Incorporating End-of-Course Exam Timing into Educational Performance Evaluations

Eric Parsons
Cory Koedel
Michael Podgursky
Mark Ehlert
P. Brett Xiang*

November 2013

There is increased policy interest in extending the test-based evaluation framework in K-12 education to include student achievement in high school. High school achievement is typically measured by performance on end-of-course exams (EOCs), which test course-specific standards in subjects including algebra, biology, English, geometry, and history, among others. However, unlike standardized tests in the early grades, students take EOCs at different points in their schooling careers. The timing of the test is a choice variable presumably determined by input from administrators, students and parents. Recent research indicates that school and district policies that determine when students take particular courses can have important consequences for achievement and subsequent outcomes, such as advanced course taking. The contribution of the present study is to develop an approach for modeling EOC test performance that disentangles the influence of school and district policies regarding the timing of course taking from other factors. After separating out the timing issue, better measures of the quality of instruction provided by districts, schools and teachers can be obtained. Our approach also offers diagnostic value because it explicitly separates out the influence of school and district course-taking policies from other factors that determine student achievement.

* The authors are from the Department of Economics at the University of Missouri. They acknowledge research support from the Institute of Education Sciences through the Center for Analysis of Longitudinal Data in Education Research (CALDER) and a collaborative relationship with the Missouri Department of Elementary and Secondary Education. The usual disclaimers apply.

1. Introduction

Direct performance measures based on student test scores are being increasingly incorporated into educational evaluations at the district, school and teacher levels. The large and well-documented variation in effectiveness across educational units (Betts, 1995; Chetty et al., 2011; Hanushek and Rivkin, 2010; Rockoff, 2004), coupled with the inability of researchers to consistently link performance differences between units to readily-observable characteristics (Betts 1995; Kane et al., 2008; Nye et al, 2004; Rivkin et al., 2005), motivates the use of these measures in the evaluation process.

The research literature upon which the development and use of performance measures in education is based is predominantly comprised of studies that measure student achievement on standardized exams administered in the early grades – in particular, in math and reading in grades 3-8. However, educational administrators looking to broadly incorporate performance measures into the evaluation process do not have the luxury of restricting their attention to the grades and subjects for which there is universal standardized testing. A logical first step in expanding the scope of evaluation beyond the traditional standardized-testing window is to incorporate high school subjects for which end-of-course exams (EOCs) are already administered. EOCs are currently administered in a variety of subjects in most states. In Missouri, for example, there are EOCs for courses such as algebra-I, algebra-II, American history, biology, English-I, English-II, geometry, and government.

A key challenge in moving from grades and subjects with (near) universal testing to EOCs is that the point in the schooling process at which EOCs are administered is a choice variable. The timing of the test depends on decisions by parents, students and district and school administrators. The fact that the timing of EOCs is subject to some discretion introduces standard concerns about endogeneity. From a policy perspective, the stakes are high. Recent research shows that school and district policies regarding the timing of course taking meaningfully affect student achievement and longer-term outcomes (Clotfelter et al., 2012a, 2012b).[1]

The contribution of the present study is to develop a procedure by which educational administrators can identify and separate out the effects of the timing of course taking in EOC evaluations. This separation achieves two objectives. First, it facilitates direct rewards/sanctions for schools and districts that set up effective/ineffective course-taking policies. Second, it allows administrators to better identify differences in instructional effectiveness as they relate to EOC performance by removing the influence of course-timing effects.[2]

We develop a three-part approach to incorporate EOC performance into educational evaluations, focusing initially on school districts as the units of analysis. First, we estimate value-added models separately by grade level to measure cross-district differences in instructional effectiveness conditional on the grade level in which the EOC is administered. A benefit of estimating the models separately by grade level is that they hold the timing of the test constant so as not to confound timing issues with other aspects of instructional effectiveness.

The initial by-grade-level models would be sufficient for evaluating district performance, subject to standard concerns regarding model specification (which we discuss in more detail below), if exam timing were unimportant. However, given that exam timing is important, the initial models are omitting important information. To give a concrete example, consider a district that is highly effective in instructional practice but has implemented suboptimal course-timing policies. Per Clotfelter et al. (2012a, 2012b) and the evidence we present below, one such suboptimal policy

[1] In practice, districts need not bundle test taking with course taking – for example, students could take algebra-I in grade-9 and then take the algebra-I EOC in grade-11. Our analysis assumes course taking and test taking occur concurrently, which is what we expect to be the most common circumstance. Of course, policies could be enacted to force the bundling of course- and test-taking for EOCs.

[2] Here, "instructional effectiveness" is a catch-all phrase meant to cover a wide variety of factors that may affect student learning. Obviously, teacher effectiveness is one part of this measure, but it may also include other non-teacher related factors such as curriculum choice.

would be to make grade-8 the modal grade in which students take algebra-I. A performance evaluation based only on output from the initial value-added models might indicate that this district is highly effective. However, when one accounts for the fact that a large fraction of students take algebra-I in a suboptimal grade, it may be underperforming.

We build on the initial models to take explicit account of the effects of course timing. Specifically, we use an instrumental variables (IV) strategy to isolate gaps in student achievement across districts that are plausibly attributable to differences in policies regarding the timing of course taking. We then use the IV estimates to adjust the initial performance measures by penalizing districts for students who take EOCs at the wrong times.

Finally, we allow district and school personnel (and students and parents) some flexibility in terms of deciding when students take courses by making *ad hoc* corrections to the course-timing adjustments. In short, these corrections allow for a fraction of students to take specific courses off of the path that *most* students should follow. The corrections that we apply are based on available evidence (Clotfelter et al., 2012b) but subject to simple modifications depending on policymaker circumstances and preferences.

To illustrate our approach, we use it to inform a hypothetical district-level evaluation system for algebra-I EOC performance in Missouri. We show that a small number of Missouri districts would be meaningfully misplaced in overall performance ratings if those ratings depended on grade-level-specific value-added measures alone. A significant number of students at these affected districts are taking the algebra-I EOC at suboptimal times. We also discuss how our approach can be generalized to accommodate other EOCs and other levels of evaluation – e.g., for schools and/or teachers. Accounting for course-timing effects will be important for educational evaluations of EOC performance at all levels.

2. Data

The data for this study are taken from the Missouri Department of Elementary and Secondary Education's (DESE) statewide longitudinal data system. The system includes all students who attend a public elementary or secondary school in the state of Missouri and, by virtue of a unique student identifier, allows for student records to be linked over time and across schools within the state from 2006 onward. In addition to student enrollment data, the system also contains assessment data for all EOC and Missouri Assessment Program (MAP) exams (MAP is the statewide standardized exam that is administered in grades 3 through 8). Detailed course assignment data are available for all students from 2008-09 forward.

EOCs were first administered in Missouri at the end of the 2008-09 school year. Three exams were given in that first year (algebra-I, English-II, and biology). The number of EOCs administered in the state has since grown to eight (as of 2011-12) with the addition of algebra-II, American history, English-I, geometry and government. We use algebra-I scores as outcomes for this paper because (1) they allow for direct comparisons to previous research on the timing of course taking in higher grades (Clotfelter et al., 2012a, 2012b), and (2) algebra-I is the most commonly administered EOC in Missouri. The outcome measures are taken from the 2011-12 school year to allow for a full complement of past exam scores to be used as controls in the empirical models. Summary statistics for the analytic sample are presented in Table 1.

