VALUE-ADDED
MEASUREMENT Services
at American Institutes for Research

Determining Attribution

Holding Teachers Accountable for Student Growth

AIR®
American Institutes for Research®

Determining Attribution:
Holding Teachers Accountable for Student Growth

August 2013

Hella Bel Hadj Amor, Ph.D.
Natalie Tucker
Umut Özek, Ph.D.

VALUE-ADDED MEASUREMENT Services
at American Institutes for Research

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202-403-5000 | TTY 877-334-3499
www.educatortalent.org

AIR® AMERICAN INSTITUTES FOR RESEARCH® | www.air.org

1979_08/13

Acknowledgments

The authors gratefully acknowledge useful comments received from Monica Young, Joshua Marland, and Amy McIntosh (New York State Education Department and Regents Research Fund) as well as Paul Bailey, Dan Sherman, and Mariann Lemke at AIR.

Contents

Introduction

New state and federal policies, such as the Race to the Top program, have redefined accountability to require that states develop and use performance-based teacher evaluation systems to inform decisions on professional development, employment, and compensation. These types of policies are an effort to increase classroom-level accountability by introducing new incentives for teachers to improve, providing support so that teachers can improve, examining alternative options with teachers who do not improve, and ultimately creating a higher quality pool of teachers who will in turn raise student achievement.

As one way to assess teacher effectiveness, many states and districts are using growth and value-added models. Developing and implementing growth and value-added models requires states and districts to grapple with the complexity of trying to assess individual teachers' contribution to student learning in a fair manner. In theory, linking student test scores to the teachers who taught them is straightforward. In practice, schools are dynamic environments where students may receive instruction from different teachers for varying amounts of time, and students and teachers themselves may move. In this environment, accurately capturing student-teacher relationships in administrative data and making determinations about how to appropriately account for teacher contributions to student learning can be challenging.

Some work already has been done around the nation to guide states and districts on developing business rules for defining teacher-student data links and the teacher of record.[1] This paper will contribute to the discussion by summarizing key issues that states and districts implementing growth and value-added models for teacher evaluation may wish to consider, along with practical examples of options for implementation. In essence, this paper seeks to offer different ways to address the question of "Which teachers are accountable for which students' performance?" This paper focuses on issues related to linking students to teachers (attribution) but can easily be extended to principals or other educators. We assume in this paper that student performance is measured by large-scale standardized assessments—often state accountability tests. However, many of the issues discussed here also may relate to student growth measures (other than statistical measures based on large-scale assessments like growth and value-added models), such as student learning objectives.

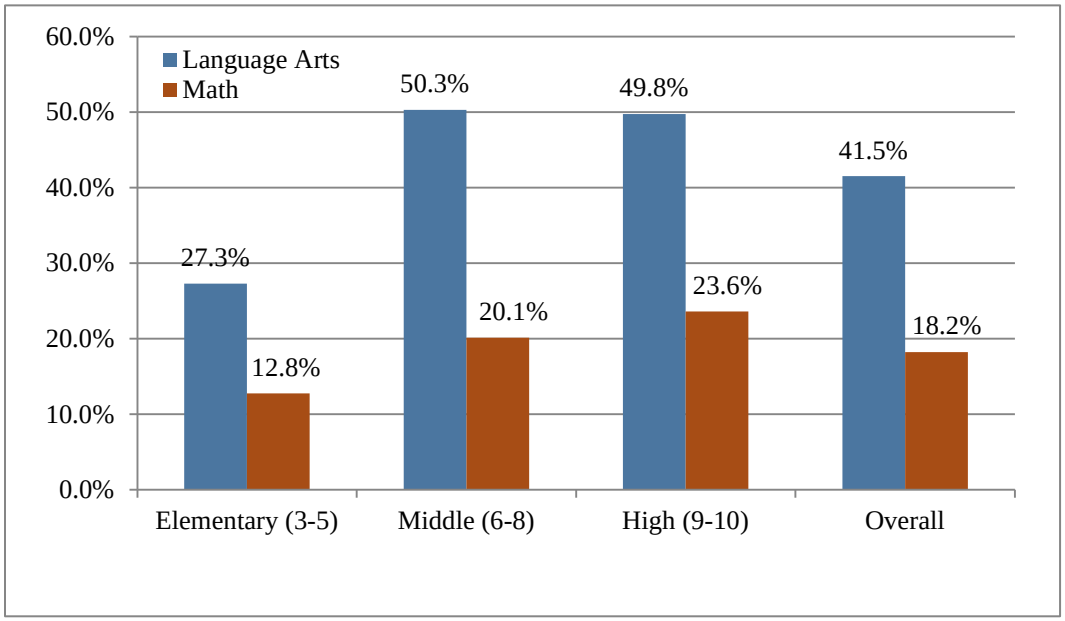
In the first sections, we illustrate some of the challenges of complex teacher-student relationships as well as the issues and options. Next, we describe data needed to attribute teachers to students and suggest potential rules for handling these data, once collected. In discussing these issues, we present analyses using data from a large state, which we were able to access for this paper. Finally, we mention related issues and offer a conclusion. Appendix A provides an overview of the data included in growth and value-added models to which attribution rules apply. Appendix B includes a decision framework for designing and implementing value-added models.

[1] See, for example, reports by the Data Quality Campaign (http://dataqualitycampaign.org/resources/details/993); resources related to the Center for Educational Leadership and Technology's pilot in five states (http://www.tsdl.org/PilotProjects.aspx) and, in particular, the *TOR (Teacher of Record) Framework* (http://www.tsdl.org/TORFramework.aspx); and select output from large applied research studies such as the Bill & Melinda Gates Foundation's Measures of Effective Teaching project, which more broadly examines issues on the fairness and reliability of evaluation systems (http://www.metproject.org/reports.php).

