

The Flaws, Fallacies and Foolishness of Benchmark Testing

Benchmarking is a great tool for improving an organization's performance...when used or identifying, then tracking (by measuring) specific variables that are proven to be "S.M.A.R.T." That is:

- Specific
- Measurable
- Achievable
- Realistic
- Time Specific

The benchmarking process leads to better organizational communication and morale. The entire organization gets behind the key improvements that are identified; and everyone takes pride (and credit) in the organization's better performance and improved profitability. Everyone shares in the accomplishment. (At least that is what is supposed to happen.)

Unfortunately, the way that school districts use the benchmarking tool is "D.U.M.B."

- Distracting, Dampens Morale
- Under funded, Un targeted, Unrealistic
- Meddlesome, Mediocre Outcomes, Menacing (to some teachers)
- Bureaucratic, Burdensome, Bothersome

The idea to use benchmark testing was adopted by school district curriculum departments, presumably after reading about how business and industry improves business performance (and the bottom line) by instituting metrics (measurements). These measurements help everyone in the business organization focus their efforts on what really matters (independent variables). The processes that are measured are those that directly relate to improving worker performance and productivity.

On the face of it, this seems to be a great idea.

Don't we all stand for efficient and productive teaching? Don't our children deserve teachers who are continually improving the process of instruction? Shouldn't teachers be fine tuning their skills to Swiss watch precision?

If benchmarking is such a great tool, and if school districts are not achieving stellar results from benchmarking initiatives, what is going wrong?

Well, maybe, benchmarking is the wrong tool for the job. Or, maybe the benchmarking tool has been used incorrectly.

© 2005 by Joseph Chmielewski All rights reserved.

<http://www.classroomtoolkit.com>

Reprints may be shared under the Reciprocal Public Materials License:

<http://www.classroomtoolkit.com/reciprocal-license.html>

The Flaws, Fallacies and Foolishness of Benchmark Testing

The keys to determining whether the results of a benchmarking program are beneficial are whether teacher performance and teacher morale improves as a result of the program. Evidence seems to suggest that benchmarking, as implemented by school districts, falls into the "D.U.M.B" rather than "S.M.A.R.T." category.

First, benchmark testing is introduced as a "cure," not as a tool for discovery of effective teacher performance.

Typically, the educators who know how to teach children are not asked what variables produce the greatest learning outcomes in their classrooms. Instead, all children are given a paper and pencil test. The test scores are supposed to help teachers and administrators. But, data derived from these paper and pencil tests are dependent variables, not the crucial, independent variables that need to be addressed.

Independent variables are the ones identified and measured that the organization can do something about.

Unless the independent variables can be identified, changed and controlled, no cause and effect relationship is proven.

What school district collect is "result data," (dependent variable) when the kind of information that needs to be collected is "what caused?" data (independent variables).

So, what is the difference between how business and industry use benchmarks for improving performance, and what school districts do with the tool?

A company, lets say a package delivery company, notices that it is aggravating customers because of delivery delays and lost parcels. The company resolves to improve its performance using the benchmarking process.

The first step in the benchmark process is to find companies that are "getting it right." Of course successful competitors would be ideal models, but direct competitors are unlikely to provide any help.

So, the company contacts non-competitors with similar processes (in this case maybe florists and pizza restaurants) and finds out how these companies measure and improve their accurate, on time delivery.

Next, the company institutes a measurement system for their delivery process, and then experiments to determine how the processes behind these measurements can be improved.

© 2005 by Joseph Chmielewski All rights reserved.

<http://www.classroomtoolkit.com>

Reprints may be shared under the Reciprocal Public Materials License:

<http://www.classroomtoolkit.com/reciprocal-license.html>

The Flaws, Fallacies and Foolishness of Benchmark Testing

For example, package sorting time may be important. So, the company tracks the time that the package arrives at their facility, and the time that the package is placed inside a delivery van, then calculates the difference. Now, everyone in the company knows that shrinking this time period is being measured. Everyone focuses upon ways to cut hours, then minutes off this lag time. All employees submit suggestions for improving the process, and effective, efficient ideas are tested, and then adopted company-wide. (It doesn't matter whether a shipping clerk, mechanic, custodian or CEO comes up with the idea. What proves to work is adopted.) Finally, once targets are met, new initiatives are studied and implemented.

The business will commit time and money to this study, and the company will commit whatever it takes to implement an effective solution.

Often, automated methods are suggested, especially the automation of busy work like tracking. In these cases, the company spends the money to implement the solution

The business will focus upon what it can change (those pesky independent variables again.)

What getting-it-right industries are educators analyzing to discover the proper metrics for improving instruction?

What getting-it-right teachers are measured so that performance standards can be applied against the performance of all other teachers?

What school districts do instead of instituting a useful benchmarking program is to give face-validity-only, paper and pencil tests to all students, and, from these "scores," determine" which teachers are doing a good job (based upon "passing rates"). These scores are supposed to assist teachers in getting ready for the mandatory high-stakes tests later in the academic year.

So, why is this process flawed?

In order to measure and replicate improvement, it is necessary to know what independent Variables created the improvement.

What the benchmark testing program discovers is that on one day, some students answered more correctly than some other students. The benchmark test data (dependent variables) do not provide a clue as to what variables were responsible for the more correct answering of some students; nor, do the benchmark test data determine if any of those variables that were responsible for more correct answers can be changed in any way by teachers in classrooms.