Table 2 shows the grade-level distribution of algebra-I EOCs in Missouri. The algebra-I EOC grade-level distribution is quite dispersed, with sizeable numbers of students taking the exam in each grade from grade-8 to grade-12. Table 2 also shows that some students may take an EOC more than once over the course of their school careers. Specifically, the first panel of the table presents the grade-level distribution for all students who take the algebra-I EOC in 2012, while the second panel is limited to students who take the exam for the first time. Comparing the two panels, it does not appear that re-takers meaningfully shift the algebra-I EOC grade-level distribution However, as one would expect, the distribution of students who retake the exam is heavily weighted towards the upper grades. In fact, 6.3 percent of grade-10 students, 22.3 percent of grade-11 students, and 11.7 percent of grade-12 students who took the algebra-I EOC in 2012 were not first-time test takers.

3. Empirical Strategy

3.1 Measuring Instructional Effectiveness Conditional on Course Timing

We begin by estimating a two-step value-added model following Ehlert et al. (2013) to produce "instructional effectiveness" measures for districts based on the algebra-I EOC.[3] The model is specified as follows, with equation (1) estimated separately by grade level:

$$Y_{idgt} = \beta_{0g} + Y_{ig(t-k)}\beta_{1g} + M_{ig(t-k)}\beta_{2g} + X_{idgt}\beta_{3g} + D_{dt}\beta_{4g} + \epsilon_{idgt} \qquad (1)$$

$$\epsilon_{idgt} = I_{idgt}\theta + \eta_{idgt} \qquad (2)$$

In equation (1), Y_{idgt} is the EOC score of student i in district d and grade g who took the test at time t. $Y_{ig(t-k)}$ is a vector of lagged MAP scores for the student (the three most recently available years of MAP examination scores in both mathematics and communication arts) where k can take on different values for students who take the algebra-I EOC in different grades. $M_{ig(t-k)}$ is a vector of dummy variables controlling for missing exam scores, X_{idgt} is a vector of student-level control variables that includes indicators for gender, race, whether the student has an individualized education plan (IEP), free/reduced price lunch (F/RL) status, English-language learner (ELL) status, exam retaking status, and student mobility, D_{dt} is a vector of district-level aggregates of the

---

[3] The exact specification for the student-achievement model is not critical to the overall approach; e.g., district fixed effects could be included directly in equation (1) if desired. Changes to the structure of the initial student-achievement model would require minor operational adjustments to subsequent steps in the process. One advantage of the two-step model as described in equations (1) and (2) is that it produces "proportional" district rankings (see Ehlert et al., 2013).

variables included in the other three control vectors, and ϵ_{idgt} is the error term.[4] By virtue of the grade-level estimation, the coefficients in equation (1) can differ across grades (g) as indicated in the equation.[5] In equation (2), I_{idgt} is a vector of indicator variables where the indicator for the district in which student i took the EOC is set to one and all other indicators are set to zero. θ is the vector of district performance measures.

Equation (1) predicts each student's EOC score based on a wide array of information about both the student and the district in which the student takes the exam. The vector of residuals taken from equation (1), ϵ, represents how well each student performs compared to her predicted score. A positive residual indicates that the student out-performed the prediction, while a negative residual indicates that the student scored below the predicted value. The estimated residuals are then used in equation (2) to produce estimates of θ.[6] A positive value for θ indicates that the average student in the district out-performed her prediction while a negative value indicates the opposite.

An important distinction between equation (1) and the first step of the value-added model presented in Ehlert et al. (2013) is that equation (1) is estimated separately for each grade level.[7] Because of this modeling structure, students are initially compared only to other students in the

[4] All MAP exam scores are standardized by year-grade-subject cell. The outcome variable (the EOC score) is also standardized to have mean zero and standard deviation of one, although its standardization is not performed separately by grade level in order to preserve cross-grade-level performance gaps in the outcome measure. For a discussion of the vector of missing score dummy variables ($M_{ig(t-k)}$), see Appendix A.

[5] The by-grade-level estimation is useful because it allows for heterogeneity in the predictive power of available covariates for students who take EOCs in different grades. As a specific example, if the model uses standardized math scores in grades 6, 7, and 8 to predict the EOC score in algebra-I, the predictive power of these prior scores is allowed to vary depending on whether students take the EOC in grade-9 or grade-10. Of course, the by-grade-level models produce results that are identical to a model that allows for all predictors to be interacted with all grade-level indicators. Perhaps the true value of the grade-level separation along this dimension, from a policy perspective, is transparency for educational administrators and interested parties (e.g., teacher organizations).

[6] Equation (2) is estimated without an intercept so that effect estimates and standard errors are calculated for every district. The effect estimates are simply the average of the residuals assigned to the given district, while the standard errors are calculated to be robust to the presence of heteroskedasticity. Shrinkage is also applied to the effect estimates via the method used in Koedel et al. (2012).

[7] Students who took the algebra-I EOC before grade-7 were excluded from the model. These students represent a very small fraction of the overall sample (<0.2%).

same grade cohort. As a result, equation (2) provides measures of how well districts are educating their students conditional on the grade in which students take the course. The modeling structure so far does not consider whether districts are placing students into the course at the right time. It is to this issue that we now turn.

3.2 Accounting for the Effects of Course Timing

Clotfelter et al. (2012a, 2013b) show that district policies regarding the grade-level placement of students into algebra-I can significantly affect exam performance and longer-term outcomes such as future course taking. Specifically, Clotfelter et al. (2012a) study an abrupt change in the algebra-I course-timing policy in the Charlotte-Mecklenburg school district and find that moderately-performing students who were accelerated into algebra-I in grade-8 score nearly a third of a standard deviation lower on the EOC than similar students who were not accelerated (and took the exam in grade-9). In a subsequent study, Clotfelter et al. (2012b) expand on their initial analysis in Charlotte-Mecklenburg to look at the 10 largest districts in North Carolina and find similar negative test-score effects of accelerated algebra. These studies point to the importance of directly accounting for course-timing effects in EOC evaluations.

Identifying the effects of course timing on test scores is challenging because the grade in which students take algebra-I is endogenous. Clotfelter et al. (2012a) provide evidence that the endogeneity of course timing is problematic and can yield misleading results if left unaccounted for. To deal with the endogeneity problem and identify the effects of course timing on student achievement, we estimate the following instrumental variables model for first-time test takers:

$$G_{idgt} = \gamma_{0g} + \tilde{Y}_{ig(t-k)}\gamma_{1g} + \widetilde{M}_{ig(t-k)}\gamma_{2g} + X_{idgt}\gamma_{3g} + \widetilde{D}_{dt}\gamma_{4g} + P_{dt}\gamma_{5g} + e_{idgt} \qquad (3)$$

$$Y_{idgt} = \delta_0 + \tilde{Y}_{ig(t-k)}\delta_1 + \widetilde{M}_{ig(t-k)}\delta_2 + X_{idgt}\delta_3 + \widetilde{D}_{dt}\delta_4 + \hat{G}_{idgt}\delta_5 + \upsilon_{idgt} \qquad (4)$$

The objective of the two-stage model in equations (3) and (4) is to identify the effects on test scores of taking the EOC in different grade levels. Equation (3) represents several first-stage regressions that combine to predict EOC timing for students in Missouri. The dependent variable in each first-stage regression, G_{idgt}, is an indicator equal to one if student i took the course in grade g and zero otherwise. Based on the sample sizes reported in Table 2 we divide students into four grade-groups based on EOC timing for the first-stage regression: (1) grades 7-8, (2) grade 9, (3) grade 10 and (4) grades 11-12. Equation (4) takes the fitted values from the first stage and uses them to identify the effects of course timing on EOC performance.

