

# Using Test Score Data to Focus Instruction

By Susan Trimble, Anne Gay, & Jan Matthews

*Moving from an initial phase of feeling overwhelmed and inadequate, Camden County teachers learned to view data objectively and to use the embedded information in test scores to improve their teaching. Each phase of their progress marked an advance in their thinking about the accuracy and usefulness of test score data. Full-time lead teachers guided the process as the teachers progressed through phases of development in accepting and interpreting data to improve their teaching.*

Advances in technology available to access test data coupled with the challenges of No Child Left Behind (NCLB) are pushing schools to grapple with the complexities of test score data. Schools now have access to more real-time data about their students' progress, and teachers are encouraged to use it to make better instructional decisions. Black and Wiliam (1998) and Guskey (2003) have documented the use of classroom assessments to improve instruction. Both *Turning Points 2000* (Jackson & Davis, 2000) and the Southern Regional Education Board (2001) recommend using data from ongoing assessments to improve classroom practices that increase learning. The importance of assessment data and access to more data is now well defined.

With the current frenzy to raise test scores, there is little attention being paid to teacher development in learning to use data to improve learning. Teachers who have not been trained, prepared, or supported well in using assessment data to modify instruction often become frustrated and resist involvement in analyzing student data. Addressing teachers' phobias

about data is an initial step in helping teachers progress through the phases of development in learning to use data.

For the past seven years, the Camden County (Georgia) School System has tackled the challenge of teacher acceptance of data usage. One of the first counties in coastal Georgia to implement curriculum alignment and mapping, this mid-size district of about 10,000 students has nine elementary schools, two middle schools, and one high school. The two middle schools together have about 2,400 students, seven administrators, and 126 teachers.

The county initiated benchmark testing in 1996-1997 to improve teaching and learning. Benchmark tests were criterion-referenced, reflecting primarily the state of Georgia's objectives. The student scores on these nine-week tests helped teachers identify which taught objectives students needed further



Teachers collaborate to interpret benchmark test scores.

PHOTO BY SUSAN TRIMBLE

---

*Susan Trimble is an associate professor of education at Georgia Southern University, Statesboro. E-mail: susatrim@georgiasouthern.edu*

*Anne Gay is an instructional lead teacher for the Camden County Board of Education, Kingsland, Georgia. E-mail: agay@camden.k12.ga.us*

*Jan Matthews is an instructional lead teacher for the Camden County Board of Education, Kingsland, Georgia. E-mail: jmatthews@camden.k12.ga.us*

instruction on before the test that “counts”—the state’s criterion-referenced test. The format of each benchmark test matched both the state’s criterion-referenced test and the nationally normed test.

In Camden County, student scores from each nine-week benchmark test were used together with scores from nationally normed tests and from the Georgia state criterion-referenced test to focus instruction. These three tests provided the teachers, administrators, and students with information on their progress, their strong content skills, and areas to improve.

By 2002 the middle level students showed marked improvement in their test scores and compared well with other middle schools state-wide. In a 2002 state ranking of 180 Georgia school systems’ scores on the state’s Criterion Referenced Competency Test (CRCT), Camden County ranked 15th, 14th, and 13th for grades six, seven, and eight respectively in language arts. In reading, the county ranked 35th, 20th, and 19th. The rankings in mathematics for grades six, seven, and eight were 26th, 20th, and 17th, respectively (Georgia Public Education Report Cards, 2001-2002).

These high rankings are notable because the district has low local revenues and low per pupil expenditures. In a state-wide ranking of school districts by the Georgia Department of Education, Camden County falls below the 20th percentile of all Georgia districts ranked by percentage of revenues from local taxes. The efficient use of these funds is evident in the district ranking 171st out of 180 systems on the 2002 Georgia ranking of per pupil expenditures (Georgia School Systems Financial Reports, 2002).

