Relationships Among Inquiry-Based Science, Elementary/Middle Grades End-of-Grade Achievement Scores, and Science-Related Student Behaviors in a North Carolina School District
Teachers and Scientists Collaborating (TASC)

Dave Smith, TASC Program Director
Charles Coble, Co-founder and Partner, The Third Mile Group
Gary Ybarra, TASC PI

Context and "Student Success"

As a partnership between multiple school districts and Duke University, TASC was subject to school district partners' constraints in defining student success and establishing measures. Student success was defined generally, by students' performance on the North Carolina end-of-grade (EOG) science test, given in grades 5 and 8, and aligned to the science content in the North Carolina Standards. The partnership selected and used research-based curriculum aligned at all grade levels (including grade levels not tested) with the state science standards. All PD focused on that alignment such that trainers referred each activity to a particular standard or its subset. In these ways, the program aligned the intervention with students' success on the standards that would be measured, generally for whole populations.

Early on, staff argued for defining and measuring science process skills as a component of student success. Pre/post-tests of teachers and students to uncover associations between treatment and gains in these skills yielded inconclusive results. In the one or two days that partner districts allowed for all PD on a curriculum unit, few teachers measurably improved these skills. Results were included in an annual report, but a low test return rate and low N values made analysis impractical. PD presenters modeled best practices, including notebook usage, coaching students for problem solving, claims and evidence, questioning skills, math integration, and science content (concepts and vocabulary), but most of training necessarily focused on the mechanics of conducting activities. Classroom observers saw few teachers incorporate practices to promote science process skills. In one or two days of PD, "mechanical usage" of curriculum units appears to be the primary skill that teachers learned.

The school environment also works against defining student success in terms of science process. EOG science tests given in grades 5 and 8 are the primary measure to which science teaching is accountable in partner districts. These tests primarily measure factual knowledge. Teachers tend to respond by focusing on vocabulary first, presenting student activities as teacher demonstrations to save time, or conducting occasional activities without wrapping up science content. All TASC districts emphasize reading and mathematics to the neglect of science such that, at most, science is taught 45 minutes per day, three days/week. The average across all districts is closer to two 45-minute sessions/week. The average teacher using a 9-week curriculum unit completes ½ - ¾ of modules. Partnership administrators generally expect that, regardless of the above conditions, exposure to good curriculum taught by a trained teacher *should* measurably increase students' recall of vocabulary. As a result of all of these conditions, the partnership defined "student success" as follows:

Student Success In the Aggregate:

- Science EOG scores for students receiving a larger amount of treatment statistically significantly above scores for students who receive a lesser amount of treatment. Treatment was measured both in the cumulative number of quarters of exposure to a research-based curriculum unit, and in cumulative exposure to trained teachers.

- Mathematics and reading EOG scores for students receiving a larger amount of treatment statistically significantly above scores for students who received a lesser amount of treatment (measured as stated above).

- Mean science EOG score for a district above the state mean science EOG score

- Students show evidence of positive change on the following desirable science-related behaviors:

    - More time on task
    - Greater level of interest in science-related activities
    - Asking more questions about science
    - Talking more about science
    - Working more cooperatively in groups
    - Making more verbal or written observations

<u>Student Success Among Groups</u>:

- Students eligible for free and reduced lunch show significantly greater increases in EOG scores (science, reading, and mathematics) than students who are ineligible for free and reduced lunch.

Design for Measuring Student Success

The research was designed to determine whether cumulative effects on student success exist where students received treatment for multiple years prior to the test year. To measure cumulative treatment, the project gathered historical student and teacher data and administered an on-line teacher survey in four districts. Each student was represented by a unique anonymous identifier, demographic data, test scores, survey data, and teacher history for the current year and the four prior years. The study also represented all teachers in the student's history with a unique anonymous identifier. Data were collected on teachers' professional development history, kit usage history, and demographics. 63% of participating teachers in the 4 research districts also returned an on-line survey. Survey questions included a request to rate changes in science-related behaviors seen throughout the program. The student data, teacher data, and teacher survey were combined to create the student profile for analysis.

Data on most of the districts has not yet been validated. Chatham County is the subject of this report because a sufficient amount of data could be validated. Chatham County Schools is a low-wealth, low-performing rural and suburban school district with a high number of Hispanic students where English is their second language, and as such represents demographic and other considerations in the three other districts for which data are gathered. Demographic data were collected to control for variables such as family income, students' math and language arts scores, teachers' cumulative experience, and fidelity of implementation. Null hypotheses related to student success are grouped under these Topics of Concern:

I. Relationship between student achievement (EOG score) and amount of PD (hypothesis1);

II. Relationship between student achievement (EOG score) and exposure to research-based curriculum and activities (hypotheses 2 and 3);

III. Relationships among student achievement family income, and treatment (hypothesis 4);

IV. Relationships among student achievement, years of teaching, and treatment (hypotheses 5 and 6);

V. Changes in student behaviors in classrooms of teachers who received TASC training (hypotheses 7 and 8).

The study set criteria to reject 8 null hypotheses. Analysis reported the probability that data support each null hypothesis. Hypotheses are as follows:

1. There is a significant correlation between the cumulative amount of PD that students' present and past teachers have in use of science kits and students' EOG scores in 2008 for: a. grade 5 students and b. grade 8 students.