Most of the right-hand side variables in (3) and (4) are defined similarly to those in equation (1) with a few important differences. First, while $Y_{ig(t-k)}$ from equation (1) contains lagged MAP scores for the three most recently available years for each student (e.g. scores in grades 5, 6, and 7 for a student who took the algebra-I EOC in grade-8), $\tilde{Y}_{ig(t-k)}$ in equation (3) contains each student's scores in grade-4, grade-5, and grade-6 regardless of the grade in which the student took the algebra-I EOC. Using these early, baseline MAP scores in equation (3) is important because they are realized prior to the algebra-I grade-placement decision for all of the students in the analytic sample.[8] Given this change in the lagged score vector, the vectors $\widetilde{M}_{ig(t-k)}$ and $\widetilde{D}_{dt}$ are correspondingly re-defined. X_{idgt} is defined as in equation (1).[9]

P_{dt} is the vector of instruments in equation (3). This vector contains variables that measure the shares of students in district d and year t who take the algebra-I EOC for the first time in each grade-group (again, the groups are 7-8, 9, 10, 11-12). The instruments are conceptually similar to those used by Clotfelter et al. (2012b) and are meant to capture variation in course-taking policies across districts. After the estimation of (3), the predicted probabilities of taking the EOC in each grade level, $\hat{G}_{idgt}$, are captured and used in place of G_{idgt} in equation (4), which is pooled across all

[8] Again, note that students who take the EOC prior to grade-7 are excluded from our analysis.
[9] Given that students who have previously taken the EOC are not included in the estimation of equations (3) and (4), X_{idgt} excludes the exam-retake indicator.

grades for estimation. Our estimates of δ_5 are presented in the second column of Table 3. We also show estimates when equation (4) is estimated via simple OLS (column 1), which are similar to analogous estimates provided by Clotfelter et al. (2012a) although slightly smaller in magnitude.

Under some assumptions, the instrumental-variables estimates presented in Table 3 represent the causal effects of taking the algebra-I EOC in different grades relative to grade-9 (the omitted category). To facilitate the exposition of our approach, we momentarily grant that these identifying assumptions are maintained. In Section 4 we discuss the assumptions underlying the instrumental-variables approach – and concerns related to the failure of these assumptions – in greater detail.

Moving forward under the maintained assumption that our instrumental-variables estimates can be interpreted causally, our estimates from equation (4) indicate that taking the algebra-I EOC prior to grade-9 has a significant, negative effect on performance, as does taking the exam after grade-10. The effect of taking the EOC in grade-10 cannot be distinguished from the effect of taking the exam in grade-9. For accelerated algebra, our estimates in Table 3 are consistent in sign, although not necessarily in magnitude, with similar estimates form Clotfelter et al. (2012a, 2012b). The estimated negative effect of delayed course taking in Table 3 – which is not addressed in any previous study of which we are aware – is consistent with pre-algebra mathematics knowledge degrading over time in high school. Skill degradation issues are more likely if there are gaps in mathematics course taking for students who significantly delay the algebra-I EOC. Comparing columns 1 and 2 of the table illustrates the importance of the IV estimation strategy – OLS estimates would wrongly suggest that accelerated algebra-I course taking improves performance.

To incorporate the influence of course timing into the larger evaluation procedure, we adjust the student-level residuals from equation (1) to account for the appropriate course-timing corrections. The adjusted residual for student i who took the algebra-I EOC in grade g can be written as:

$$\epsilon_{idgt}^{adj} = \epsilon_{idgt} + Q_g \tag{5}$$

where Q_g is the coefficient from Table 3 corresponding to the effect of taking the exam in grade g. Note that equation (5) imposes performance penalties on students who take the exam in grades 7-8 and grades 11-12.[10] Students who take the exam in grades 9 and 10 are not penalized (recall the effect of taking the course in grade-10 cannot be statistically distinguished from taking it in grade-9). Once the adjusted residuals are calculated we use them as outcome variables in a revised version of equation (2), producing a set of district performance measures modified to account for grade-placement effects:[11]

$$\epsilon_{idgt}^{adj} = I_{idgt}\lambda + u_{idgt} \tag{6}$$

Keeping in mind that equations (1) and (2) estimate student performance within grade level, equation (6) produces comparable estimates that additionally account for the fact that some students would have performed better had they taken the course in a different grade. In this way, a comparison of the unadjusted to the adjusted estimates provides an indication of how district course-timing policies are promoting or inhibiting student performance on the algebra-I EOC.

Figure 1 illustrates how the correction in equation (6) alters the district performance measures. In the first panel of the figure, the unadjusted measures as estimated by equation (2) are plotted against the percentage of students in the district who take the algebra-I EOC in grades 9 and 10. The low correlation (not statistically significant) between these two values is a result of the fact that the unadjusted estimates do not account for the effects of grade-placement policies on exam

[10] Students who retake the EOC in grades 11-12 are exempted from the performance penalties.
[11] Note that the course-timing adjustment parameters are treated as deterministic in equation (5). The fact that the adjustment parameters are estimated with error can be accounted for directly if desired.

performance. In contrast, the second panel in the figure plots the adjusted measures (from equation 6). The result is that there is now a positive correlation between the performance measures and the percentage of students in the district who take algebra-I in grades 9 and 10. The red and green dots indicate cases where districts change status in terms of whether they are identified as being statistically different from average, with red indicating a decline in status and green indicating an improvement. Districts that pursue more effective course-timing policies improve relative to their peers after the adjustment.

Finally, we briefly note an operational issue with regard to implementing the course-timing adjustment. It is important that the adjustment parameters be estimated using data that pre-dates the evaluation system that incorporates information on course timing. Estimating these values concurrently with an evaluation system that takes them directly into account is problematic because the estimates will be affected by district behavioral responses to the evaluation.

3.3 Allowing for Practitioner Discretion

One limitation of the course-timing adjustments so far is that they are implemented uniformly for all students without discretion. That is, the procedure thus far does not account for differences in student aptitude, etc., that might justify different course-taking paths for some students.[12] On one end of the spectrum, students with lower prior mathematics achievement may benefit from delaying course taking beyond grade 9 or 10, particularly if the delay allows them to develop the skills required to pass the algebra-I EOC. At the other end, high ability students who are

[12] We attempted to estimate the course-timing models separately for different groups of students, e.g. by decile or by quintile of prior achievement, which would have allowed for some variation in the grade-placement corrections across the student achievement distribution. Unfortunately, the declining sample sizes led to results that were generally not significant and often logically inconsistent – at least some of the estimation problems we encountered are related to applying instrumental variables estimation to small samples. Clotfelter et al. (2012b) find that the negative effects of accelerated algebra are not monotonic by achievement quintile. For example, students in the lowest quintile of prior achievement see a negative effect that is smaller (in absolute terms) than those in the fourth quintile, and much smaller than those in the second and third quintiles. Given that precisely identifying the effects of course timing across the achievement distribution will be challenging under most circumstances, in this section we propose an *ad hoc* procedure that builds general flexibility into the evaluation system.

ready to take algebra-I in grade-8 may benefit from the accelerated course path, as it would allow them to take higher-level math courses sooner, such as pre-calculus and calculus. Evidence from Clotfelter et al. (2012b) suggests that these are important concerns, and they provide direct evidence that some high-achieving students benefit from accelerated coursework. In particular, although students from every quintile of the prior-achievement distribution have lower algebra-I EOC scores if they take the course in grade-8 or before, students in the top quintile are more likely to pass geometry by grade-11 if they accelerate their algebra coursework.