The Challenge

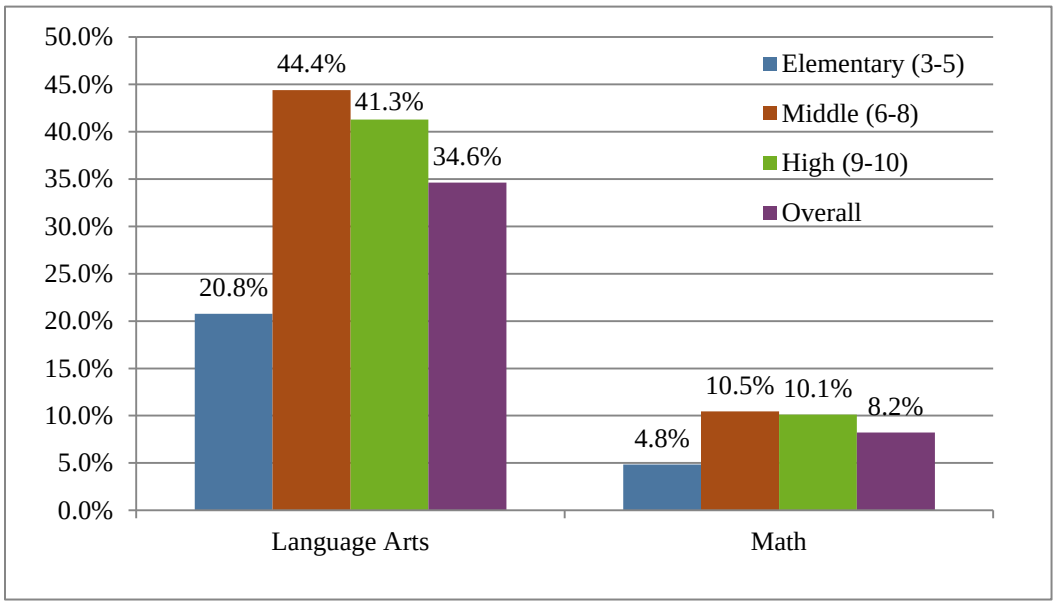
Although it may seem uncomplicated from a policy perspective to seek to hold teachers accountable for the learning of the students they teach, the reality of student assignment and instruction is complex. One important complication is that students may have more than one teacher in a given grade and subject during the year, making it more difficult to attribute learning gains among teachers. Using data from one state, we see that more than 40 percent of students are taught by more than one teacher in the school year in reading, and nearly 20 percent of students are taught by more than one teacher in mathematics (see Figure 1).

Figure 1. Percentage of Students Taught by More Than One Teacher Since the Start of the School Year, by Subject and Level



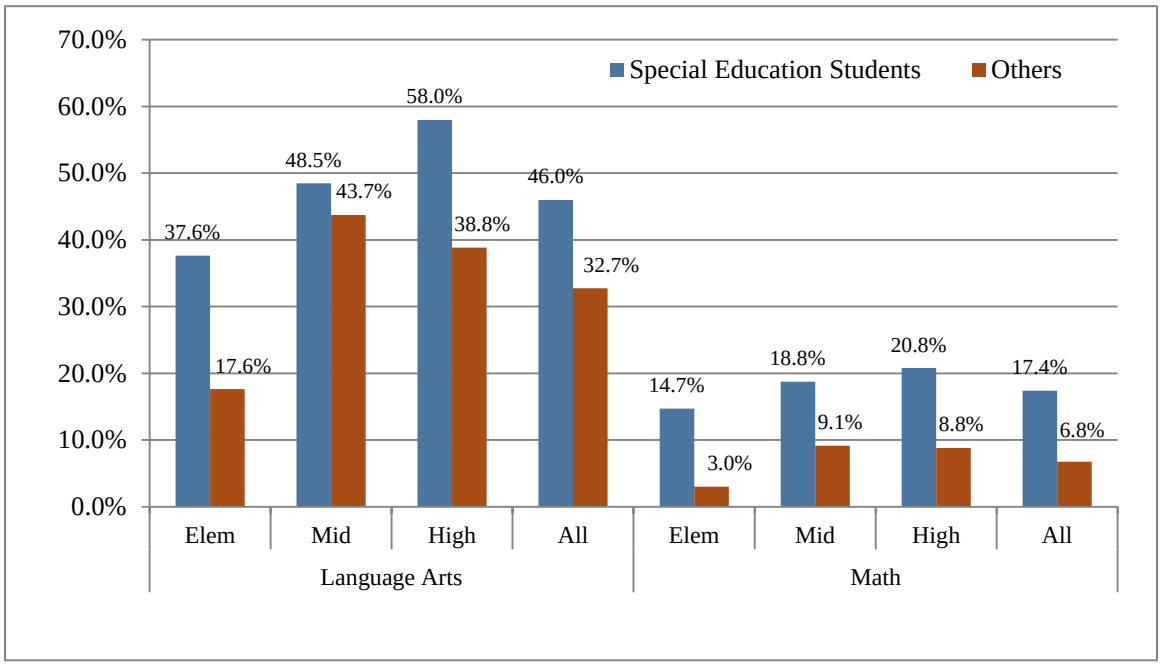
Especially at the higher grade levels, one might expect that students would be instructed by more teachers; in these grades, student schedules typically become more complex and students may switch teachers from one semester to the next. As illustrated in Figure 2, however, many students are receiving instruction from more than one teacher even within a semester, particularly in reading. These data suggest that students may be enrolled in multiple courses in a subject area, coteaching arrangements, or other situations, and they highlight the importance of accurately identifying courses relevant to assessment data and of capturing simultaneous teacher assignments.