The district attributes benchmark testing as a major reform that contributed to improvement in student learning. Beginning in 1997, every nine weeks, students in grades one through five were tested to generate information on their progress in language arts. As a result, teachers had the incoming test data about the students they were currently teaching. In the subsequent years, grades six through eight were added, and all middle and elementary schools began to implement benchmark testing every nine weeks in math, science, and language arts. In addition to the state’s criterion-referenced test, nationally normed tests continued to be administered annually.

To help in the process of implementing benchmarking and using assessment data, the county created the new position of lead teacher and hired 13 full-time teachers to fill the positions. In addition



PHOTO BY SUSAN TRIMBLE

*Benchmark test data help teachers align their curricular areas to better instruct young adolescents.*

to local revenues, funding for the new positions came from grants and Title 1 money, without increasing class size. These lead teachers devoted most of their time to teacher support and training, benchmark testing and data analysis, and communication between the district and schools. Monthly, they met as a team with their district-level lead teacher coordinator. Each lead teacher was assigned to one or two schools with a ratio of about one lead teacher to 63 middle level teachers. The lead teachers met with subject-specific teams of six to eight teachers to align instruction, implement benchmark assessments, analyze benchmark test scores, and to incorporate the findings into meaningful instruction. Because the teachers felt uninhibited working with peer teachers who did not evaluate their work, they expressed their feelings openly. The lead teachers noticed that with each succeeding nine-week benchmarking period, the teachers’ acceptance of the data changed.

---

***These high rankings are notable because the district has low local revenues and low per pupil expenditures.***

---

From successive meetings of lead teachers with subject area teachers to analyze the test scores, a pattern emerged of distinct phases of teachers’ reactions to the data. Teachers appeared to progress through five phases in accepting the implications of student test scores and using these data to modify their teaching. Each phase lasted two to three months. These phases in teachers’ approaches to the student

PHOTO BY SUSAN TRIMBLE



Using timely benchmark test data empowers teachers to target student learning needs.

test data appeared when local benchmark testing was implemented. Prior to benchmarking, teachers had access only to state-sponsored standardized test data.

## Teachers' Initial Experiences with Standardized Test Data

The Camden County teachers were already familiar with two groups of standardized data—nationally normed tests and the state criterion-referenced test. The Camden County School System administers the Stanford Achievement Test, Ninth Edition (SAT9), a nationally normed instrument, to students in grades one through eight. The state test, the Georgia Criterion Referenced Competency Test (CRCT), is also given in grades one through eight. These two tests generate valuable data about how well students master curriculum objectives, with the disadvantage that the test scores arrive at the school during the summer or the following year when the teachers no longer teach those students.

Using the grade level reports of test scores from the SAT9 and the CRCT, the teachers looked at data by domains. For example, for Reading Vocabulary (SAT9) the domains include Synonyms, Context, and Multiple Meaning. For Math Procedures the domains are Computation/Symbolic Notation, Computation in Context, and Rounding. The reports give a Mean National NCE which teachers used to compare one domain to another to address those listed as "below average." In a similar manner, the state test (CRCT) provides scores by domain,

such as these for reading: Reading for Vocabulary Improvement, Reading for Locating and Recalling Information, Reading for Meaning, and Reading for Critical Analysis. But the results are reported as percentages of students, categorized as "Meets Expectations," "Does Not Meet Expectations," or "Exceeds Expectations." Grade level data were also disaggregated by race and gender. Using these two tests' results, the teachers searched for overall trends by domain and noticed apparent gaps in the curriculum identified by the tests. They noted general strengths and weaknesses in content domains.

The teachers, however, were not emotionally involved with these data sets. Using test data from last year's students to reveal strengths and weaknesses in content area skills, teachers related the grade level test data to the curriculum in general, not to teacher-specific behaviors. Because of the delay in receiving these test results, teachers never felt much connection to the data. They commented, "This information isn't about *my* kids. This is about the kids I had *last* year, and even then, it was just my homeroom. I didn't teach these kids!"