To address this need for flexibility, we allow for "penalty forgiveness" for some students. Based on Clotfelter et al.'s (2012b) findings, we exempt students in the top quintile of the grade-6 math achievement distribution from the penalty if they take the algebra-I EOC prior to grade-9. We symmetrically allow students in the bottom quintile of the prior achievement distribution to be exempted if they take the EOC in grades 11 or 12. Hence, districts receive no penalty for letting some high-performing students take the exam early, nor do they receive a penalty for allowing some low-achieving students to take the exam late.[13]

Applying "penalty forgiveness" as described in the previous paragraph does not induce a large change in the effect estimates overall (the correlation between the district performance estimates with and without penalty forgiveness exceeds 0.99). However, it does meaningfully alter the substantive results for several districts. To illustrate, consider dividing the school districts in Missouri into three groups based on their total performance measures: (1) statistically below average, (2) statistically indistinguishable from average and (3) statistically above average. After we allow for penalty forgiveness, thirteen districts see an improvement in their status while another ten see their status change for the worse. The reason for these changes is apparent in Table 4, which shows the

[13] Of course, these cut-offs could be modified depending on the goals of policymakers and/or other adjustments could also be made. For example, course-timing penalties could be eliminated entirely for students who take the course after grade-9.

percentage of students receiving accelerated and delayed course-taking penalties with and without penalty forgiveness for the thirteen districts that experience an improvement in status.[14] As can be seen in the table, a large portion of students in these districts receive penalty forgiveness. In fact, the average district in Table 4 went from having 36.6 to 12.2 percent of its students receiving a course-timing penalty, a 66.7 percent decrease.[15]

4. Identification of the Course-Timing Effects Using Instrumental Variables

We use the percentage of students in each district who take the algebra-I EOC in each grade, P_{dt}, to instrument for the grade-level indicator variables in equation (4). The district enrollment shares by grade level are correlated with student-level grade placement (instrument relevance) as indicated by Table 5, which reports estimates from the first-stage regression shown in equation (3). Note that the instrument corresponding to the grade-level regression being estimated (in the highlighted cells) is always the most predictive.

Turning to the issue of instrument validity, the conceptual appeal of the instruments is that the identifying variation reflects district grade-level-placement policies – precisely the policies that evaluators will want to consider. These policies are exogenous for individual students conditional on district-of-attendance. For example, holding all else equal, a student who attends a district where students typically take algebra-I in grade-8 will be more likely to take algebra-I in grade-8 herself. Furthermore, the IV parameters are estimated conditional on individual and district-aggregated measures of achievement and student demographics, which limits first-order concerns about confounding variables related to the endogenous selection of course-timing policies by districts and endogenous student sorting.

[14] Districts with fewer than 20 students are excluded from Table 4.

[15] For the declining districts, the opposite holds true. These districts have the vast majority of their students taking the course in the optimal grades and, as such, do not receive much in the way of penalty forgiveness.

Still, it is unlikely that a compelling defense of instrument validity – one strong enough to convince a steadfast skeptic – can be mounted in our application. As just one example of a threat to instrument validity that we cannot rule out, it may be that conditional on all of the observable information we have about students and school districts, districts with higher-quality teachers are more likely to push for earlier algebra-I course taking.[16] However, it is important to recognize that even an instrument for which the exclusion restriction must be relaxed can still be useful (see Conley, Hansen and Rossi, 2012). This is particularly likely to be the case if (1) the direction of the likely bias can be signed and (2) outside evidence is available to support the notion that the instrument is providing useful information. Both of these conditions are met in this case.[17]

On point (1), if we operate under the assumption that there is some bias in the IV estimates, it is worthwhile to consider its direction. Table 6 shows the average characteristics of districts with modal grade-8 course-timing policies and modal grade-9 course-timing policies. In line with what one might expect, modal grade-8 districts are positively selected, particularly along the dimension of MAP achievement. Although we can deal with the observable differences in the table by directly conditioning on this information in the models, one might infer from Table 6 that any lingering unobserved differences between districts with different course-timing policies are similar directionally (e.g., see Altonji, Elder and Taber, 2005). This would suggest that districts more likely to see higher conditional EOC performance are also more likely to accelerate algebra-I course taking. Noting that available evidence shows that the causal effect of accelerating algebra-I course

[16] Even this story does not seem particularly likely. Our use of district-level course placement percentages rather than school-level percentages means that the teacher quality differentials would have to be district-wide to invalidate the instruments. Most of the variance in teacher quality occurs within schools (Hanushek and Rivkin, 2012). Furthermore, the fact that our models condition on district characteristics means that the cross-district variance in teacher quality must not be highly correlated with observable district characteristics in order to confound our instrumental-variables estimates.

[17] Our work could be extended to formally apply the techniques laid out in Conley, Hansen and Rossi (2012). They provide a rigorous framework for examining the sensitivity of the IV estimates to deviations from the exact exclusion restriction.

taking on achievement is negative (Clotfelter et al., 2012a, 2012b), any such bias would imply that the "course-timing penalty" terms that we apply in equation (5) are too small (but per Table 3, still signed properly).

From the perspective of administrators, course-timing penalties that are directionally accurate but attenuated can still be quite useful. They will still incentivize more effective policies, even if the incentives are not as strong as would be the case if the instruments were truly exogenous. Also note that administrators may prefer undersized penalties in equation (5) if they view the costs of over-penalizing districts as higher than the costs of under-penalizing districts.

Returning to point (2) from above, Clotfelter et al. (2012a, 2012b) provide estimates that we can compare to our estimates in Table 3, at least for accelerated algebra-I course taking. Our estimate of the effect of district policies that accelerate algebra-I to grade-8 relative to grade-9, -0.220 as reported in Table 3, is roughly two-thirds the size of analogous estimates reported in their studies.

One possible explanation for the discrepancy is that there is lingering bias in our estimates driven by the failure, to some degree, of the exclusion restrictions for the instruments. However, it is also possible that both estimates are correct, in which case the discrepancy can be explained by the fact that Clotfelter et al. (2012a, 2012b) aim to identify the effects of sharp changes in course-taking policies within school districts, while our model is designed (under the identifying assumptions) to capture "steady-state" differences in algebra-I course-timing policies across districts. The reason that this is important in our context is that a sharp policy change to accelerate algebra-I course taking may not have the same effect as a steady-state accelerated algebra-I policy. For example, in the latter case districts may be better able to tailor lead-in courses to accommodate students taking algebra-I in grade-8, where a sharp policy change will be less accommodating in this regard (this caveat to their findings is noted by the authors).

Although we cannot precisely resolve the discrepancy in our estimates and the estimates available from Clotfelter et al. (2012a, 2012b), a comparison of our study to theirs suggests that our approach provides an estimate for the accelerated course-taking penalty that may be too small, but is properly signed and of a magnitude that will be useful for incentivizing districts to structure the timing of algebra-I course taking effectively.[18]

Regardless of how one interprets the strength of our case for using the penalties for accelerated algebra-I from Table 3, the case for using the delay penalties is weaker. There are two reasons for this. First, using a symmetric argument about bias from unobservables, it seems likely that if the IV estimates are biased the direction of the bias is *negative* for districts that delay algebra-I, which will lead to estimated penalties in Table 3 that are too large for delayed course taking. A second problem, which compounds the first, is that we have no external study to which we can compare our findings for delayed course taking.[19]

Policymakers may be concerned about over-penalizing districts where students take algebra-I after grade-10. In the face of this concern, one possible remedy would be to remove the penalty for delayed course taking entirely, or similarly, scale the penalty down by an appropriate factor. The approach described in Conley, Hansen and Rossi (2012) may be useful in the latter case. It should also be noted that empirically speaking, concerns over delayed algebra-I course taking are secondary to concerns about accelerated course taking – this can be seen in Table 2, which shows that

[18] An added advantage of the method presented in this paper from the standpoint of designing an evaluation system is that no student records are systematically excluded from the model. This is in contrast to the method used in Clotfelter et al. (2012b) in which district-by-prior-achievement cells are removed from the analysis if they do not have enough variance over time to rule out random enrollment fluctuations, a procedure that was implemented to potentially help limit endogeneity concerns and improve the case for the instruments being valid. Educational administrators and policymakers often place considerable weight on "inclusion" considerations for political reasons. Such considerations are typically of less importance to researchers.

