to identify root causes of failures

TABLE 1
Sample Root-Cause Table

1. The System	2. The Student	3. The Teaching
1a. Test requires specific test-taking skills not provided in curriculum	2a. Student has test-taking anxiety.	3a. Concepts on test are not sufficiently discussed.
1b. Test response format is confusing or prone to user error (for example, fill-in-the-dot or computerized multiple choice).	2b. Student does not take test seriously.	3b. Presentation of concepts is not effective for all students (different learning styles, lack of higher-order learning).
1c. Test questions or answers are poorly written.	2c. Student has not studied material presented in classes.	3c. Example test questions are not sufficiently discussed (that is, we need to teach to the test).
1d. Test timing is poor (for example, too early in year or same day as other tests).	2d. Student has issues that are not related to mathematics. (English-language learner, reading difficulty, vision problems, and so on).	3d. Presentation of example test questions is not effective for all students.
1e. Results are not properly recorded or are lost.		

should be capable of submitting a work sample, it is not the best way to teach students to keep track of and submit work samples.

Another system root cause effectively addressed by simple remediation and subsequent retesting is root cause 1a, test requires specific test-taking skills not included in the curriculum. Students may be unfamiliar with effective test-taking techniques, particularly those associated with multiple-choice tests. If students are given tips on taking multiple-choice tests and have more practice, they can use the techniques and pass.

Every teacher should be part of the solution

When deciding whether a remediation is appropriate, teachers should remember the goals of benchmark testing at the school and ask whether the remediation supports the goals. An interesting example is root cause 3c, example test questions are not sufficiently discussed. In other words, the teacher did not teach to the test. I know that these words are dirty ones in many circles. However, information from several educators puts teaching to the test in a different light. The new slant on specifically teaching test questions is that it can be no different from embedding work samples in the curriculum. The particular questions on the test have been selected as ones that all high school students should be able to answer. We want to teach students how to apply concepts that they understand to work these problems. The problem should be familiar, not surprising or foreign in appearance. Obviously, the merits of this slant depend on how it is implemented.

WHAT TEACHERS CAN DO FROM THE TRENCHES

First, I cannot overemphasize the importance of a nonpunitive environment for teachers and students. And let it begin with me. . . . Here are some suggestions:

- We must provide direct and specific feedback to students. Work samples are a good place to start. Naturally, we should emphasize specific things that students can reasonably do rather than what they did not do. ("For heaven's sake, listen in class!" is not productive.)
- We should request help from school administrators in raising benchmark-score awareness to the student body. We can post congratulatory lists in prominent places and try to keep a positive spin on passing rather than a negative spin on not passing.
- We should use the Feinway method in the SIP and ask the school district to consider using it in the CDIP.
- We should ask the school district for statistical analysis of test data to replace guessed or estimated percents of root causes, and we should organize the department, colleagues, and school.
- We should identify what we can do in the remediation process. Teachers should use the one-plate principle, which goes like this: My plate of "things I can do in this lifetime" is this big and it is always full. If I take on task A, then task B must fall off. Easier said than done (My colleagues will agree that I am one of the worst offenders. However, in my district, passing work-sample and benchmark tests is a graduation requirement. Fixing these problems has a high priority, as does my sanity.)
- Every teacher should be part of the solution.

I appreciate thoughtful discussions with the following people: Aeylin Summers, Karen Phillips, Camille Wainwright, Tom Swanson, and my colleagues at Clackamas High School. Writing this article was supported by Write On! a writing retreat facilitated by the Oregon Collaborative for Excellence in the Preparation of Teachers (OCEPT) and funded by National Science Foundation grants DUE-9996453 and 0222552.

MT