Figure 2. Percentage of Students With More Than One Teacher in Spring, by Subject and Level



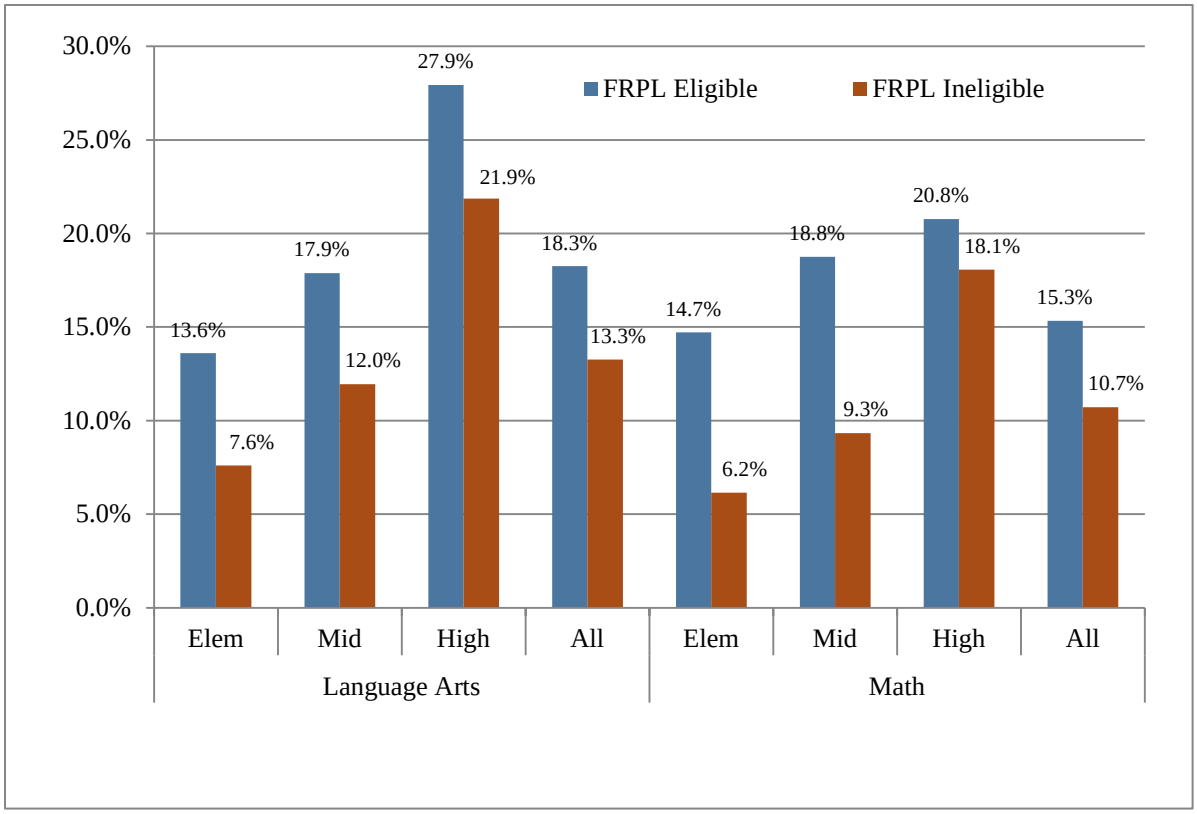
Students receiving special services (such as English language learners or special education students) may be particularly likely to have multiple teachers. Figure 3 shows that nearly 50 percent of students receiving special education services received instruction from more than one reading teacher and just under 20 percent from more than one mathematics teacher (compared with about 30 percent of non-special education students for reading and fewer than 10 percent for mathematics).

Figure 3. Percentage of Students With More Than One Teacher in Spring, by Subject, Level, and Special Education Status



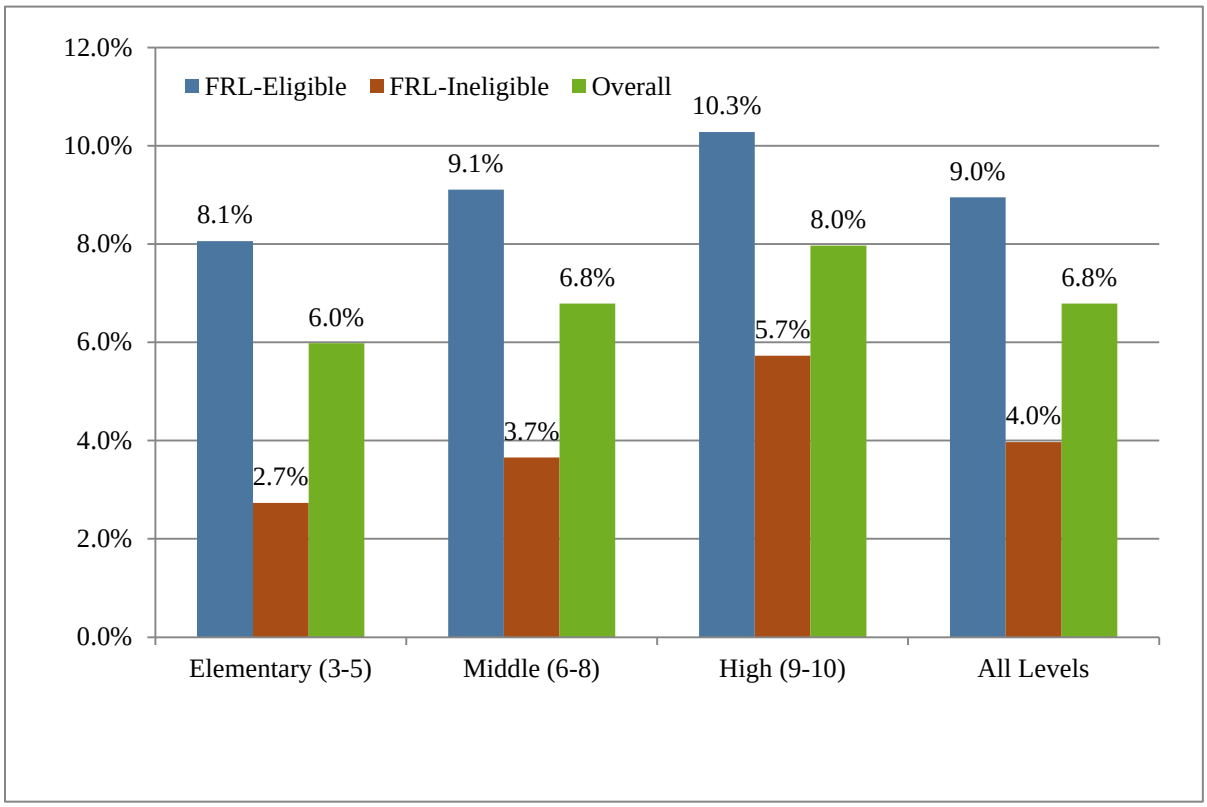
As another example, students eligible for free or reduced-price lunch (FRPL) are much more likely than the general student population to be instructed by multiple teachers in the course of a semester (see Figure 4) or the school year (see Figure 5).