[19] Clotfelter et al. (2012a, 2012b) only consider the effects of accelerated, not delayed, algebra taking.

approximately 50 percent more students take algebra-I in grade-8 than take it in grades 11 and 12, a differential that is even larger when one considers first-time exam takers only.[20]

5. Diagnostic Value of the Model and Other Concerns

Diagnostic Value of our Approach

Although accounting for the effect that district-level grade-placement policies have on student achievement is the primary motivation for the design we have described above, the multi-part structure of the model also provides valuable diagnostic information that can be used by both policymakers and practitioners to improve student outcomes. For example, consider a district that has implemented a policy whereby most of its students take algebra-I in grade-8. Suppose that instructional quality in this district is high, and as such, the students in the district are performing better on the exam than other grade-8 algebra-I students in the state (although worse than they would have performed if they had taken the course in grade-9, all else equal). The high quality of instruction delivered by the district is captured by the unadjusted district effect estimates produced by equation (2). Districts that promote good instructional strategies (e.g. better teachers, improved curricula, enhanced tutoring services) can be identified and used as models for other districts in the state in this regard.

On the other hand, this hypothetical district's grade-8 policy is actually harming student achievement, a problem that should not be ignored and that the above-outlined procedure is designed to identify and address. In this case, the district's adjusted effect estimate would decline markedly from its unadjusted estimate. The adjusted and unadjusted effect estimates, which could be reported side-by-side, provide valuable diagnostic information to policymakers and practitioners. Districts with effective instruction and ineffective grade-placement policies can be made aware of

[20] Students retaking the exam in grades 11 and 12 are exempted from the delayed course-taking penalty.

this situation and then work to remedy it (a relatively easy policy fix), while districts with ineffective instruction but effective grade-placement policies can focus on instructional issues.

Extensions to School- and Teacher-Level Models

The diagnostic nature of the model also points to how the district-level model might be adapted to both school- and teacher-level evaluations. The grade-placement policies that the model is designed to account for are largely out of the control of individual teachers. As such, the instructional quality measures, unadjusted for the grade-placement corrections, are the natural measures to use to form the foundation for teacher-level evaluations. That said, substantial challenges remain in developing teacher-level performance measures based on student EOC performance beyond simply accounting for course-timing effects. Perhaps the central concern is how to deal with student tracking, an issue discussed in recent studies by Anderson and Harris (2013) and Jackson (2013).[21]

Schools, on the other hand, likely lie somewhere between districts and teachers in their ability to influence the grades in which students take specific courses. For example, a school with active leadership may accelerate courses for their students even in the absence of a formal district policy of that nature. As such, a school-level model could build in grade-placement penalties for sub-optimal deviations from district course-placement policies if schools are presumed to have considerable influence in this regard. This would hold schools accountable for their own internal policies, but not larger district policies.

Incentivizing Course Taking

Another important difference between standardized exams in grades 3-8 and high school tests is that students need not take courses that are tied to EOCs. Whether this is of concern

---

[21] Of course, the initial value-added models we use in Section 3.1 remove the effects of cross-grade tracking on student achievement by construction.

depends on the specific course and on students' educational plans. By making EOC performance a part of the larger evaluation system, however, it is important to be cognizant of the potential to create incentives that inadvertently encourage districts to keep some students from taking certain courses.

The model presented above is flexible enough to account for students who never take the EOC. One remedy that has already been mentioned is to simply remove all penalties for late-taking students. This would reduce district incentives to discourage marginal students from taking the course. To directly incentivize districts to get students to take key EOCs, however, students who never take the test can be directly incorporated into the performance measures.[22] Generally, the first instinct in such situations is to impute an exam score for the missing students. However, the student-level measures used to determine the district effect estimates are the residuals from equation (1), i.e. the deviations of students' actual exam scores from their predictions. Hence, by definition, any student with an imputed score would have a zero residual, and their inclusion in the model would simply pull the district estimates toward the mean.

An alternative is to assign a negative value for each student who does not take the exam, purposefully building in a penalty to districts for these students. A natural value to use for this purpose is the sum of the district effect estimate and the grade-level penalty for students who take the course in grade-12 (-0.202 from Table 3).[23] For EOCs that are required for all students (like algebra-I in Missouri), the penalty is assigned for all students who never take the test, even those that might have otherwise received penalty forgiveness had they taken the exam in grade-12. The lack of

[22] Determining the list of students who never take algebra-I and, more importantly, assigning them to accountable districts, is challenging; but reasonable options are available. See Appendix B for further discussion of this issue.
[23] For high performing districts, this sum might be positive. To deal with this issue and provide proper incentives to district officials, we assign students who never take the test residual values of zero plus the grade-12 penalty in districts with positive performance measures. In addition to properly aligning incentives, this implicitly assumes that students who never take the course are likely worse than those who take it in grade-12. A similar assumption is made by Clotfelter et al. (2012a, 2012b) in imputing exam scores for students who never take the EOC.

forgiveness for students who never take the test provides strong incentives for districts to encourage students to take the course.[24] A discussion of the technical details of the adjustment for students who never take the test and results showing how the inclusion of these students affects the evaluation procedure overall are provided in Appendix B.

6. Conclusion

One limitation of the current use of standardized exams as part of educational evaluation systems is that they are limited to tested grades and subjects, typically mathematics and communication arts in grades 3 through 8. The increased availability of EOC data provides an opportunity to extend the reach of test-based performance evaluations into what have, up until this point, been considered non-tested grades and subjects. However, the extension of models designed to analyze student performance on (nearly) universally administered standardized tests is problematic when extended to EOCs for two reasons. First, the grade in which the course is taken is a choice variable and is correlated with unobserved student-level characteristics such as academic aptitude. Second, recent research suggests that district- and school-policies that affect the grade in which courses are taken can meaningfully impact student achievement (Clotfelter et al., 2012a, 2012b). The procedure developed in this paper attempts to deal with these issues within a framework well-suited for use in educational evaluation systems.

The first step in our approach ignores the course-timing issue and produces district performance measures conditional on the grade levels in which students take EOCs. These measures capture district "instructional effectiveness" conditional on when students take courses. We then explicitly incorporate the effects of course-timing policies to provide a direct accounting for the role that these policies play in determining student achievement. As a final adjustment, we introduce

[24] This method for dealing with students who never take the test is conceptually more difficult in non-required courses. In these cases, one alternative would be to empirically determine a likelihood of success in the course based on prior achievement and then exclude students below some threshold value from the model.

flexibility in imposing the course-timing penalties. This allows some students to take courses in grade levels that our models indicate are suboptimal for *most* students without penalty, thereby facilitating student, parent and practitioner discretion in determining when individual students take courses.

