Can School District Classroom-Based Assessment Improve Performance on State Tests?

Nancy Baenen and Chuck Dulaney
Wake County Public Schools, Evaluation and Research

Paper presented at the annual meeting of the
American Educational Research Association, New Orleans, April 2000

Background

The North Carolina End of Grade (EOG) tests have been given each spring at grades 3-8. Third graders are also given a short pretest in the fall. The stakes connected with this testing have risen dramatically since they were first instituted in 1993, with school ratings, teacher bonuses, and student retention screening now tied to the results. In addition, the Wake County Public School System (WCPSS) has adopted a goal that 95% of our students in grades 3 and 8 will score "on grade level" (Levels III or IV) by 2003. About 77-87% currently score at those levels in grades 3 and 8. A major instructional assistance effort (with a budget of about $7 million) has been designed to provide additional assistance for students.

Given this context, the initial challenges for Evaluation and Research (E&R) were:

- How can we identify students who need additional assistance at the early grades (1-3) before the full EOG tests are given?

- What should teachers focus on when providing assistance, since EOG does not provide diagnostic information at the individual student level (and is not even very statistically reliable at the individual level)?

The obvious solution appeared to be the use of results from another measure used in WCPSS--classroom profiles that had been developed in reading, mathematics, and writing by the Curriculum and Instruction department staff with input from teachers. These profiles were based on the state curriculum and profile models (although WCPSS profiles are far more comprehensive than the state's model). While the WCPSS profiles have a great deal of face validity, one limitation is that E&R was not asked to be involved until after the instruments were developed. Reliability and validity checks were not considered in the developmental process. E&R was asked after the system was being implemented to design data capture scan forms, collect and analyze the data, and

help set cut scores for targets at the early grades. We had concerns based on the first year's data about whether teachers within and across schools were truly implementing the system in a consistent way, even after three days of training. Some of the analyses reported here provide data related to that concern.

A second more specific set of questions arose once the decision was made to try to use the profile results. Our results address these questions.

1. Do spring results on the grade 3 classroom profile data correlate well with the results of the third grade EOG tests?

2. Incoming third graders are a critical group for accomplishment of the school system achievement goal. Is the state's EOG fall third grade assessment or the spring of second grade classroom profile data a better measure to use in identifying students who may need help?

3. Should another EOG measure be provided to schools to identify progress mid-year in terms of the success of assistance provided?

Do spring results on the grade 3 classroom profile data correlate well with the results of the third grade EOG tests?

While EOG testing data have been available since 1993, the systematic use of standardized classroom literacy performance assessments began in a group of pilot schools in WCPSS in 1996-97. During 1997-98, training in the use of the assessment process was provided for primary grade teachers system-wide, although some schools did not fully participate in training until 1998-99. The form is included as Attachment 1.

A key component of the K-3 literacy assessment system is a system of "leveled" books and the use of "running records" to identify the level of book most appropriate for each student. Running records are conducted for each student at least quarterly with several pieces of data recorded on a profile card that is kept with a student's cumulative record. During the primary grades, most students move from Level 1 to Level 32. For purposes of screening for at or below grade level status, the key piece of information was identified as the book level at which a student could read with 90% accuracy and a "retelling" score of 3 or 4 on a four point scale.

In math, an observational matrix is provided for each grade that represents seven strands in the curriculum. Students who show on grade level performance in at least 6 of the 7 strands are considered to be "on grade level." A sample from grade 2 is included as Attachment 2.

In 1997-98, teachers were asked to use a scanable form to code the results of the running record conducted near the end of the school year. A sample form is included as Attachment 3. These scan forms were used to compile a system-wide summary of students' book levels. This summary was studied by the district reading instruction specialists, and a level of 23 or higher was chosen to represent "at or above grade level" at the end of second grade. A level of 29 or higher was identified as "at or above grade level" at the end of third grade. The application of these cut-points identified approximately 25% of students as below grade level at the end of second grade and approximately 35% of students as below grade level at the end of third grade. EOG test results at the end of third grade identified 21% of students as below grade level in 1998.

The Pearson correlation between the reading book level and the EOG test scores for the 6,009 third grade students for whom we had both pieces of information from spring 1998 was .64. This was slightly lower than the correlation between the third grade fall pretest and the EOG reading test (.74).

The data collection process was repeated in 1998-99 and showed that:

- Approximately 20% of second grade students were below grade level according to the literacy assessment profile;
- Approximately 28% of third grade students were below grade level according to the literacy assessment profile;
- Approximately 20% of third grade students were below grade level according to the spring EOG results;
- A Pearson correlation of .63 existed between book level and EOG reading score for 6,944 third grade students; and
- For students scoring below grade level (Level 29) on the literacy assessment, the correlation with EOG was lower at .51.