Figure 4. Percentage of Students With at Least One Different Teacher in Spring Compared With Fall, by Subject, Level, and FRPL Eligibility



This fact may be due in part to higher student mobility among FRPL-eligible students, as shown in Figure 5, and points to the necessity of rules that handle the situations of student mobility.

Figure 5. Percentage of Students Who Changed Schools After the Start of the School Year, by Level and FRPL Eligibility (Reading and Mathematics Combined)



Hock and Isenberg (2012) similarly report on the complexity of student assignment, noting that in one district, "Of the teachers in the tested grades who had at least 15 math students, 21 percent taught students who were also educated in math by another teacher that year" (p. 1).

Issues and Options

"Teacher of Record" Guidance and Data—The Role of Teachers, Schools, Districts, and States

Many discussions of how to hold teachers accountable for student performance are framed around the issue of "teacher of record." The questions then are how does one determine who is a "teacher of record" for a student, and who is therefore responsible for student performance (and will receive a growth or value-added score).

Many states and districts already have adopted "teacher of record" policies, which reflect a general description such as that developed by the Center for Educational Leadership and Technology (2013): "A teacher of record is an 'educator' who is responsible for a 'specified proportion' of a student's 'learning activities' that are within a 'subject or course' and are aligned to 'performance measures.' "

From this general definition, states and districts must then determine how each of these terms is defined, what data will be collected to support them, and who will decide what is reported to meet the data requirements.

Some states or districts may provide guidance on how to determine if a particular teacher should be reported for a given student, while others may require that additional data about student-teacher-course relationships be reported to provide a more complete picture of the relationship. For example, the state of Michigan requires that districts report data about which teachers taught which students in which courses. The state provides guidance to districts about which students and teachers should be reported in their data but offers local discretion, noting the following about student and teacher data:

- "If the student was not enrolled for a period of time long enough to warrant the reporting of the course on the student's academic record, then that course does not need to be reported in TSDL [Teacher Student Data Link]. It is up to each district to determine when a course should be reported" (Center for Educational Performance and Information, 2011, p. 9).
- "The system allows up to three PICs [personnel identification codes] to be reported for each course in the student's record. All teachers who are responsible for assigning grades in the class should be reported. A teacher who is participating in the class in a support capacity but is not determining assignments or grades does not need to be reported. (For example, a special education teacher who comes into the class to provide extra support for students with IEPs [individualized education programs].) It is a local decision whether to report these additional support resources in the TSDL [Teacher Student Data Link] Collection. If they are reported, growth data will be returned accordingly" (Center for Educational Performance and Information, 2011, p. 11).

Districts may provide similar guidance to schools about what to capture in local student information systems.

In contrast, the state of New York endeavors to capture all teacher-student relationships, along with additional data elements that allow for more fine-grained determinations of "teacher of record" to be made. For instance, New York requires that districts report information related to instructional responsibility weight—that is, to what extent a teacher is responsible for a student's instruction. All teachers involved in instruction are reported, and an additional data element describing the extent of a teacher's responsibility is included. So rather than a yes/no determination about a teacher being responsible for a student's instruction, this additional data enables the state to determine for how much of a student's growth a teacher is accountable.

Typically data for determining "teacher of record" is gathered through state-level administrative data collections, in which districts report to the state. Districts, in turn, gather information from schools, often using local student information systems. One important question is how to ensure that the data reported by schools and districts are accurate and consistently reported across a given district or the state.

However, two important questions are how to ensure the quality of the data and who makes determinations about what is reported. In particular, since growth and value-added scores will ultimately be reported about teachers, states and districts may need to consider the role of teachers in ensuring that data reported about their instruction is accurate. One mechanism currently used in a number of states and districts is roster verification, in which teachers are able to review data about their students and verify the accuracy of the data. This type of activity can help build confidence in a value-added system because teachers know they will be held accountable only for students accurately linked to them.

Roster verification also can serve as a mechanism to collect information directly from teachers, rather than through a district-to-state data collection. Roster verification may be needed or useful in the short run if existing administrative systems do not capture needed data or capture such data with insufficient quality. Incorrect attribution can have serious consequences. Hock and Isenberg (2010) compared data from administrative systems to data that were collected through a roster verification process in a large urban district. They found that although teacher value-added scores did not vary much when using one or the other set of data *for teachers who are correctly placed in classrooms*, teacher rankings based on these scores varied significantly, with the rankings of one in 12 teachers changing by more than 10 percentile points. Further, not all teachers are correctly placed in classrooms according to administrative data. In this same study, Hock and Isenberg also found that about one in seven teachers was mismatched with entire classrooms of students in the administrative data because they did not teach mathematics, reading, or both to their students.

Whether or not roster verification is used to collect or review data, the next section explores potential data elements and how to use them in additional detail.

Data Needed to Link Students to Teachers

Linking Courses to Student Performance Measures

Although it may be useful to have a complete picture of teacher-student linkages in general, for the purposes of growth and value-added modeling, we need to understand which teachers taught students the content and skills that are relevant to the tests that students take. We also must know which courses students took that prepared them for these tests and which teachers taught them during those courses.

There are a number of different types of situations to consider:

- **Courses directly linked to the standards assessed by tests that students take.** In the elementary grades, for instance, students may be assigned to a course called "reading" or "mathematics."
- **Specialized courses with some relationship to the standards assessed by tests that students take.** For instance, students may be enrolled in a journalism course. If a student is tested in reading, to what extent are reading standards covered in a journalism course? Should a student's reading test score count toward a journalism teacher's growth or value-added score? These types of determinations about courses may require input from curriculum or instructional specialists. Another example is a mathematics resource course. A student in this course is likely to be receiving additional instruction in mathematics, which enriches her subject knowledge. When this student takes the mathematics test, should a value-added score be calculated for the resource course teacher if the bulk of the student's instruction was received from another teacher? In addition, how do we think about the fact that this student receives perhaps twice as much instruction as other students in mathematics—are there additional statistical controls that should be in place to account for these students and ensure that student growth is fairly evaluated?