The end result is a district performance measure that provides diagnostic value. For example, districts can use the results from the "instructional effectiveness" portion of the procedure to determine if they need to replace or refine their instructional methods, while they can infer from the course-timing adjustments whether their course-timing policies are in the best interest of students. A final advantage of our approach is that it provides policymakers with a wide degree of flexibility in precisely how to apply the grade-placement penalties, which can be adjusted depending on the policy objectives being pursued.

References

Altonji, Joseph G., Todd E. Elder and Christopher R. Taber. 2005. Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy* 113(1), 151-184.

Anderson, Andrew and Douglas Harris. 2013. Bias of Public Sector Worker Performance Monitoring: A Structural Model and Empirical Evidence from Secondary School Teachers. Working Paper.

Betts, Julian. 1995. Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth. *Review of Economics and Statistics* 77(2), 231-250.

Chetty, Raj, John N. Friedman and Jonah E. Rockoff. 2011. The Long-Term Impacts of Teachers: Teacher value-added and student outcomes in adulthood. NBER Working Paper No. 17699.

Clotfelter, Charles T., Helen F. Ladd and Jacob L. Vidgor. 2012a. The Aftermath of Accelerating Algebra: Evidence from a District Policy Initiative. NBER Working Paper No. 18161.

-- 2012b. Algebra for 8th Graders: Evidence on its Effects from 10 North Carolina Districts. NBER Working Paper No. 18649.

Conley, Timothy G., Christian B. Hansen and Peter E. Rossi. 2012. Plausibly Exogenous. *Review of Economics and Statistics* 94(1): 260-272.

Ehlert, Mark, Cory Koedel, Eric Parsons and Michael Podgursky. 2013. Selecting Growth Measures for School and Teacher Evaluations: Should Proportionality Matter? CALDER Working Paper No. 80.

Hanushek, Eric A and Steven G. Rivkin. 2010. Generalizations about Using Value-Added Measures of Teacher Quality. *American Economic Review* 100(2), 267-271.

Hanushek, Eric A and Steven G. Rivkin. 2012. The Distribution of Teacher Quality and Implications for Policy. *Annual Review of Economics* 4, 131-157.

Jackson, Kirabo C. Forthcoming. Teacher Quality at the High-School Level: The Importance of Accounting for Tracks. *The Journal of Labor Economics*.

Kane, Tom J., Jonah E. Rockoff, and Douglas O. Staiger. 2008. What Does Certification Tell us about Teacher Effectiveness? Evidence from New York City. *Economics of Education Review* 27(6), 615-631.

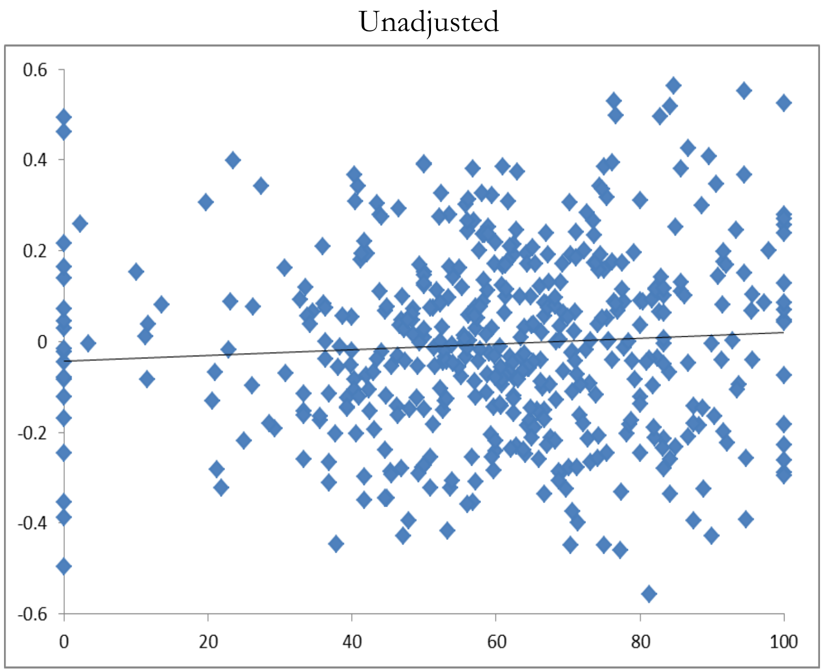
Koedel, Cory, Rebecca Leatherman and Eric Parsons. 2012. Test Measurement Error and Inference from Value-Added Models. *The B.E. Journal of Economic Analysis & Policy* 12(1). (Topics).

Nye, Barbara, Spyros Konstantopoulos and Larry V. Hedges. 2004. How Large are Teacher Effects? *Educational Evaluation and Policy Analysis* 26(3), 237-257.

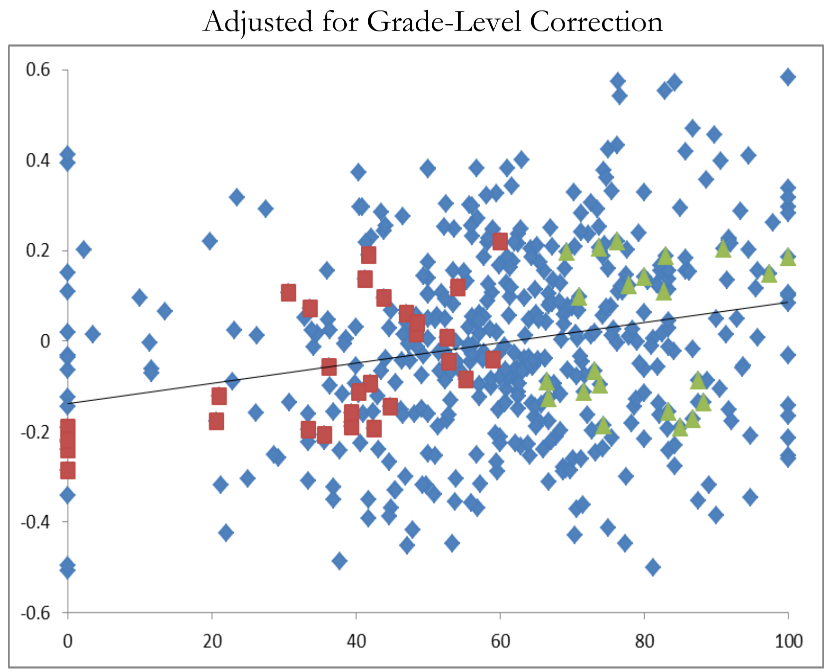
Rivkin, Steven G., Eric A. Hanushek and John F. Kain. 2005. Teachers, Schools and Academic Achievement. *Econometrica* 73(2), 417-58.

Rockoff, Jonah. 2004. The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review (P&P)* 94(2), 247-252.

Figure 1. Unadjusted and Grade-Level-Correction-Adjusted Estimated EOC Performance Measures for Missouri School Districts Plotted against the Percentage of Students in the District who Take Algebra I in Grade-9 or Grade-10.



Correlation: 0.066 (p-value ≈ 0.15)



Correlation: 0.223 (p-value < 0.01)

Notes: The right-hand panel of the figure shows adjusted estimates prior to the flexibility we introduce in the assignment of course-timing penalties in Section 3.3 and prior to accounting for students who never take the algebra-1 EOC. After incorporating our entire procedure, the correlation between the district performance measures and the share of students who take the EOC in grades 9/10 moderates slightly - falling to approximately 0.18 (p-value < 0.01). A graph analogous to what we show here displaying this final correlation is available upon request.

Table 1. Summary Statistics.