Since teachers were still being trained in 1998-99 in the use of the literacy performance assessment system, the correlation between performance assessment and EOG multiple-choice score was considered encouraging. We hope that as teachers gain

expertise in the administration and interpretation of running records, the correlation will rise and judgements based upon the literacy profile data will be made with greater confidence.

Incoming third graders are a critical group for accomplishment of the school system achievement goal. Is the state's EOG fall third grade assessment or the spring of second grade classroom profile data a better measure to use in identifying students who may need help?

We ran Pearson correlations between the spring 1999 grade 3 EOG results and the fall of grade 3 EOG results as well as spring 1998 grade 2 classroom profile results.

- In reading, correlations with spring grade 3 EOGs were .72 with fall 3rd grade EOGs and .67 with spring of grade 2 classroom profiles.

- In math, correlations were .76 with the fall 3rd grade EOG and .59 for the spring of 2nd grade classroom profile.

Thus, both had reasonable correlations with the third grade spring EOGs, with some evidence that the fall EOG was a better predictor, especially in math. Of course, the EOG fall and spring tests are both multiple choice and similar in structure and format, and the tests are closer together in time, so this seems reasonable from that standpoint. However, the standard error on the fall EOG test is larger than the EOG, primarily because the test is much shorter, so the results are somewhat surprising. The next two charts illustrate the percentage of students who were classified correctly and incorrectly on the spring EOG based on the fall EOG and spring classroom profile data. Even though the correlations with the third grade EOG appeared acceptable for both instruments, about 16% of the students were misclassified with both instruments in terms of where they ultimately scored on EOG in the spring of their third grade year.

The good news is that a higher percentage of students scored on grade level after scoring low on the pretest or class profile than the reverse (in which students scored at or above grade level initially and fell below grade level on the spring of grade 3 EOG). The difference in percentages was more positive for the fall of 3rd grade pretest than the classroom profiles. We hope students who improved their status did so as a result of assistance provided, although we suspect some was due to measurement error.

We plan to re-run these analyses once spring 2000 data are in and hope correlations and student classifications improve with an extra year of use of the instruments (especially the classroom profiles).

Spring EOG 3rd Grade Results in Relation to Fall EOG 3rd Grade Results:

Reading

	EOG Status Spring Grade 3, N=7,221	
Fall Grade 3 Status	On Grade Level: III or IV	Not on Grade Level: I or II
Level III or IV	5,064 (70.1%)	393 (5.4%)
Level I or II	785 (10.9%)	979 (13.6%)

Mathematics

	EOG Status Spring Grade 3, N=7,248	
Fall Grade 3 Status	On Grade Level: III or IV	Not on Grade Level: I or II
Level III or IV	5,117 (70.6%)	650 (9.0%)
Level I or II	512 (7.1%)	969 (13.4%)

Spring EOG 3rd Grade Results in Relation to Spring Grade 2 Class Profile Results:

Reading (Book Levels)

	EOG Status Spring Grade 3, N=5,747	
Spring Grade 2 Status	On Grade Level: III or IV	Not on Grade Level: I or II
Above Cut (Level 23)	4,109 (71.5%)	398 (6.9%)
Below Cut	547 (9.5%)	693 (12.1%)

Mathematics

	EOG Status Spring Grade 3	
Spring Grade 2 Status	On Grade Level: III or IV	Not on Grade Level: I or II
Above Cut*	4,463 (69.5%)	514 (8.0%)
Below Cut	575 (9.0%)	874 (13.6%)

*Level III or IV on 6 or 7 of 7 instructional strands

We ran correlations of the fall of grade 3 EOG results for 1999 and the spring of grade 2 data capture results for 1999 and found the two measures (fall 3rd grade EOG and spring 2nd classroom profiles) had a moderate positive correlation with each other in reading (.63) and in math (.54). As shown in the next figures, about 17%-18% of the students were low on only one or the other of the two measures (about 1,000 students in reading and 1,200 in math).