State and district capacity to identify relevant course codes is likely to vary. Expanding subject areas or grade levels gradually will increase the number of teachers eligible for a growth or value-added score. From most straightforward to most complex—and perhaps, accordingly, from what can be done immediately versus what can be put in place in subsequent years—options include the following:

- **Including only directly relevant courses, such as reading and mathematics courses.** At a minimum, this option will require identifying reading and mathematics instruction within homerooms in elementary school.
- **Adding courses that are clearly closely aligned with reading and mathematics and are likely to include standards tested on the reading and mathematics standardized assessments.** These courses will vary across districts and states but potentially could include word study, oral reading, spelling, or algebra.
- **Adding high school courses for which the eighth-grade reading and mathematics standardized tests can serve as a pretest (e.g., creative writing, calculus).** These

courses also will differ from one district or state to the next, depending on standards taught in the course.

- **Engaging curriculum and instruction staff in a review of courses and standards taught (using, for example, syllabi and frameworks that connect courses and standards) to determine which courses can be associated with existing standardized tests.** This review may begin with courses that may be more likely to include teaching of the reading and mathematics standards (such as social science and science) and then continue with electives and on to all courses—including resource, support, and supplemental courses.

Notably, all districts in a state and even all schools in a district may not consistently identify or name similar courses. If it does not infringe on district or school autonomy, requiring consistent names and numbers for courses (many states use those course names and numbers identified by the National Center for Education Statistics, for example) or creating lists of courses with different names and/or numbers that are in fact the same course would help reduce the need for identifying reading and mathematics courses on a case-by-case basis.

The best processes will not prevent cases of missing course assignments. If there is no record of a student being associated with any course but test scores are reported for this student, the student can be excluded from analyses or the student's scores can count toward his or her school's growth or value-added score and not toward the growth or value-added score of any particular teacher. The latter option is unlikely to add much complexity to the process; that option, however, will ensure that more student scores are included in models and that fewer students are at risk of not counting at all.[2]

Linking Students to Courses and Teachers

After we have identified the courses that are relevant to a test that a student takes, we may want to identify how much time students spend in those courses with teachers. We may want to account differently in a value-added model for students who spent an entire school year with a teacher than for those who spent only a short time.

Policy priorities will guide the decision of whether to collect data that are not currently in administrative databases and how much new information should be collected. For example, priorities can be as follows: keeping the process simple; minimizing the burden on teachers to enter data and on principals and/or district officials to verify it; enhancing the accuracy and precision of what we know about teacher-student links to increase fairness in the process; reassuring teachers that the process is thoughtful and reflects the work they do; collecting more

[2] Although this paper focuses on the teacher level and much can be applied to the school (principal) level, it is useful to note two related issues at the school level here. First, there can be missing school assignments. If there is no record of a student being associated with any school but test scores are reported for this student, the student can be excluded from analyses or the student's test scores can count toward his or her district's (if reported) growth or value-added score and not toward the growth or value-added score of any particular school or teacher. This situation would apply only to systems that compute scores for districts. Second, students also may be assigned to multiple schools. A student may have a record of attending one school, but his or her test scores may be associated with another school; or a student may have records of attendance and testing associated with two schools. Toward which school's growth or value-added score will the student's test scores count?

information than may be needed to estimate models so that the additional information can be used to check the information included in models (what statisticians commonly refer to as face validity); or putting incentives in place (e.g., to maximize student performance on standardized tests, to focus on all students equally or on some students more, or to raise attendance).

These elements may include the following:

- **Enrollment.** How long—whether reported in minutes, days, or other metrics—were students enrolled in a teacher's course?
- **Attendance.** How often—again, whether reported in minutes, days, or other metrics—did students attend a course in which they were enrolled?
- **Multiple teacher assignments.** When multiple teachers are assigned to the same course with students, how should this situation be captured or accounted for in a growth or value-added model?

The next section provides possible rules for states to consider in using these data elements if collected. See Appendix B for a summary list of questions and issues to be addressed, as described in the rest of the paper.

Rules for States to Consider in Using Data Elements

Rules for Managing Enrollment and Attendance Data

Minimum Student Enrollment and/or Attendance Requirement

One approach to refining the links between teachers and students is to create a link only if a student was enrolled for a minimum amount of time in a teacher's course during the school year. In this case, a student is either in or out—either counts or does not count for a teacher's growth or value-added score.

If we are going to evaluate a teacher based on how much of an impact his or her instruction had on a student's academic growth, the teacher needs time to instruct the student. For example, consider a new student who enrolls in a class eight weeks into the school year. In this case, the teacher has less of an opportunity to impact the student's learning or contribute to the student's growth, compared with students who enrolled on the first day of school. Following are three examples of how enrollment or attendance data can be used to set a minimum requirement for students to be linked to—that is, "count" in a teacher's growth or value-added score.

- **One specific day of enrollment for inclusion.** One could measure enrollment by including any student present on a specified day, such as the student count day for auditing purposes or the first day of state assessment testing. If the student is enrolled in a teacher's course that day, then that student's growth counts toward that teacher's growth or value-added score. The data requirement is minimal, but this approach runs the risk of including in teacher scores the achievement of students who were only briefly in their classrooms.
- **Two specific days of enrollment for inclusion.** One could measure enrollment by including any student present on two specified days, such as the student count day *and* the first day of the state assessment testing. If the student is in the teacher's course on those two days, then that student's growth counts toward the teacher's growth or value-added score. Although this approach is an improvement over the previous measure in that there is more of a chance the student stayed in that classroom for the entire period, this approach still runs the risk of including in teacher scores the achievement of students who were only briefly in their classroom, as students could have moved in and out of the school or class between those two dates.
- **Minimum number of days of enrollment or minimum enrollment or attendance proportion of a course for inclusion.** Alternatively, one could measure enrollment by choosing a minimum number of days that a student was enrolled in a course with a teacher. This measure could require that a student be enrolled for a minimum number of days or continually between two dates, for example, between the student count day *and* the first day of the state assessment testing. Or one could set a minimum enrollment duration—for instance, that a student must have been enrolled for at least 85 percent of a course's duration in order to count for a teacher. One could go even further and distinguish days of continuous enrollment versus all days of enrollment. A similar type of rule can be established for attendance. For example, if a student meets a certain enrollment threshold, a similar attendance threshold also could be applied.