Analytic Sample Size	
Number of Districts	487
Number of Schools	848
Number of Students	68,621
Student Characteristics	
Percent Female	49.5%
Percent Free/Reduced Price Lunch Eligible	42.9
Percent Minority	21.4
Percent English as a Second Language	2.0
Percent with an Individualized Education Plan	10.0
Percent Mobile	4.8

Notes: A student is defined as mobile if she does not attend the school in which the exam was taken for the entire school year.

Table 2. Grade Distribution of the 2012 Algebra-I EOC.

	Grade Level of EOC							
All Students	Missing	< 7	7	8	9	10	11	12
No. of Students	13	57	694	14488	32124	11634	5090	4521
Percent of Students	0.0	0.1	1.0	21.1	46.8	17.0	7.4	6.6
First-time Test Takers	Missing	< 7	7	8	9	10	11	12
No. of Students	10	57	693	14470	31314	10906	3955	3991
Percent of Students	0.0	0.1	1.1	22.1	47.9	16.7	6.1	6.1

Table 3. Grade-Level Coefficients from Pooled Grade-Level Models (Equation 4).

	OLS	IV
Grades 7 and 8	0.149**	-0.220**
	(0.008)	(0.026)
Grade 10	-0.176**	0.012
	(0.008)	(0.032)
Grades 11 and 12	-0.510**	-0.202**
	(0.0112)	(0.044)
Student-Level Controls		
Grade-4, 5, and 6 Exam Scores (Both Subjects)	X	X
Missing Exam Score Indicator Variables	X	X
Demographics	X	X
District-Level Aggregates of Student-Level Controls	X	X

Notes: ** represents statistical significance at the 0.01 level.

Table 4. The Effect of Grade-Level Correction Forgiveness on Student Residuals (Districts with Significantly Improved Effect Estimates).

	Percentage of Student Residuals Receiving an Accelerated Course-Taking Penalty		Percentage of Student Residuals Receiving a Delayed Course-Taking Penalty	
	Before Forgiveness	After Forgiveness	Before Forgiveness	After Forgiveness
District 1	21.2	4.0	10.1	4.0
District 2	31.1	9.5	8.1	1.4
District 3	28.7	10.2	5.4	1.8
District 4	25.5	9.9	6.8	2.4
District 5	26.5	11.8	20.6	11.8
District 6	11.8	3.4	25.2	10.9
District 7	23.4	3.2	13.8	6.4
District 8	25.5	3.6	19.1	9.5
District 9	12.9	1.4	30.1	19.2
District 10	20.7	1.7	9.8	3.5
District 11	35.0	10.5	2.3	1.1
District 12	0.0	0.0	33.2	10.7
District 13	18.8	4.2	10.4	3.1
Simple Average	21.6	5.6	15.0	6.6

Table 5. Results from the First-Stage of the Grade-Placement Instrumental Variables Regressions.

	Dependent Variables: Student took the EOC in Grade:		
	Grades 7/8	Grade 10	Grades 11/12
Instruments			
Share in District Taking Exam in Grades 7 and 8	**0.010****	0.000*	-0.000
	(0.000)	(0.000)	(0.000)
Share in District Taking Exam in Grade 10	0.001**	**0.010****	-0.000
	(0.000)	**(0.000)**	(0.000)
Share in District Taking Exam in Grades 11 and 12	0.000*	0.001**	**0.006****
	(0.000)	(0.000)	**(0.000)**
Other Controls			
Grade-4, 5, and 6 Exam Scores (Both Subjects)	X	X	X
Missing Exam Score Indicator Variables	X	X	X
Demographics	X	X	X
District-Level Aggregates of Student-Level Controls	X	X	X

Notes: * represents statistical significance at the 0.05 level.
** represents statistical significance at the 0.01 level.

Table 6. Characteristics of Districts with Grade-8 and Grade-9 Modal Algebra-I Course Assignment.

	Modal Grade for Algebra-I Couse Taking									
	Grade-8 (n = 71)					Grade-9 (n = 360)				
	Mean	Std. Dev.	Quartile 1	Median	Quartile 3	Mean	Std. Dev.	Quartile 1	Median	Quartile 3
Avg. Grade-6 MAP Math Score	0.389	0.458	0.085	0.323	0.555	0.077	0.308	-0.113	0.095	0.260
Avg. Grade-6 MAP Com Arts Score	0.288	0.373	0.043	0.183	0.531	0.048	0.287	-0.118	0.080	0.240
Percent Female	50.3	14.9	44.0	50.0	56.5	48.8	9.0	44.4	49.4	53.8
Percent F/RL	40.8	22.1	26.7	38.7	52.2	49.3	18.5	36.9	48.5	60.9
Percent Minority	10.5	21.9	0.0	2.0	7.3	11.0	20.2	0.0	4.6	10.7
Percent of Students with an IEP	6.5	6.6	0.0	4.7	11.8	10.0	7.7	5.2	9.1	13.4
Percent ESL	0.5	1.6	0.0	0.0	0.0	1.3	5.4	0.0	0.0	0.8
Percent Mobile	3.3	5.1	0.0	1.6	5.1	4.8	5.0	1.6	3.9	6.6

Appendix A
Controlling for Incomplete MAP Score Histories

The inclusion of three years of lagged scores in the model described in this paper combined with the fact that, in some cases, these lagged exam scores may be up to six years old (for students taking the exam in grade-12), increases the incidence of missing data in the EOC model. The general method used to control for this issue parallels that in Ehlert et al. (2013) – that is, missing exam scores are set to zero (the standardized mean) and indicator variables are initialized to indicate missing data. However, the length of the lagged score vector combined with the fact that some algebra-I EOC takers have no prior MAP records complicates matter.

By way of comparison, in the model presented in Ehlert et al. (2013), students are *required* to have a same-subject lagged exam score to be included in the model estimation. Hence, only one missing exam score indicator variable is needed in the model. However, to control for every possible combination of missing lagged exam score data in the EOC model would require $2^6 = 64$ dummy variables, many of which could not be included in every by-grade regression as no students in the given grade would have that missing score combination. In addition, some students have no prior MAP exam scores to use as predictors.

As a result, a simplified approach is taken whereby only four missing exam score dummies are included in $M_{ig(t-k)}$ – one to indicate if the student had no lagged MAP records, a second to indicate if the student was *only* missing the lag 3 exam scores (both subjects), a third to indicate if he/she was missing the lag 2 *and* lag 3 exam scores (both subjects), and a fourth that captures any other missing exam score combination. The first three of these dummy variables most likely capture student migration and transfer, i.e. students who moved in from out-of-state at some point over the course of their grade-3 to grade-8 careers or students who transferred from private to public schools. This is due to the fact that, for any given year, both of the student's exam scores are either present or missing. In contrast, the last missing-test-score(s) indicator variable likely captures attendance issues during the week of exams, potentially combined with student mobility issues. Overall, these more broadly-defined controls work well for the algebra-I model presented in this paper. They also have the benefit of being easily adaptable to other EOCs. The distribution of these dummy variables by grade is presented in Table A.1.

Table A.1. Missing Test Score Percentages by Grade and Category.

	Grade					
	7	8	9	10	11	12
No Missing Scores	85.6%	91.3%	86.2%	83.0%	77.4%	61.7%
Missing MAP Lag 1, 2, 3	3.6	2.0	5.8	7.0	10.1	13.8
Missing MAP Lag 2, 3	5.2	2.5	2.6	2.7	2.7	4.8
Missing MAP Lag 3	3.0	3.3	2.9	3.5	3.7	10.9
Missing MAP Lag - Other	2.6	1.0	2.5	3.8	6.1	8.8
Total N	694	14488	32124	11634	5090	4523

Appendix B
Accounting for Students who Never Take Algebra-I

Although it is conceptually simple to deal with students who never take algebra-I in the manner described in the main body of the paper, implementation is challenging. An important question is how to determine who qualifies as a student who never takes algebra-I and, additionally, how to assign these students to districts. There are a several possible ways this can be accomplished, all of which have costs and benefits. This appendix discusses our preferred method, which we call "fractional grade-9 forward" assignment.