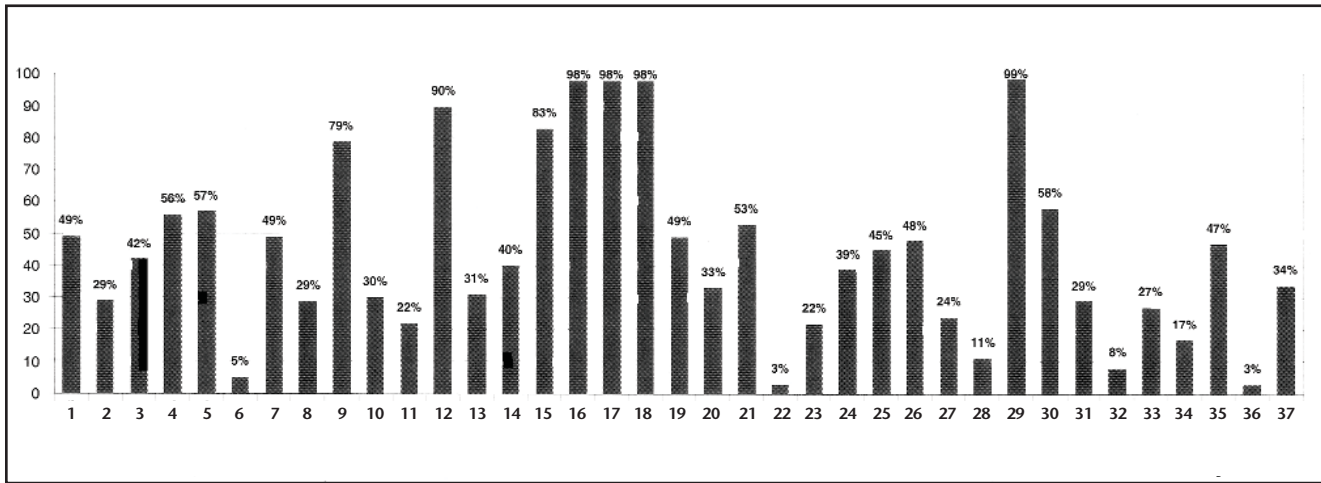
## Reactions to the Real-Time Benchmark Test Data

Given every nine weeks, the benchmark tests generated real-time data that applied to current content, current teachers, and current students. When faced with these test scores, teachers initially voiced strong emotional reactions and were resistant to the test data.

### Phase 1: Confusion and overload

Almost without fail, teachers' first reaction to the local benchmark test results was a sense of confusion. The teachers were overwhelmed by the amount of information, and vented their frustrations with so many numbers, columns, graphs, and percentages. They could not make sense of the reports, which provided an enormous amount of specific information. Their confusion and frustration were justified, as the reporting format for the test results differed from the format of the national and state tests. When the district aligned the curriculum and hired a consulting firm for two years to develop the benchmark tests, the test makers blended the skills measured by both standardized assessment instruments and renamed the skills for each content objective. To teachers, the new labels and format introduced new vocabulary. Although the reports

Figure 1  
Test Item Report for Sixth Grade Mathematics



revealed a variety of information about their own students whom they were currently teaching, the teachers shook their heads and said, “This is too much. I can’t understand any of it. I have enough to do. I was hired to teach, not do statistics.”

In this phase, the lead teachers listened to the concerns of the teachers. Working with six to eight teachers in content-area groups, they responded after teachers expressed their frustration. The lead teachers remembered their own struggles with the data and realized that the data had to be simplified to be understood. To provide teachers with easier-to-read data, they redesigned the data reports, which disaggregated data by objectives in each area, and then orally interpreted the reports in subject-area team meetings. Lead teachers realized that this much information needed to be given to the teachers one report at a time, allowing the implications of each report to be realized gradually. For example, they explained the Test Item Report (see Figure 1) during an hour-long session, pointing out that individual items categorized by objective showed proficiency or nonproficiency. As teachers realized the numbers showed how students performed on individual skills they had taught, they moved rapidly into Phase 2.

