Technical Report 1B
Procedures and limitations for linking teachers and students using the School Activity Report (SAR)
February 12, 2009

Data collected by schools do not provide a direct link between teacher and student. Such a link would permit innovative analyses of the relationships between teacher characteristics and student outcomes. The School Activity Report, described below, is sometimes used as a proxy for classroom roster data. This approach has been explored as a means of linking individual teachers to students in classrooms, since classroom roster data are not available to date. Researchers should be aware of significant limitations and caveats to this approach.

The Data
The School Activity Report (SAR) contains data for all personnel employed by the public school system who have direct student contact in a classroom or non-classroom activity for which a state course code or personnel assignment type exists (See SAR codebook, attached). Activities include traditional academic classes as well as non-class events, such as study hall, lunch period, or counseling. This file includes activities that meet all year, such as English 1, and those that meet for only part of the year, such as a month-long drama workshop.   The data for personnel include fields for last name, first name, middle name, and social security number as well as the district and school in which that person worked.[1] The NCERDC has already created a randomized identifier for each person in this file that allows researchers to follow a teacher over time and across the School Activity Report and Teacher Salary and Licensure data. In this document, we describe NCERDC efforts   to assign that identifier to the instructor name in the ABC data so that researchers can link teacher information to student data.

The ABC data are the End of Grade and End of Course tests with a record for each student. The data include a field called "instructor name" which is the name of the person monitoring the exam. *We do not know whether the exam monitor is the teacher for those students. Although it is often the case that teachers monitor the tests for the classes that they have taught,, there are other instances where, for example, all the eighth graders take a test together, or where a fourth grader takes math in a class from someone other than her homeroom teacher. Additional validation of teacher-student links needs to be done by the NCERDC and by individual researchers.* However, matching these files is the only way we have to begin to link teachers and students. This matching is an essential first step in the process.

---

[1] One can reliably link the SAR data to the Licensure data because the social security number in SAR is validated against that in the Licensure data.

The Matching Process

On the ABC test forms, instructors did not have guidelines in completing the instructor name field, so the format of the names has a great deal of variation (last name-first name; first name-last name; last name-nickname; last name-middle name and so forth). This dataset has only one field for name, did not distinguish between first and last names, and some scanning errors did occur. Due to these variations, the NCERDC used many different methods for matching these files. The basic approach is as follows: First, the SAR and ABC data are limited to one record per LEA, school code and instructor name. This is the level at which the data are matched. Then, the NCERDC started with the most conservative match (e.g., full name), set those names aside into a matched data set and excluded them from further iterations of matches. With the names that did not match the first time, the NCERDC tried to match using other information (e.g., last name, first initial), and set those matches aside, and so forth. See below for more detail.

After matching these names, the NCERDC reviewed all of the matches to ensure that the same teacher id was not assigned to different people in the ABC data; NCERDC then created a *matchtype* variable that indicates the method used to identify the match. With this variable, researchers can choose whether to accept all methods of matching names.

Those That Will Not Match

Not all schools in the ABC data are in the SAR. For example, SAR does not include DHHS schools, DOJ schools, and many charter schools. On the ABC form, some people left the *instructor name* field blank or completed it with something other than a teacher's name such as "Visitor" "Grade 8" or "Longview School." Furthermore, because schools complete the SAR in the fall and collect the ABC data in the spring, teachers who transferred schools or changed their names during the year will not match.

Overview of Limitations:

For elementary school students, who are supposed to have only one teacher, it is often assumed that the teacher listed on the End of Grade test is in fact that student's teacher. This may not always be the case. In some schools, students move across classrooms for instruction in different subjects, particularly in fourth and fifth grades. If this is the case, the teacher administering the EOG test may not be the instructor for reading and/or math for the students in that classroom.

For middle school and high school students, only a fraction of the teachers listed on the End of Grade/End of Course tests are actually the students' teachers in that subject. For End of Grade exams in middle grades where one teacher teaches English (reading) and another teaches math, only one teacher is identified on the test record that contains both reading and math scores. End of Course exams include the instructor proctoring the exam, and this is not necessarily the teacher of that course. In some cases, using the classroom-level data, we can verify that the teacher attached to the exam is in fact the teacher of the course. We have tested this with Algebra I, identified about 20% of instructors on the Algebra I exam as students' Algebra I teachers.

Limitations in Student-Teacher Matching, End of Grade Tests

Simple matching by TEACHID:
Matching teacher records and student records by TEACHID will result in a link between the teacher who administered the test and the student taking the test**. Note:** This teacher will not necessarily have taught the student.

End of grade records contain scores for both math and reading. Especially in middle grades, it is unlikely that the student had the same teacher for both subjects, however only one teacher is identified in the test record.

In 2006, the number of teachers matched to students in each End of Grade test who (as found in the SAR data) taught a self-contained classroom declines significantly after grade 5:

> Grade 3: 4,314 teachers identified in the Grade 3 EOG taught self-contained classes (SAR SUBJCT=0000);
> Grade 4: 3,746
> Grade 5: 3,526
> Grade 6: 185
> Grade 7: 26
> Grade 8: 26

Note: Even teachers reported as having a self-contained classroom are likely to have students moving across classrooms for instruction in different subjects, particularly in the 4th and 5th grades.

An Example
In one project, researchers were interested in the effect of licensure and certifications of middle school math teachers on their students. We address the question of whether we can match middle school math teachers to their students.

To answer this question, we examined the math course offerings at one middle school:

End of Grade Data:
There are 359 students in the grade 7 EOG data from this school.

SAR (Classroom level) Data:
There are 23 sections of 7th grade math (SUBJCT=2001 7) according to the SAR student count file, totaling 355 students.

Now let's look at the teaching pattern of one teacher:

According to the SAR personnel file, this teacher taught 4 math courses: (2) 2001 7 (7th grade math) and (2) 2001 8 (8th grade math).
SAR student count shows the composition of these 5 courses to be:
      2001 7 00006: 23 students (4WF, 6WM, 7BF, 3BM, 1HF, 1AF, 1MF)
      2001 7 00007: 20 students (8WF, 2WM, 4BF, 3BM, 1HF, 1HM, 1AF)
      2001 8 00007: 15 students (2WF, 2WM, 1BF, 5BM, 3HF, 1AF, 1AM)
      2001 8 00008: 24 students (7WF, 7WM, 3BF, 3BM, 3HM, 1AM)

EOG 7:
This teacher was linked to19 students (4WF, 4WM, 6BF, 1BM, 1HM, 1HF, 1AF, 1MF), which appears to be **2001 7 00006**. However, this teacher is not linked to the other section of 7th grade math and does not appear in the EOG 8 dataset.

Out of the 82 students taught by this particular teacher, it was only possible to link this teacher with 19 students.

What we know: It appears that the math teacher listed in the End of Grade file is in fact the math teacher who taught the students in course 2001 7 0006 during the current year.

What we don't know: For students taking the End of Grade test with a teacher other than their math teacher, we cannot tell which teacher they had.
Summary for one county:
- 25,425 students took an end of grade exam for middle school grades (6-8).
- 4,849 student end of grade records were linked to 197 math teachers (are we sure that these students are linked with the teachers who taught them math?). The other 20,576 most likely took the test in a different class (with an English teacher, for example).