We start by defining the cohort of students entering grade-9 in 2009 (and, hence, who are expected to graduate in 2012). For each student the district of attendance is identified as the first district in which the student is enrolled for at least ten percent of the 2009 school year. This selection rule was chosen because course taking for mobile students in receiving districts is likely to be largely driven by the student's course schedule in the sending district. Students are then tracked forward through 2012 with the district of attendance in each year assigned according to the same rule.

During the tracking process, student EOC and course records are analyzed to determine if each student takes algebra-I over the course of their academic careers. To account for the fact that mobile students may lack an algebra-I EOC score if they take the course outside of the Missouri public school system, students are flagged as having taken algebra-I if *any* of the following criteria are met:

1. The student has a record for *any* of the mathematics EOCs. In Missouri, these include algebra-I, algebra-II, and geometry.

2. Course assignment records indicate that the student took algebra-I in any academic year.

3. Course assignment records indicate that the student took a higher mathematics class in any academic year. In Missouri, these courses include algebra-II, geometry, math analysis, trigonometry, algebra-trigonometry, analytic geometry, calculus, and pre-calculus.

Conversely, students who do not meet any of the above criteria are flagged as having never taken algebra-I.

Once students who never take algebra-I are flagged, they are added to the residual list for each district to which they are assigned from 2009 to 2012 with the estimated adjusted residual for student i in district d equal to:

$$\hat{\epsilon}_{idnt}^{adj} = \bar{\hat{\epsilon}}_d + f_{idnt} \times Q_{12} \qquad \text{(B.1)}$$

if $\bar{\hat{\epsilon}}_d \leq 0$ and

$$\hat{\epsilon}_{idnt}^{adj} = f_{idnt} \times Q_{12} \qquad \text{(B.2)}$$

if $\bar{\hat{\epsilon}}_d > 0$. Here, $\bar{\hat{\epsilon}}_d$ is the average estimated adjusted residual for all students in district d (with penalty forgiveness applied), f_{idnt} is the fraction of student i's high school career that is assigned to district d, and Q_{12} is the grade-12 adjustment from Table 3.

The fractional penalties assigned by equations (B.1) and (B.2) are designed to (a) account for student mobility across districts and (b) error on the side of a conservative penalty in cases where it is uncertain whether missing student data is the result of student mobility across state lines or student drop-outs. For example, if a student appears in the Missouri data for all four years but is assigned to one district the first two years and another district the final two years, that student is added to the residual list for *both* districts. However, the penalty assigned in each case is one-half of the grade-12 adjustment. Similarly, a student who only appears in the data for a single year is attached to a single district's residual list with an additional penalty that is one-quarter of the grade-12 adjustment. In cases where drop-outs can be easily distinguished from out-of-system transfers, a reasonable alternative would be to set $f_{idnt} = 1$ and give the full weight of the penalty to the district where the drop-out occurred. This would provide a further incentive for districts to lower their drop-out rates, while not penalizing districts that do a good job of keeping marginal students in school.

Data on students from the 2009 grade-9 cohort who never take algebra-I are presented in Table B.1. The first thing to note is that most of the students disappear from the state data before completing high school. In fact, over 40 percent of the students only appear in the state during 2009, while less than 20 percent were present in the state for all four years. Furthermore, for those students who appear in the data for multiple years, there is also a large amount of cross-district mobility. This high degree of mobility among students who never take algebra-I, both internal and external of the Missouri public school system, is what motivates the development of the assignment method described above.[25]

Given the addition of the students in Table B.1 to district rosters, Table B.2 presents a list of the districts for which the performance rankings are significantly changed (relative to the overall mean) by the adjustment, as well as the percentage of the residual counts for these districts that consist of students who never take algebra-I.[26] The adjustment for students who never take algebra-I affects relatively few districts, especially when compared to changes resulting from the grade-placement adjustments.

[25] Of course, alternatives to the "fractional grade-9 forward" method exist. One option would be to simply assign the full penalty to each student's grade-9 district. Although this method would be blind to student mobility both within the state and across state lines, an argument can be made that this is the most logical place to assign the penalty, as grade-9 has been empirically determined to be the optimal grade for algebra-I course taking. Another option would be to define the cohort in grade-12 and track them backwards. One argument in favor of the grade-12 backward assignment mechanism is that the entity ultimately responsible for ensuring a student takes the required courses is that student's grade-12 district, as this district represents the student's last point-of-contact with the secondary school system. A drawback of the grade-12 backward method, however, is that it excludes students who drop out of high school prior to their senior year. It can also potentially punish districts that receive out-of-state students (or private school transfers) who take algebra-I in the sending school but do not take any further mathematics courses.

[26] The addition of students who never take algebra-I to the district rosters does not provide any more information regarding the precision of the effect estimates. Therefore, the standard errors used to calculate the confidence intervals in Table B.2 are based on the analysis preceding the inclusion of these students.

Table B.1. Mobility for Students Who Never Take Algebra-I.

Number of Years that the Student Appears in the Missouri Public School System	Number of Districts Attended	Frequency Count	Percentage of Total
1	1	3534	42.7%
2	1	1466	17.7
	2	603	7.3
3	1	705	8.5
	2	450	5.4
	3	99	1.2
4	1	967	11.7
	2	314	3.8
	3	105	1.3
	4	29	0.4
Total		8272	100.0

Table B.2. Districts that Experience a Status Change as a Result of the Adjustment for Students who Never Take Algebra-I.

		Distribution of EOC Exams Across Grades					
		Grade 7/8	Grade 9	Grade 10	Grade 11/12 1st taker	Grade 11/12 Retaker	Never Takers
Improving Districts							
Negative Significant to Not Significant	District 1	32.3%	26.5%	9.8%	19.3%	0.7%	11.5%
	District 2	0.0	3.2	17.5	46.0	9.5	23.8
	District 3	0.0	60.0	14.3	20.0	0.0	5.7
	District 4	30.8	42.0	11.1	10.2	1.4	4.5
	District 5	0.0	40.0	47.5	3.8	0.0	8.8
	District 6	0.0	35.3	31.4	17.6	2.0	13.7
Not Significant To Positive Significant	District 7	25.0	61.6	5.4	3.6	0.0	4.5
	District 8	17.4	46.1	0.9	30.4	0.0	5.2
	District 9	0.0	55.6	22.2	14.1	0.0	8.1
	District 10	20.2	41.2	20.2	8.4	0.0	10.1
Declining Districts	District 11	23.8	48.5	3.3	3.3	1.6	19.5
Positive Significant to Not Significant	District 12	12.9	10.1	23.4	30.1	0.0	23.4
	District 13	0.0	66.7	9.5	4.8	0.0	19.0
	District 14	0.0	28.1	14.0	7.0	0.0	50.9
	District 15	0.0	23.9	20.0	33.2	4.9	18.0
	District 16	0.0	42.1	31.6	13.2	0.0	13.2
Not Significant to Negative Significant	District 17	11.8	17.6	22.7	25.2	1.7	21.0
	District 18	23.4	16.0	23.4	13.8	0.0	23.4