All of this information is generally available in administrative data. More detailed information may be available or systems may be designed to collect it, such as an exact record of days enrolled. The more refined the information, the more likely it will be for teachers to agree that it is fair. Myriad options are available. Yet collecting more refined information also is more complex and time-consuming, thus at risk for error. Also, the most refined information may in the end not make a significant difference in model result—in the sense that it may not affect teacher growth, value-added scores, or teacher rankings based on those scores. One consideration for policymakers is the numbers of students and teachers for whom data will be reported based on different business rules. Stringent rules may exclude many students, such that few teachers are accountable for their performance, while more lenient rules may result in teachers perceiving that they are held accountable for students for whom they had no opportunity to influence learning. Once again, how detailed to go is a policy decision based on balancing fairness and simplicity. Estimating models with various options may provide empirical guidance for making these types of decisions.

Weighting Student Enrollment or Attendance

Rather than a simple yes/no decision for including students in a teacher's growth or value-added score, an alternative approach is to weight students in a score based on how much time they spent with a teacher—a "dosage" or sliding scale type of approach. This measure could apply both to enrollment and attendance, with students being weighted based on one or the other. Alternately, minimum requirements could be set for enrollment, with a weighting based on attendance.

Statistical Controls Versus Linking

The rationale for considering attendance criteria in generating teacher growth and value-added scores is to ensure that teachers are fairly held accountable for students. An alternative to using attendance data in linking students to teachers would be to include it as a statistical control in a growth or value-added model. By including attendance rates in a growth or value-added model, a teacher's score will be adjusted for a student who tends to have poor attendance. Value-added modeling experts typically advise policymakers to use prior year attendance because student attendance rates tend to be steady from one year to next. Using current or outcome-year attendance is statistically undesirable because it may be correlated or confounded with the outcomes that we are trying to measure.

Teacher Attendance

Teacher attendance also may need to be considered. If a teacher is absent for a significant portion of a school year, such as for a medical leave, but this absence is not represented by a break in the teacher's assignment to a course, the teacher does not have an opportunity to contribute to the learning of students in the classroom that was assigned to him or her. A state or district will want to determine what the minimum number of days a teacher should be present to receive a growth or value-added score. Beyond that, the teacher's contribution to student learning and how student test scores count toward the teacher's growth or value-added score will be based on how much instruction students receive from that teacher during the rest of the school year using the rules discussed next (see the Rules for Multiple Teacher Assignments subsection). For example, a

teacher who was absent long-term may count as a coteacher or support teacher and still be associated with his or her students.

When a student is taught by a substitute teacher while the regular teacher is absent, an important factor is whether the substitute teacher is evaluated in the same manner as the regular teacher. Short-term substitutes generally are excluded from standard teacher evaluation processes, so the students' scores from the course are not likely to be used for a short-term substitute teacher's evaluation. If long-term substitutes are included in evaluations, which may be the case particularly when long-term substitutes are certified and/or licensed teachers, then the same rule for inclusion in a regular teacher's growth or value-added score is likely to apply. Alternatively, a state or district may decide that student scores will contribute only to the growth or value-added score of the teacher who taught most of the course, possibly with a weight attached to the student scores (see the following section on Rules for Multiple Teacher Assignments). What constitutes long- or short-term will be up to the state or district, which may have rules in place in that regard or may choose to create a definition for this purpose—perhaps based on a minimum number of days, which allows for significant contribution to students' instruction.

Rules for Multiple Teacher Assignments

The data needed and approaches for handling such data can become even more complex when multiple teachers are linked to a student in a course—as in coteaching, resource, or support teacher situations.

One potential approach to handle these situations is to use weights that represent a teacher's relative responsibility for teaching students in a course. These weights can be used to reduce the contribution of student scores to a growth or value-added score for a teacher who has less of an opportunity to contribute to those students' learning. This approach may be especially useful in situations where resource teachers may target their instruction to particular students in a class rather than to a whole class.

For example, if we define a teacher's growth score as the mean of student growth scores, and a teacher teaches reading to 20 students, but 10 of those students are pulled out to receive instruction from an English language learner (ELL) teacher 20 percent of the time, we could create a weighted mean where the teacher's score is the sum of the growth scores of the 10 students receiving ELL services multiplied by 0.80 plus the sum of the growth scores of the 10 other students, all divided by 20. As such, the students who the teacher does not teach full-time do not count as much in his or her growth score.

These situations can become even more complex at the secondary level, when assessments may not be taken concurrently with relevant courses (e.g., in cases where test retaking is allowed) or when a variety of courses may be considered relevant to a test (e.g., English 1 and 2 may be relevant to an end-of-course English language arts test).

Data availability will in part determine whether coteaching experiences can be taken into account. Incentives are an additional important point to consider when examining these situations. For example, if the students who a teacher teaches count more toward her score than

the students she coteaches, she may pay more attention to the students she teaches without a coteacher.

Collecting these data and answering these questions will help policymakers devise attribution schemes that reflect teaching and learning in schools and are fair to teachers—by, for instance, excluding students or teachers from the student-teacher data links or by assigning weights to reflect smaller or larger contributions.

Additional Approaches to Attribution

There are a number of ways to measure refined student-teacher data linkages, depending on what information is available or can be collected through existing or new data systems. For example:

- Links can be measured as minutes (e.g., length of a course).
- Links can be measured as percentages:
  - "In a typical week, I teach this student 100 percent of the time, 75 percent, and so on."
  - "In a typical week, I contribute 100 percent, 75 percent, and so on of the reading instruction this student receives."
- In addition, each of these measures can be more or less detailed:
  - Deciles
  - 0–25 percent, 26–50 percent, 51–75 percent, 76–100 percent
  - 0, 100, other values

Rather than presenting teachers with numbers, the district could ask teachers to assign a percentage of time to a given student based on simple scenarios. For example, a teacher could choose between five scenarios for each student:

- "I was the only teacher for this student in this course."
- "I provided most of the instruction to this student in this course, but the student did receive additional support from another teacher."
- "Another teacher and I jointly taught this student in this course."
- "I provided support to this student in this course."
- "I did not teach this student in this course."