Fall EOG 3rd Grade Results in Relation to Spring Grade 2 Class Profile Results:

Reading

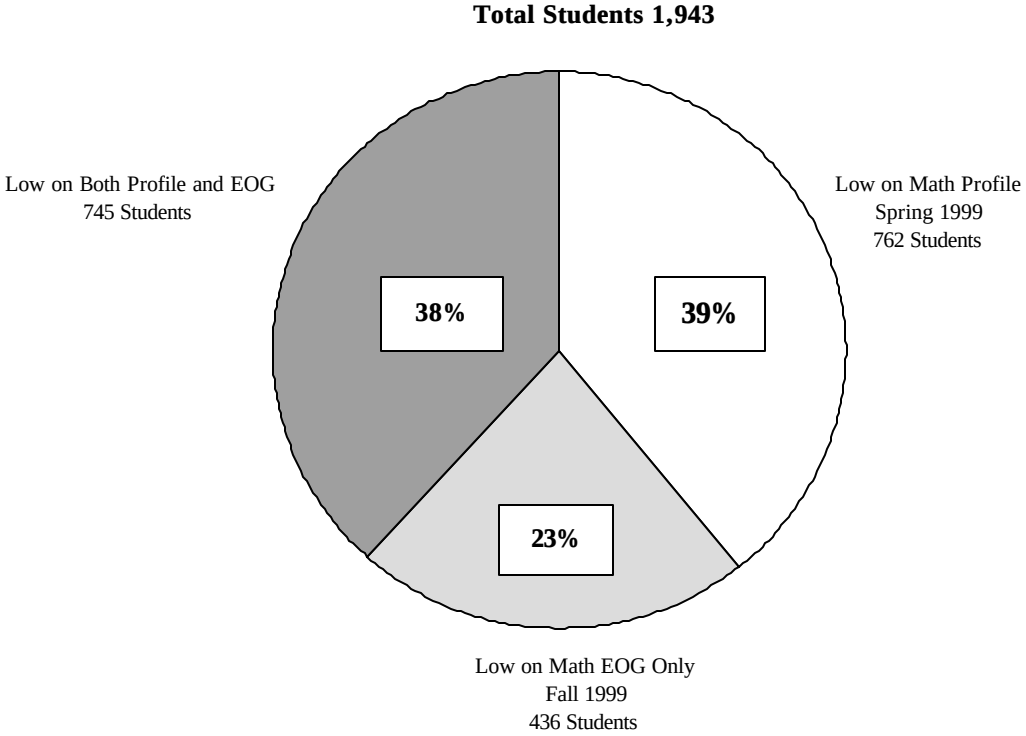
	EOG Status Fall Grade 3, N=6,473	
Spring Grade 2 Status	On Grade Level: III or IV	Not on Grade Level: I or II
Above Cut (Level 23)	4,732 (73.1%)	650 (10.0%)
Below Cut	406 (6.3%)	685 (10.6%)

Mathematics

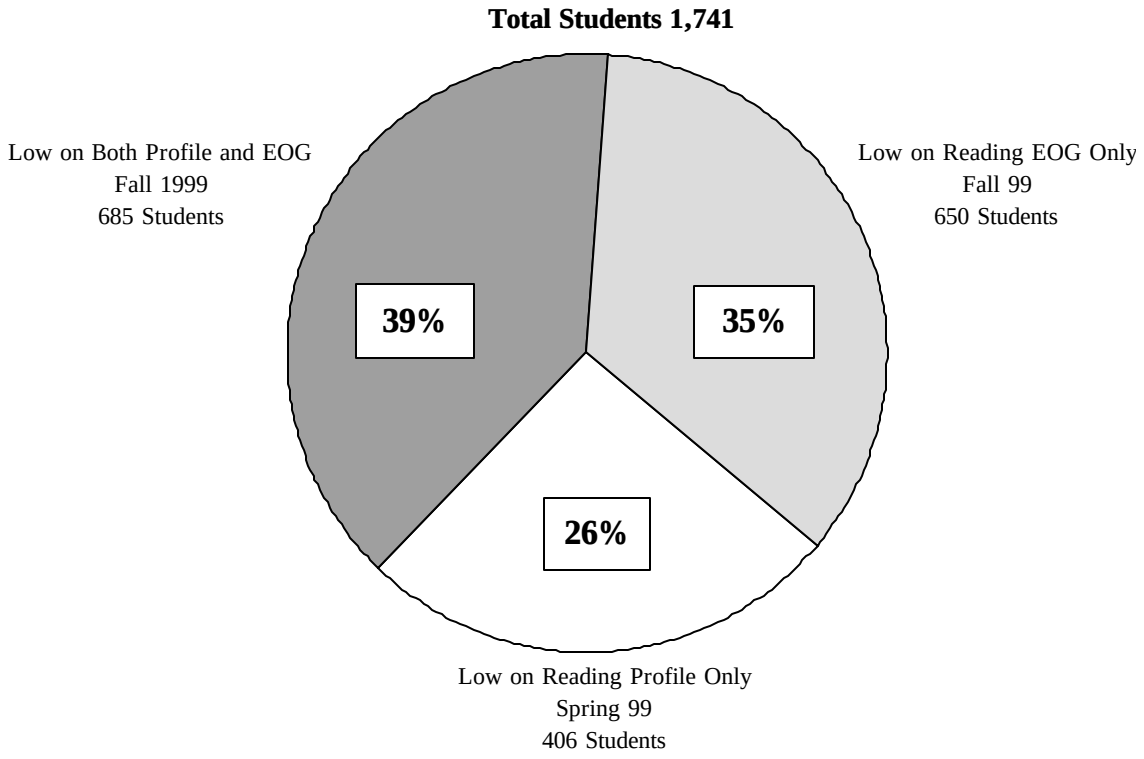
	EOG Status Fall Grade 3, N=6,642	
Spring Grade 2 Status	On Grade Level: III or IV	Not on Grade Level: I or II
Above Cut (Level 23)	4,699 (70.8%)	436 (6.6%)
Below Cut	762 (11.5%)	745 (11.2%)