### Phase 2: Feeling inadequate and distrustful

On seeing the test reports of the current nine weeks, the teachers were alarmed to realize how poorly their students performed on some of the content that they had taught. Teachers were not pleased with the low proficiency scores and felt they were being judged unfairly. Their first impulse was to blame the instrument. They commented, “How can

two questions on a test possibly establish mastery of an objective? These questions are terrible! We don’t use this format, vocabulary, terminology, etc. in our classes. I don’t teach it that way!” By maintaining that the data were not valid, they felt justified in rejecting it. They denied their involvement with the results and found fault with the test items.

In sessions with lead teachers, the teachers continued to vent, this time about the implications of the data. One lead teacher reported that some of these meetings with content teachers required her to wear a coat of armor to protect her from an onslaught of projected anger. She also heard teacher comments such as: “These scores can’t be right. I taught this concept for a whole week. ... Something is wrong.” Such comments indicated that teachers wanted to investigate the cause of the low scores. Moving into Phase 3, the teachers first challenged the test items.

### Phase 3: Challenging the test

In this phase, teachers asked to examine the tests. Their mood suggested a need to avoid personal responsibility and to identify the test as the cause of the low scores. They examined the number of test items for each objective, the test questions missed by the most students, and focused especially on the wording of these often missed questions. For example, an eighth grade English teacher pointed out, “No wonder those items about business letters scored low; look at this test question. We don’t use this word in class—‘editing the letter for *publication*,’—publication? There’s the reason right there for those low scores.” As teachers critiqued and criticized the test items,

they become aware that their students probably had the same difficulty with unfamiliar words on the standardized tests. The lead teacher asked, "Is this a problem with the test question or the instruction?" The team decided it was a problem with the test but added that they could start using the word "publication" in their lessons.

The teachers also looked for mistakes their students had marked that were common to all students. For example, reports in all classes showed a high percentage of students chose the same incorrect answer on a test item that measured mastery of plural forms in language arts. The question: "Which word BEST fills in the blank in the sentence below?"

"The cows moved their \_\_\_\_\_ forward when they saw Peggy."

---

***Teachers could now answer two questions, "What students need extra help and in what topic?" and "What topics do I need to re-teach in different ways?"***

---

The correct answer was *ears*. Forty-five percent of the students answered it correctly, but 30% gave *ear's*, as their answer. The teachers groaned and commented, "But we've taught plurals and possessives." The lead teachers asked, "Is this a test question problem?" The group agreed that there was not a problem with the question.

Lead teachers helped teachers in this phase by continuing to listen, offering suggestions, and guiding the discussion to keep them on the task of scrutinizing the data for information about each class's progress. The lead teachers passed out teacher copies of the benchmark tests that contained the test items, answers, and the learning objectives that the items assessed. They continually asked, "Do the scores reveal a problem with the test item or the instruction?" Teachers had never before scrutinized test scores, test items, and their own teaching in the detailed manner in which they now proceeded.

#### **Phase 4: Examining the results objectively and looking for causes**

At the next phase, teachers examined their classes' benchmark test reports without the skepticism and resistance of previous phases. They looked for patterns

among the students' responses, considering similarities among the high scoring items (e.g., Figure 1, items 12, 16, 17, 18, & 29) and brainstorming possible causes for low scoring items (e.g., items 6, 22, 28, 32, & 36). In contrast to previous phases, teachers demonstrated a more analytical approach. They looked at all the variables, such as the types of students in their classes, the time of day, the time of the week, the sports schedules, the number of attempts to reach parents, sibling rivalry, and their own teaching strategies. One teacher commented, "Of course, they scored high on using guidewords in the dictionary. That's when we had that word contest in class to see how quickly we could locate a given word." Another teacher noted, "They got that objective on measuring angles when we built those geo-creatures." Other teachers noted differences in test scores among their classes: "All my low scoring classes were in the afternoon; we are all tired then." Another teacher asked for suggestions about how to teach converting decimals, commenting, "Such a difficult thing for me to get across."