Note that each of these options can be phrased more or less precisely to help a teacher understand and capture what happens in schools in a given district or state. For example, the first statement could be rewritten as follows: "I was the only teacher for this student in this subject," "I was the only teacher for this student in this course for most of the school year," or "I was the only teacher for this student in this course in a typical week."

These statements could then be automatically associated with percentages in the data—say 100 percent, 75 percent, 50 percent, 25 percent, 0 percent, respectively. Or to keep it simple or if

the data warrant, the three middle categories could all be associated with 50 percent to represent partial contribution.

The approach can be further refined to account for multiple course assignments. Students may be enrolled in the same relevant course in multiple time periods with the same or different teachers. They also can be enrolled in multiple courses in the same subject. How will students be attributed to those courses or teachers? Some options include the following:

- A student can be associated with each of those courses or teachers, and the student's score counts once toward the teacher's growth or value-added score with a proportion of 100 percent.
- The student can count multiple times based on the number of periods.
- The contribution of the student's score can be weighted as follows:
    - By the number of periods
    - By the proportion of time spent with each teacher prior to testing, which can be measured as follows:
        - By the number of minutes each period lasts
        - By the number of days the course meets every week
        - By the number of days the student is enrolled in each period
        - By the number of days the student attends in each period
        - As described by linkage, course duration, and adjustment to linkage duration data

Some of these questions also will be relevant in situations where a student enrolls in a course, drops it, and then re-enrolls—for example, if the student switched schools and then came back to her original school.

Whichever approach is chosen, there is a risk of some students not counting toward the growth or value-added score of any teacher. Highly mobile students may not be enrolled in any teacher's class long enough to be included in calculations. By using more refined attribution rules as described earlier, there is greater likelihood that students will count even for a small proportion. It also may be possible for students to count toward the growth or value-added score of a school even though they may not count toward the score of a particular teacher.

Experimenting with these methods first is ideal, such as by conducting a pilot study, hosting focus groups with teachers, and/or implementing growth or value-added models with different numbers and options. Policymakers could put different sets of percentages of linkage to students in front of teachers and get their reactions; what makes them feel their work is accurately described? Policymakers could estimate models with 0-25-50-75-100 percent and 0-50-100 percent breakdowns of student-teacher linkage and see if and how teachers' scores change. Policymakers could estimate teacher scores using two days of required enrollment and with a continuous enrollment number of days, then a continuous attendance number of days. Are there differences in these results? The data will help state and district officials make decisions and document, justify, and defend them.

Related Issues

Although the workings of growth and value-added models are not the subject of this paper, there is one important underlying concept that is relevant to the topic of attribution: the number of students that can be associated with a teacher or school. Larger numbers of students make for more reliable value-added scores, and value-added scores will become more a precise measure of a teacher's value added score as they are associated with more students. A state or district should then consider what the minimum number of students will be, below which growth or value-added scores will not be reported for a teacher or school and used for the purpose of evaluation—in other words, a number that will make teachers and other users of the data feel comfortable that enough students contribute to the growth or value-added score for the score to be a meaningful representation of a teacher's effectiveness.

At the same time, setting the minimum number of required students too high may prevent many teachers and possibly even some schools from receiving the additional measure of effectiveness that a value-added score provides. Also, this approach will likely affect certain types of teachers more than others—for example, teachers who typically teach in smaller class-size settings, such as is often the case for teachers of special education students or of ELLs.[3] In addition, it will disproportionately affect teachers in schools with high student mobility, which often are among the more disadvantaged schools. In turn, students of these teachers in these settings may be affected if this system provides an incentive for teachers to focus more on students in classrooms or courses where they are more likely to count toward their evaluation.

There is no theoretical or standard minimum number of students required or even recommended for a teacher to receive a growth or value-added score. Rather the following two options, at a minimum, should be considered:

- **Option 1.** Examine data and experiment with different thresholds, such the following:
  - How many teachers are excluded with different minimum numbers of students?
  - Who are these teachers? Where are these teachers located? (Are the teachers of students in special education much more likely to be excluded?)
  - How does this situation affect educator ratings? (Are there many more teachers with the lowest growth or value-added scores when a particular minimum sample size is used?)
  - How do the answers to these questions differ when examined in the context of different attribution rules (e.g., depending on the minimum number of days set, as discussed earlier)?
- **Option 2.** Take statistical precision into account in using the data. Instead of setting a minimum number of students, report standard errors, which will give users of the data a sense of the range in which their score falls. (The smaller the number of students associated with the teacher, the larger the range of possible scores will be.)

[3] Note that growth and value-added models can measure growth only for students taking standardized tests or tests with the minimum properties of measuring performance over time, which excludes a number of students in special education. Issues that arise in the context of nontested grades and subjects are beyond the scope of this paper.

Conclusion

The goal of this paper is to contribute to the growth and value-added efforts of state and district leaders and to provide guidance on designing attribution rules based on assembling the right data and asking the right questions. Because using growth and value-added models to measure teacher effectiveness is a relatively new endeavor, stakeholders may have many legitimate questions and concerns. For instance, administrators may question how best to evaluate their teachers, or teachers may be concerned because they recognize the potential implications of these measures—both the risks and the rewards. Considering these types of questions will result in a model that best reflects teachers' contributions to student learning and informs a fair evaluation system that will reward teachers and provide them with the support that they need to provide the nation's children with opportunities to thrive academically.