However, when we looked at only those students identified as low by either instrument to see how much they overlapped, the same students were identified as low by both instruments for only 39% of the cases in reading and 38% in math.

STUDENTS IDENTIFIED AS LOW ACHIEVERS IN MATH, FALL 1999



STUDENTS IDENTIFIED AS LOW ACHIEVERS IN READING, FALL 1999



Our decision for this year at least was to provide schools with the combined list of students who might need additional assistance this year and ask them to assess whether the students truly needed additional help to reach grade level on the EOG in the spring of third grade. We did provide recommendations based on whether students were low on both instruments (definitely provide assistance) or just one (consider class work from fall plus additional classroom profile information that was current). This appeared to be the only prudent decision given the 95% goal. We thought it would be better to have too many students considered for help than too few. However, this was an expensive decision, since schools were given allocations for the new Accelerated Learning Program (ALP) based on students who scored low on the assessments. The number of students potentially eligible almost doubled for grade 3 compared to selecting one instrument or the other.

Should another measure be provided to schools to determine mid-year status, progress made, and areas for further assistance?

This year WCPSS added yet another measure of performance--mid-year optional testing using "secure for local use" forms of the EOG tests. The optional testing was provided in response to multiple requests over several years, and was only possible because the state released some forms of the state tests used in prior years for local use on a secure basis. We had concerns about mis-use of data from administration of one form of the EOG, since the form would not cover the full curriculum in any great depth, but training highlighted appropriate and inappropriate uses of the test. Results provided an item analysis for groups of students and rosters of student responses to test items, but no total score.

The primary benefit of the optional mid-year testing appears to be increased teacher familiarity with the design of NC EOG test items and the structure of the EOG tests. Principals report that teachers have a new understanding of the expectations of the test and the need to target instructional time to the goals and objectives of the North Carolina Standard Course of Study. Hopefully, students have benefited from an opportunity to discuss specific test items in the classroom.

The item analyses and student rosters do provide some diagnostic information to teachers based upon patterns of questions missed and incorrect answers chosen by students. However, it is questionable whether additional testing provides greater incentive and/or ability to identify low-performing students in need of special assistance.

Conclusions

What do our results suggest about whether multiple measures are a boon, bane, or baloney? Or, to throw in another “b”, are multiple measures a boondoggle—a pointless, unnecessary waste of time and money?

From the viewpoint of most teachers, collecting classroom profile information to the EOGs is a positive move (a boon) for several reasons.

- Schools see the EOG tests as "traditional", non-diagnostic, and subtracting from instructional time. For the most part, they see the classroom profiles (performance assessments), which actually take more time, as instructional and diagnostic, therefore less wasteful of their time.

- While both assessments measure students' mastery of the North Carolina curriculum, they are different in that:

  - EOG is primarily multiple choice items, with some open-ended items at some grades and writing assessments at grades 3, 5, 8. Multiple choice results are returned quickly, but open-ended and writing results take months to be returned.

  - The math, literacy, and writing profile cards are performance assessments primarily, and rely on teacher observation, student class work, and some standardized methods and instruments (e.g., running records for reading).

Our impression from a system-level perspective at this point is also that multiple measures are a positive addition—albeit an expensive one. If our goal is to determine whether students are truly mastering the material we desire, measures that are closer to classroom practice and which provide teachers with clues about what students do not understand provide more useful information than our statewide assessment that is not diagnostic at the individual level. The information can be quite helpful. It would clearly be nice if these measurements were a little more exact, and if they were validated before being placed in use, but they do provide more clear-cut standards than teacher judgement alone. The expensive part is that we identify more students as potentially needing help than we might if the measurements were more reliable. We would appreciate input on other ways to narrow the field of students likely to benefit from assistance.

Attachments

The web copy of this report does not contain Attachments 1, 2 and 3. For paper copies, call Wake County Public School System, Department of Evaluation and Research (919) 850-1840.