Fewer teachers now felt that the tests were a threat or implied a judgment about their abilities. Teachers realized that low test scores might reflect content the students had not yet been taught. Mayo and Shotts (2004) stated that results from benchmark tests must not threaten teachers' security by being used to imply teacher incompetence, but rather must be seen as a valuable tool in instructional planning. Realizing the benefits of using test results, the teachers could now answer two questions: "What students need extra help and in what topic?" and "What topics do I need to re-teach in different ways?" They approached test data as a tool to help them focus their teaching on content, the students, and their teaching styles.

Even though the lead teachers avoided distributing classroom test results in a manner that would encourage comparisons of one teacher's results to another's, they compared their students' scores. The teachers wondered why some scores were higher than others on the same topic. They began to share teaching techniques that worked. For example, a seventh grade teacher asked a fellow math teacher, "What are you doing with symmetry that gets those results?" She replied, "Maybe it was using those geo-mirrors at the beginning." Another teacher interrupted, "It's not that. It's the math journal writing you have them do at the end of each class. The kids really have to think and use the appropriate vocabulary when they put it

into their own words. I know, they write them in my class.”

To keep the focus on data analysis, lead teachers kept asking, “What do these scores tell us?” At this stage, the teachers were ready to dissect the scores to see what they revealed about the individual students and their own teaching. The positive outcome of this phase was the renewed focus on the relationships among the elements of the learning environment. The teachers were now asking the questions often voiced by the lead teachers, questions that helped to direct teacher thinking from negative reactions of blaming the test to a more positive approach in analyzing the test results. Did the test measure the learning objective? If the test item was adequate, how was my teaching? Did I teach the objectives in a manner that showed connections among the concepts and had relevance for these students? Did I give them enough opportunity, enough variety, to allow them to master the content? Was the content interesting to them? The intent here was not to teach the fragmented skills of individual objectives, but rather to make connections among the concepts with an eye to the “whole” or the big picture of interconnections.

The gradual change during this phase from challenging the test to focusing on their own instruction marked a change in the teachers’ approach to using the test data. Teachers began to comment that they felt they needed to present their material in other ways, using different formats, strategies, and terminology that helped the students make connections among the concepts and to their own lives. They accepted the value of the data for improving their teaching and realized that if their students were not understanding the material, perhaps their teaching strategies needed to be relevant to the lives of their students or to emerge from their students’ concerns and interests.

### **Phase 5: Accepting data as information, seeking solutions, and modifying instruction**

Teachers assumed a more accepting attitude about the information they were receiving after the third or fourth benchmarking period—about a full school year of benchmark testing each nine weeks. Following the previous meetings to analyze benchmark data, teachers implemented additional strategies focused on the curriculum objectives. As a result of reteaching and refining teaching strategies, teachers saw evidence that their teaching had caused the proficiency scores to rise. Lead teachers noted at this stage that the conversation among the teachers

moved toward a search for more effective teaching strategies. The lead teachers reminded teachers of the purpose: to glean information from the test results to find strategies for integrating the weak skills into meaningful contexts for the students who needed help.

In a team meeting with language arts teachers where the percentage of incorrect answers revealed that possessive nouns and plural nouns were specific skills that needed additional attention, teachers moved beyond awareness of the problem to discussions of strategies to address the weakness. The lead teacher encouraged integrating the skills into larger products that were meaningful to the students. One teacher reminded the group of her favorite strategy, “Get ‘em to write their own stories, and then go back and have them add a list of crazy, touchy-feely items as plural nouns and as possessives for each of their story characters, items like “fuzzy fingernails,” “goeey goose bumps,” and “wormy hair locks.” They love it!”

The teachers also looked for time in their teaching schedules and student assignments already on their teaching agendas where the students could apply the skills. They searched for cooperative group activities and remembered ideas from workshops where possessive nouns were acted out by some students while another student wrote the correct possessive form on the white board. They shared project ideas, such as letter writing to local newspapers and advertisements for garage sales, to help each other integrate the skills into meaningful contexts with other objectives they were currently teaching.