2. Students' cumulative exposure to science kit usage throughout their academic histories significantly correlates with students' scores on EOG science tests in 2008 for: a. grade 5 and b. grade 8.

3. In 2008, a. grade 5, b. grade 8 students with high cumulative exposure to science kit usage through their academic history had a statistically significant higher mean EOG science score than the state mean science score.

4. Correlations between cumulative exposure to science kits and state end-of-grade (EOG) scores on a.) science, b.) mathematics and c.) reading tests for students eligible for free and reduced lunch are significantly higher than those correlations for students who are ineligible for free and reduced lunch.

5. Among students with cumulative exposure to science kits ≥ 3 quarters of instruction, there is a significant difference in science EOG scores between students taught (in the year the student took the test) by a teacher with ≥ 5 years of teaching experience and those taught by a teacher with < 5 years of teaching experience.

6. Among students with cumulative exposure to teachers whose kit training totals ≥ 3 days in science kit usage, there is a significant difference in EOG scores between students taught in the year that they took the test by teachers with ≥ 5 years of teaching experience and those taught by teachers with < 5 years of teaching experience.

7. Teachers who had ≥ 3 days of professional development report significantly more positive change in desirable science-related behaviors than teachers with < 3 days of professional development on science kits. Desirable science-related behaviors are:

    - More time on task
    - Greater level of interest in science-related activities
    - Asking more questions about science
    - Talking more about science
    - Working more cooperatively in groups
    - Making more verbal or written observations

8. Teachers who used science kits for ≥ 3 quarters report significantly more positive change on desirable science-related behaviors than teachers who used science kits for < 3 quarters (see behaviors above).

Findings

Data were analyzed separately for grades 5 and 8 in 2008, the year for which complete data could be gathered and verified. Not all hypotheses were supported by the data and analyses at both grade levels. Analysis of Topics of Concern and related hypotheses is summarized below.

Student Achievement (EOG score) and Amount of PD

High-Low T-tests of grade 5 results showed a positive and significant correlation between students' 2008 EOG scores and cumulative PD that students' present and past teachers had in use of science kits. Grade 8 in that year did not show significant correlation. The pool of teachers in Chatham County who had not had significant TASC PD was low, making a statistically fair comparison difficult. Therefore, TASC makes no claim that PD for current and past teachers of grade 8 students is associated with higher grade 8 EOG scores. However, 61.6% of Chatham County grade 8 students scored at proficiency on EOG test in science, whereas statewide, 52.3% of students were proficient.

Relationship Between Student Achievement (EOG Score) and Exposure to Science Kits

Looking at students' science EOG scores in 2008 for grade 5 and 8 versus cumulative exposure to science kit usage throughout their academic histories, the study found significant correlatations both within district and in comparison with state mean science EOG score.

Comparing EOG scores for students who had high exposure to science kits with those who had low exposure, grade 5 students had a small pool of students with low exposure, making the T-test non-significant. However, high and low exposure populations differed by about 3 science scale scores. The

grade 8 test of significance showed a positive and significant correlation between cumulative kit exposure and student's scores on EOG science test.

Regarding a comparison between the district's mean science EOG score versus the state mean, Chatham County grade 5 students' mean performance was 0.951 points below the state mean, and statistically significant. On the other hand, Chatham County grade 8 students' mean performance was 3.77 points above the state mean, and also significant. The district's 2008 overall proficiency rate for both grades combined was 48.91% compared with the statewide 2008 proficiency rate of 45.89%.

Relationships among Family Income, Student Achievement, and Treatment

The study also compared two populations – those on free and reduced lunch, and those not on free and reduced lunch – to determine whether test scores and treatment correlate differently in these two groups. The treatment measured was cumulative exposure to science kits, and again the measure of student success was state end-of-grade (EOG) scores, this time in grade 5 and 8 science, mathematics, and reading. Researchers hypothesized that treatment and test scores would correlate better for students eligible for free and reduced lunch than for students who are ineligible for free and reduced lunch. Such evidence might support an assertion that TASC contributed to closing achievement gaps between students in high and low income families. However, the study found no significant difference between income groups with regard to correlation between exposure to science kits and EOG test scores in science, reading or mathematics.

Relationships among Years of Teaching, Student Achievement, and Treatment

The study also examined the impact of teaching experience of the teacher who taught the student in the year of the test on EOG scores. Grade 5 teacher data were used because too few Chatham County grade 8 teachers had taught < 5 years. Only students who had ≥ 3 cumulative quarters of kit usage by grade 5 were included. A second related question examined effects that the teacher's years of experience may have on science EOG scores for students whose past and current teachers have cumulative exposure to PD of ≥ 3 sessions. Findings were significant. Years of teaching experience held by the teacher in the year of the test strongly correlated with students' science EOG scores in grade 5.