© 2005 by Joseph Chmielewski All rights reserved.

<http://www.classroomtoolkit.com>

Reprints may be shared under the Reciprocal Public Materials License:

<http://www.classroomtoolkit.com/reciprocal-license.html>

The Flaws, Fallacies and Foolishness of Benchmark Testing

In order to make improvements in instruction through the use of benchmarks, a school district would need to measure specific variables that go into the performance of teachers. There is little direct benefit of measuring the global symptoms that occur sometime later using the paper and pencil benchmark test.

For example, the reason that math test scores are low could be that the new math text book is confusing students, or, that the book (that teachers were ordered to use) does not match the benchmark test time frame for objectives tested, i.e., the benchmark test didn't test what the teacher taught.

An independent variable affecting test scores (that is beyond the control of teachers) might be the many cases of influenza, with even the teacher absent for several days, before the benchmark test was administered.

Measuring students' answers on a paper and pencil test is similar to the package delivery company measuring package size and package shape to determine how effectively their delivery system is operating. That is, not directly relevant.

What an effective educational benchmarking process would have to do to actually improve instruction would be to:

- Determine exactly what materials, methods and behaviors result in actual students'
- Improved performance outcomes
- Determine exactly what methods and behaviors can be measured for each teacher's performance
- Measure the teacher's behaviors and performance as instruction is being delivered
- Analyze the metrics and prescribe changes for the teacher before instruction gets too far off track

To be effective, curriculum experts would need to know exactly what teachers have to do to teach each specific content area subject.

Next, these experts would have to identify specific, observable, measurable behaviors and events.

Then, these experts would need to find a way to collect these measurements (without adding yet another burden on teachers).

Finally, these experts would have to measure exactly how well each teacher performed, and measure how well each teacher improved, based upon these measures.

© 2005 by Joseph Chmielewski All rights reserved.

<http://www.classroomtoolkit.com>

Reprints may be shared under the Reciprocal Public Materials License:

<http://www.classroomtoolkit.com/reciprocal-license.html>

The Flaws, Fallacies and Foolishness of Benchmark Testing

For those teachers who bristle at the thought of being "placed under a microscope, our advice is to "relax."

First, the curriculum experts can't devote the time and money required to discover the specific, measurable, observable "stuff" that is required to actually gauge teacher performance, class-by-class, and tell how well each teacher is doing.

However, even if curriculum experts knew what information to collect...specific-measurable-observable assessment of each teacher's performance...who is going to collect the data?

We can't give the job of collecting performance data to teachers (yet another busy work task that is unrelated to instruction), can we?

If we did require teachers to collect this data, could we trust that teachers wouldn't "pad" the data to make themselves look better?

And, if observable performance behaviors are important, who is going to observe so many teachers? No district can spend the kind of funds required to observe every teacher. No school district has that much extra money.

In educational practice, the benchmark testing program seems to degenerate to "let's make teachers work harder by increasing 'accountability' and 'productivity' without spending any money, without providing resources or retraining our teachers."

This will sound politically correct and it will seem like those in charge are doing something.

It is easy to say, "We'll test all the students, and if our teachers don't deliver higher test scores, we'll replace those "losers" with someone who can get the job done." (Of course, the replacements won't do any better, but the people in charge are vindicated by their initiatives."

So, benchmarking in education will continue to provide the candy-bar, potato chip, and Soda-water "lunch." Benchmarking in education seems to taste good, but contains empty calories that provide noxious and detrimental after effects and leave a "bad taste" in teachers' experience.

"Doing benchmarking right" in education would be so expensive that I can't imagine any politician agreeing to make the funds available.

© 2005 by Joseph Chmielewski All rights reserved.

<http://www.classroomtoolkit.com>

Reprints may be shared under the Reciprocal Public Materials License:

<http://www.classroomtoolkit.com/reciprocal-license.html>

The Flaws, Fallacies and Foolishness of Benchmark Testing

District benchmarking initiatives demonstrate what happens when a good tool, used incorrectly, is adopted as a cure for education's ills.

Note:

This article is a bit controversial.

We will be happy to print comments and critiques that focus on improving instruction and benefiting teachers.

If anyone knows of a school district that has used the benchmarking process effectively, please let us know, and we will showcase them.

If anyone knows of a teacher who likes the benchmark testing program and thinks that, the results are beneficial, please refer this person. We would like to interview them to discover what we believe would be a unique perspective. Who knows? Out of the hundreds of thousands of teachers in this country, there might be one that believes that the process is useful.

For an example of what it takes to do benchmarking right, visit the Center for Creative Leadership.

<http://www.ccl.org/leadership/about/index.aspx?pageId=10>

This site will provide clues to what it would take to do classroom benchmarking right, and provide clues to the range of costs that would be required.

Hint: this organization has spent 15 years studying leadership, only one of many skills that a teacher employs daily. And, a single course can cost \$1,295US. Do the math and see what will continue to keep school district from "doing benchmarking right."

You also might be interested in a free report about leadership and emotional intelligence that is available at this site.

http://www.ccl.org/leadership/pdf/assessments/skills_intelligence.pdf

© 2005 by Joseph Chmielewski All rights reserved.

<http://www.classroomtoolkit.com>

Reprints may be shared under the Reciprocal Public Materials License:

<http://www.classroomtoolkit.com/reciprocal-license.html>