The lead teachers witnessed teachers gradually realizing the relationships among the isolated skills



*Benchmark test data brought teachers together to coordinate instructional modifications.*

PHOTO BY SUSAN TRIMBLE

of their lessons, the amount of time available to them for teaching the objectives, their teaching strategies, and what was tested. Lead teachers reported that at this phase teachers spent more time clarifying the alignment of their objectives with the test items. They realized that some content areas were not being addressed because they had not been a priority for the individual teacher; they faced the dilemma of omitting favorite units or making a judgment about the time devoted to those topics. They objectively weighed whether the activities involved actually addressed the concepts their students were required by the state of Georgia to master. They confirmed what the literature says: Curriculum objectives must become "the focus of daily instruction, even if this means dropping or revising favorite units" (Thomas, 2004).

---

### *The students were invited to analyze their thinking, their errors, and their own way of learning.*

---

At this stage teachers used the information from benchmark tests to change their own teaching strategies, often creatively and effectively. One teacher created individual practice packets of activities for each child on his or her specific areas of weakness. Another teacher conferenced with each child to determine an achievable personal goal for the next benchmark assessment. For example, one child who scored 30% was not expected to attain 70% on the next test but was encouraged to strive for 40%. A third teacher critiqued test items with his students, helping them to compare their individual responses to the answer key, and asking them to share with him why they marked the answer they chose. The class then discussed why answers were correct or incorrect and how they might like to learn the material next time it was taught. The students were invited to analyze their thinking, their errors, and their own way of learning. The teachers used the benchmark data in a variety of ways to relate more effectively to their students.

Teachers now saw the value of more frequent assessments in their classrooms. Grade level teams were encouraged to create brief quizzes and assessments to use weekly or biweekly. These mini-assessments covered only a handful of objectives and used a variety of assessment activities. They

were integrated into regular instruction as a very normal and non-stressful part of the lesson. By analyzing the results of these quick assessments, teachers received immediate feedback about the quality of their instruction.

One lead teacher commented, "Prior to implementation of formative assessments such as our benchmark testing, teachers were shooting in the dark. They were standing at the goal line in a dark gymnasium taking aim at a hoop they could not see. Now, we can see the goal and have a much better chance of ringing the basket!"

But using data to change instruction did not come easily. To sustain efforts of using formative test data to guide instruction required support and patience from administrators, lead teachers, central office personnel, and the teachers themselves. The Information Age with its copious amounts of data creates a high learning curve for all stakeholders who incorporate data to help students learn. The process pays big dividends if teachers act with sensitivity and with support through the phases of development.

### References

- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessments. *Phi Delta Kappan*, 80, 139-148.
- Georgia Public Education Report Cards. (2001-2002). System rankings by selected criteria. Retrieved June 9, 2003, from <http://accountability.doe.k12.ga.us/Report02/>
- Georgia School Systems Financial Reports. (2002). Revenue and expenditure reports. Retrieved May 17, 2003, from [http://db1.doe.k12.ga.us:8001/ows-bin/owa/fin\\_pack\\_revenue.entry\\_form](http://db1.doe.k12.ga.us:8001/ows-bin/owa/fin_pack_revenue.entry_form)
- Guskey, T. R. (2003). How classroom assessments improve learning. *Educational Leadership*, 60(5), 6-11.
- Jackson, A. W., & Davis, G. A. (2000). *Turning points 2000*. New York & Westerville, OH: Teachers College Press & National Middle School Association.
- Mayo, C. R., & Shotts, C. T. (2004). Test results: Creating a road map for NCLB. *Principal*, 83(5), 22-25.
- Southern Regional Education Board. (2001). *Making middle grades work*. Retrieved May 26, 2004, from <http://www.sreb.org/main/Publications/catalog/srebcatalog.asp>
- Thomas, R. S. (2004). Aligning schools to state standards. *Principal*, 83(5), 16-20. Retrieved May 26, 2004, from <http://www.naesp.org/ContentLoad.do?contentId=1248>