NATIONAL CENTER for ANALYSIS of LONGITUDINAL DATA in EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

A program of research by the American Institutes for Research with Duke University, Northwestern University, Stanford University, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington

AIR

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

ERIC PARSONS
CORY KOEDEL
MICHAEL PODGURSKY
MARK EHLERT AND
P. BRETT XIANG

WORKING PAPER 137 • August 2015

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

Eric Parsons
University of Missouri

Cory Koedel
University of Missouri

Michael Podgursky
University of Missouri

Mark Ehlert
University of Missouri

P. Brett Xiang
University of Missouri

Contents

Acknowledgements

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A120310 to the American Institutes for Research. The views expressed are those of the authors and should not be attributed to the American Institutes for Research, its trustees, or any of the funders or supporting organizations mentioned herein. Any errors are attributable to the authors.

The authors are from the Department of Economics, and Koedel is also from the Truman School of Public Affairs, 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.

CALDER working papers have not undergone final formal review and should not be cited or distributed without permission from the authors. They are intended to encourage discussion and suggestions for revision before final publication.

CALDER • American Institutes for Research
1000 Thomas Jefferson Street N.W., Washington, D.C. 20007
202-403-5796 • www.caldercenter.org

Incorporating End-of-Course Exam Timing into Educational Performance Evaluations
Eric Parsons, Cory Koedel, Michael Podgursky, Mark Ehlert and P. Brett Xiang
CALDER Working Paper No. 137
August 2015

Abstract

There is increased policy interest in extending test-based evaluations 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 a variety of subjects. 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 like advanced course taking. We 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 separates out the influence of school and district course-timing policies from other factors that determine student achievement.

1. Introduction

It is increasingly common for direct performance measures based on student test scores to be 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, Friedman and Rockoff, forthcoming; Hanushek and Rivkin, 2010; Konstantopoulos, 2006; Rockoff, 2004), coupled with the inability of researchers to consistently link performance differences between units to readily-observable characteristics (Betts, 1995; Kane, Rockoff and Staiger, 2008; Nye, Konstantopoulos and Hedges, 2004; Rivkin, Hanushek and Kain, 2005), motivates the use of these measures in the evaluation process.

The research literature upon which the development and use of test-based measures in education is based is predominantly comprised of studies that measure student achievement on standardized exams administered in the early grades – in particular, math and English/language arts in grades 3-8. However, educational administrators looking to broadly incorporate these 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 being administered. EOCs are currently available 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 students take EOCs 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, Ladd and Vigdor, 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 course timing in EOC evaluations. This separation achieves two objectives. First, it facilitates direct rewards/sanctions for schools and districts that set up effective/ineffective course-timing 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 grade-specific 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 critical and policy-relevant information. To give a concrete example, consider a district that is highly effective in instructional practice but has implemented suboptimal course-timing policies. Based on the findings from Clotfelter, Ladd and Vigdor (2012a, 2012b), and the evidence we present below, a

[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 like curriculum choice.

suboptimal policy 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 policies on student outcomes. Specifically, we use an instrumental variables (IV) strategy to isolate gaps in student achievement across districts that are 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.[3]

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 the data indicate *most* students should follow. The corrections that we apply are based on available research evidence (Clotfelter, Ladd and Vigdor, 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- specific value-added measures alone. A significant number of students at these affected districts are taking the algebra-I EOC in grade-8. We also discuss how our approach can be generalized to accommodate other EOCs and other levels of evaluation – e.g., schools and/or teachers. Accounting for course-timing effects will be important for evaluations of EOC performance at all levels.

[3] It is implicit in our analysis that course-timing policies are largely at the discretion of districts. This view is consistent with the variation in course-timing policies that we observe across Missouri districts (see Figure 1 below) and supported by two studies by Clotfelter, Ladd and Vigdor using data from North Carolina (2012a, 2012b).