Changes in Student Behaviors in Classrooms of TASC Teachers

The study explored two possible impacts on students' science-related behaviors. One question examined associations between the amount of a teacher's cumulative PD and their perception of their students' science related behaviors. The second question explored associations between the teacher's amount of usage of research-based curriculum and their perceptions of students' science related behaviors. Data were gathered through an on-line survey rated changes in six science-related behaviors:

a. time on task
b. level of interest in science-related activities
c. asking questions about science
d. talking about science
e. working cooperatively in groups
f. making verbal or written observations

Survey response options and their assigned weights were as follows: Much Less (1), Less (2), No Change (3), More (4), Much More (5). Data were tabulated and analyzed for reported changes in student behaviors in the six areas above. For analysis, responses were grouped into low change ("Much Less", "Less," and "No Change") and high change ("More," and "Much More.") Each question was treated as separate sub-hypothesis.

Regarding amount of PD that teachers received v. perceptions of students' science-related behaviors, teachers reported significantly more positive change in the following:

- exposure to science kits correlated at p = 0.0127 with "time on task."
- exposure to science kits correlated at p = 0.0256 with "student interest in science"
- exposure to science kits correlated at p = 0.0044 with "amount of student-generated questions about science."

Regarding amount of usage of research-based curriculum v. perceptions of students' science-related behaviors, teachers saw more positive change in the following:

- amount of PD that teachers had significantly correlated at p = 0.0392 with "work in groups."

No other items were significantly correlated.

How Findings Inform Our MSP Work

Findings focus staff on issues in achieving program goals in the context described above. Areas for improvement lie in PD design and administrative changes. Teachers in the early years of the profession need more support for their students to achieve higher performance. All teachers need more knowledge and skills in how to close achievement gaps between levels of family income. Attention and support are needed to help teachers use science kits to generate positive science-related behaviors.

During 2009 and 2010 PD, trainers shifted somewhat from "mechanical usage" of kits to "unpacking content," spending twice the time on notebook prompts, vocabulary review after (and connected to) activities. Science EOG scores and proficiency rates in Spring 2009 increased. Until districts provide more time for professional development, training time on "unpacking content" will come from time on setting up activities. This will be a positive for experienced teachers, but will pose difficulties for new teachers and those using kits for the first time. TASC staff will continue this shift in emphasis, but more study is needed to learn whether these changes may contribute to gains.

With regard to policy and the North Carolina end-of-grade science test, TASC staff continue to seek allies in the state to advocate for raising the level of thinking skills assessed in science. The current emphasis on testing at knowledge level works against inquiry-oriented science.

Evaluation/Research Design

The research was begun when state and school district data collection systems were in transition and somewhat disorganized. Validation of these data has been extremely troublesome because of discrepancies between data coding systems, discrepancies between school districts, difficulty in gaining permissions, and inconsistent student and teacher identification.

The data currently gathered do not leave open the possibility to disaggregate test data by item. However, in some cases, data disaggregated by item are available to teachers when they receive test results. In addition, all of the districts under study have begun using science benchmark tests for which items are routinely disaggregated by topic. If it becomes possible to work with districts on the disaggregated data, topic-by-topic, either on EOG results or on benchmark results, it will be possible to refocus professional development more narrowly to emphasize concepts and knowledge that students have more difficulty in learning. It may also be possible to refocus training on teaching practices more narrowly. If so, the prospect of gains that could be realized may be sufficient for districts to allow teachers to take additional days of science professional development. This will be particularly attractive if linkages to mathematics and language arts can be strengthened through such training. The case has already been made that best practices for mathematics and science are closely allied.

Insights

Although broad analysis of correlations between treatment and EOG scores has been of interest to districts, difficulties in validating such data have been overwhelming. Studies undertaken so far do quantify associations between trained teachers using research-based curriculum and their students' scores. In North Carolina, few districts are building district-wide inquiry-based science programs, and the question of efficacy of such programs for test scores is the primary question for administrators deciding

whether or not to take that step. In that regard, the current study provides associations, not evidence of cause, but it will be informative nonetheless.

The program has reached a level of maturity in the districts where it operates such that more specific training for both teachers and administrators is warranted and welcomed, but these districts will have to allocate the time to accomplish this. It will likely be possible to work with districts on disaggregated data, topic-by-topic, either on EOG results or on benchmark results. If so, this more tightly focused study will guide professional development to emphasize the specific concepts and knowledge that students have more difficulty in learning. Such future studies may also guide training more narrowly focused on teaching practice. Training sessions in 2009 and 2010 began to explore more linkages to mathematics and language arts, and the case has been made that best practices for mathematics and science are closely allied. The prospect of the gains from such training, informed by study of results, may prompt districts to allow teachers to take additional days of science professional development.