

***North Carolina Education Research Data Center
Technical Report #5: Linking Teachers in the Course Membership Data to
Teachers in the School Activity Report.
February 12, 2009***

New data collected by schools provides a direct link between teachers and students, permitting innovative analyses. Before 2006, it was possible to link students to classroom teachers using the School Activity Report (SAR), which contains information that may be used as a proxy for classroom roster data. The use of this proxy information for linking individual teachers to students in classrooms is described in detail in Technical Report #1B, along with the limitations and caveats of this technique. One of the main limitations of linking the SAR with test score data is that it is not possible to establish whether the teacher listed in a student's test score record was actually that student's teacher for that test subject. Conversely, the teacher listed in a student's Course Membership record is the teacher who taught that particular course. Therefore, the linking of teacher information from the Course Membership files is a critically improved resource for research that seeks to link specific teachers to specific student outcomes. The NCERDC uses information from the School Activity Report to assign a teacher ID to each student's course information in the Course Membership file. This file, containing both individual student and teacher IDs, can then be linked to student with the student test score files.

This report documents the methods the North Carolina Education Research Data Center (NCERDC) employed to link the instructors in the course membership (CM) data to the personnel file in the data to the School Activity Report (SAR). Using the method described below, the NCERDC obtained (approximately) an 80% matching rate.

The Matching Process

The only information identifying teachers in the course membership data is a single teacher name variable. Unlike the SAR personnel data, which includes formatted first and last names, the format of the names in the student-level data has a great deal of variation (last name-first name; first name-last name; last name-nickname; last name-middle name and so forth). Because this dataset has only one field for name and did not distinguish between first and last names, the NCERDC used many different methods for matching these files. The basic approach is as follows: First, the teacher and student datasets 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 data and created a match type 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 student data are in the SAR. For example, SAR does not include DHHS schools, DOJ schools, and many charter schools. In some cases, the *teacher name* field is blank or contains something other than a teacher's name such as "Visitor" or "6th Remedial Math." Furthermore, because the data are not collected at the same time, teachers who transfer schools or change their names between collection points will not match.

Finding Matches

All matches are done within fiscal year, LEA, and school code. To address the various formats of teacher name in the course membership, a variety of match criteria were employed. Note that if there are multiple records in the SAR data based on the criteria, the teacher name is excluded from the possibility of a match (i.e., if there is an Alex Smith and an Ann Smith in one school, then match types 4, 5, 7, 10, 11 would not be possible).

The following 13-character teacher names would match the SAR instructor

First = ALEX

Middle = JAMES

Last = SMITH

Matchtype 1: Last, First	“SMITH, ALEX”
Matchtype 2: Last First	“SMITH ALEX”
Matchtype 3: Last First Middle Initial	“SMITH ALEX J”
Matchtype 4: Last First Initial	“SMITH A”
Matchtype 5: Title First Last	“MR A SMITH”
Matchtype 6: First Last	“ALEX SMITH”
Matchtype 7: First Initial Last	“A SMITH”
Matchtype 8: Last First Middle	“SMITH ALEX JA”
Matchtype 9: First Middle Last	“ALEX JAMES SM”
Matchtype 10: Title Last	“MR SMITH”
Matchtype 11: Last	“SMITH”
Matchtype 12: Middle Last	“JAMES SMITH”
Matchtype 13: Last, Close-First	“SMITH, ALEXAN”
Matchtype 14: First Close-Last	“ALEX SMITT”
Matchtype 15: Close-First Last	“ALEXA SMITH”
Matchtype 16: Close-Last First	“SMITT ALEX”
Matchtype 17: Last Middle	“SMITH JAMES”