References

Center for Educational Leadership and Technology. (2013). *Teacher of record (TOR)* [Webpage]. Retrieved from http://www.tsdl.org/teacherofrecord.aspx

Center for Educational Performance and Information. (2011). *Michigan student data system: Teacher student data system: Frequently asked question* (Version 1.2). Lansing, MI: Author. Retrieved from http://www.michigan.gov/documents/cepi/TSDL_FAQ_351862_7.pdf

Hock, H., & Isenberg, E. (2010, March). *Teacher and student misclassification in value-added models.* Paper presented at American Education Finance Association's annual conference, Richmond, VA.

Hock, H., & Isenberg, E. (2012). *Methods for accounting for co-teaching in value-added models* (Working Paper). Washington, DC: Mathematica Policy Research. Retrieved from http://www.mathematica-mpr.com/publications/PDFs/education/acctco-teaching_wp.pdf

Appendix A. Overview of Growth and Value-Added Model Data

This appendix provides an overview of the data included in growth and value-added models to which attribution rules apply. At a minimum, growth and value-added models require student achievement data. They also call for data that allow for the linkage of students to teachers—at a minimum, data or a business rule that links at least one teacher with each student. Finally, some types of models can take into account additional variables at the student or classroom level. Figure A1 illustrates required and optional data for growth and value-added models.

Figure A1. Examples of Required and Optional Data for Growth and Value-Added Models

Student Achievement Data	Additional Student- or Classroom-Level Data	Student to Teacher to Course Linkage Data
▪ Current year's achievement data in reading and mathematics ▪ Prior year's achievement data in reading and mathematics ▪ *Optional:* Achievement data in other subjects ▪ *Optional:* Additional years of prior achievement data	▪ *Optional student-level:* Indicators of poverty, ELL status, disability status, immigrant status, homeless status, gifted status, race/ethnicity, age, gender, grade repetition, attendance, suspensions ▪ *Optional classroom level:* Percentages of those characteristics within a classroom	▪ Student course schedules ▪ Teacher course assignments, including period, section, or other relevant information ▪ District course catalogs, including course names and identification numbers, for all reading and mathematics courses ▪ *Optional*: Additional information such as weights collected through roster verification

Student Achievement Data

The first basic requirement for any growth or value-added model is the availability of assessment data over time in order to measure the change in student performance over time—thus, achievement data for at least two points in time are needed. Generally, these assessment data are from standardized tests administered throughout a set of grades, often Grades 3–8, and available in reading and mathematics. Some states also have created tests in more subjects and grades, and some states have created pretests for existing assessments or are exploring ways to use existing assessments that do not have a sequence across years and grades.[4]

Additional Student- and Classroom-Level Data

Some models can take into account characteristics of students or of the classroom that may have an effect on a student's growth trajectory. States and districts may select to include student- and/or classroom-level variables that may affect student outcomes while being out of the control of the teacher. Student-level variables generally considered for inclusion in models include indicators of poverty, ELL status, disability status, immigrant status, homeless status, gifted

[4] For example, Georgia; Hillsborough County, Florida; and New York state.

status, race/ethnicity, age, gender, grade repetition, attendance, or suspensions. Classroom-level variables typically are created as a percentage of students in the classroom with a given characteristic—for example, the percentage of students in the classroom identified as ELLs. These variables take into account peer effects in the classroom and how these peer effects may affect student growth.

The general argument against including student-level characteristics in the estimation is that it places different expectations on those students who have been reported to have a particular characteristic. The general argument for including student-level characteristics is that students with these characteristics may pose additional teaching challenges or be easier to instruct, and a teacher should not be penalized or rewarded based on the composition of his or her classroom(s). For example, a teacher may serve students with an attention deficit disorder, who may show learning growth at a slower rate than other students, or who may struggle more than others in test-taking situations. By including an indicator in the model for special education services, these students may be eligible to give that teacher a similar chance to receive a high value-added score as another teacher. Finally, states may have legal requirements to include or exclude certain variables.

Appendix B. Quick Guide for Developing Growth or Value-Added Models

The following questions and checklists include the key points discussed in this report for developing growth or value-added models.

1. **What data do you need to correctly attribute students to teachers?**

   Required:

   _____ Current year's achievement data

   _____ Prior year's achievement data

   _____Unique student and teacher IDs

   _____ Student course schedules

   _____ Teacher course assignments, including period, section, or other relevant information

   _____ District course catalogs, including course names, and identification numbers

   Optional:

   _____ Achievement data in other subjects

   _____ Additional years of prior achievement data

   _____ Student-level data (e.g., indicators of poverty, ELL status, special education status)

   _____ Classroom-level data (e.g., percentages within a classroom of student-level data)

   _____ Additional information such as weights collected through roster verification

2. **Is roster verification or additional data collection needed?**

   Consider the following questions to determine if roster verification or additional data collection may be warranted:

   - What data are readily available in existing systems? How much do we already know about who teaches what to whom and for how long (e.g., courses that students take and with which teacher(s), subject/standards taught in the course, teacher rosters, frequency and dates of assignment updates)?
   - How reliable are these data? Do districts certify them before reporting them to the state? Do teachers certify them before the district reports them to the state? Do coteachers certify one another's entries?

- What data are missing? For example, roster verification can give teachers the opportunity to flag whether they are the sole teacher for a course or define the coteaching situation in which they are. Often, administrative data (e.g., homeroom label) do not specify whether elementary school teachers teach reading, mathematics, or both, for example.

3. What is the minimum number of students below which growth or value-added scores should not be reported for a teacher and used as part of their evaluation?

To determine the minimum number of students required, experiment with different thresholds of students and answer the following:

- How many teachers are excluded with different minimum numbers of students?
- Who/what are they (e.g., are teachers of students in special education much more likely to be excluded)?
- How does this situation affect ratings (e.g., are there many more teachers with the lowest growth or value-added scores when a particular minimum sample size is used)?

Alternatively, instead of setting a minimum number of students, standard errors can be reported that will give users of the data a sense of the range in which their score falls. (The smaller the number of students associated with the teacher, the larger the range of possible scores will be.)

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202.403.5000

www.air.org

Value-Added Measurement Services, part of AIR's Educator Talent Management services and resources, can assist states and districts in finding the right approach to measuring how educators contribute to growth in student achievement.

1979_08/13