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 test 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 the first year (algebra-I, English-II and biology). The number of EOCs administered in the state has since grown to eight (as of 2012-13) 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, Ladd and Vigdor, 2012a, 2012b), and (2) algebra-I is the most commonly administered EOC in Missouri. The outcome measures are taken from the 2011-12 and 2012-13 school years 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 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 take the EOC more than once over the course of their schooling careers. As one would expect, the distribution of students who retake the exam is heavily weighted towards the upper grades – 7.4 percent of grade-10 students, 19.9 percent of grade-11 students, and

11.2 percent of grade-12 students who took the algebra-I EOC in 2012 and 2013 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. (forthcoming) to produce "instructional effectiveness" measures for districts based on the algebra-I EOC.[4] The model is specified as follows and estimated separately by grade level:

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

$$\epsilon_{idgt} = I_{idgt}\theta + \eta_{idgt} \tag{2}$$

In equation (1), Z_{idgt} is the EOC score of student i in district d and grade g who took the test at time t. $Z_{ig(t-k)}$ 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 indicator variables controlling for missing lagged 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 variables included in the three previously-described control vectors, T_t is an indicator for the 2012-2013 school year, and ϵ_{idgt} is the error term.[5] By virtue of the grade-level estimation, the coefficients in equation (1) can differ across

[4] 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., forthcoming).

[5] All MAP exam scores are standardized by year-grade-subject cell. The outcome variable (the EOC score) is also standardized by year 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 lagged score dummy variables ($M_{ig(t-k)}$) see Appendix A. Exam re-takers are

grades (g) as indicated in the equation.[6] 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) represents how well each student performed 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 residuals are used as outcome variables in equation (2) to produce the estimates of θ.[7] 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. (forthcoming) is that equation (1) is estimated separately for each grade level.[8] This ensures that students are initially compared only to other students in the same grade. As a result, equation (2) provides measures of how well districts are educating their students conditional on the

---

included in the analytic sample that we use to estimate equations (1) and (2), and there is an indicator for re-taking status included in X_{idgt}. Note that the inclusion of these students in equations (1) and (2) does not change our findings with regard to the course-timing effects, which are estimated separately using a procedure described in the next section (that excludes re-takers).

[6] 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. The differing gaps between the lagged exam scores and the outcome variable may affect the precision of the estimates in the higher grades, but the by-grade-level estimation should limit concerns about bias, particularly at the district level.

[7] 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, and the standard errors are calculated to be robust to the presence of heteroskedasticity and are clustered at the student-level to account for re-takers. Shrinkage is applied via the method used in Koedel, Leatherman and Parsons (2012).

[8] 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 (≈0.1 percent – see Table 2).

grade in which students take the course.[9] 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, Ladd and Vigdor (2012a, 2012b) show that district policies regarding the grade-level placement of students into algebra-I can significantly affect exam performance and longer-term student outcomes such as future course taking. Clotfelter, Ladd and Vigdor (2012a) study an abrupt change in the algebra-I course-timing policy in the Charlotte-Mecklenburg School District. They 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, Ladd and Vigdor (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.[10]

Identifying the effects of course timing on test scores is challenging because the grade in which students take algebra-I is endogenous. Clotfelter, Ladd and Vigdor (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:

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

[9] Limiting comparisons to be between students taking the course in the same grade is also important for models at the school and teacher levels (we elaborate on this point in Section 5.2).

[10] A separate issue is that the EOC is administered up to three times during the academic year in Missouri (fall, spring, summer). We do not take up the issue of "within-academic-year" test timing in this study because supplementary analysis suggests it is a second-order issue. One reason is that the vast majority of students take their EOCs in the spring (in 2011-12 and 2012-13, 93.6 percent of Missouri students who took the algebra-I EOC took it in the spring, 5.4 percent took it in the fall, and 1.0 percent took it in the summer). In results omitted for brevity, we also directly estimated the effect of within-academic-year timing on achievement using an approach analogous to the one outlined below for our main analysis of grade-level timing (focusing on the fall and spring test dates) and found that within-academic-year timing is not an important determinant of achievement. More information is available from the authors upon request.

$T_t\gamma_{6g} + e_{idgt}$

$$Z_{idgt} = \delta_0 + \tilde{Z}_{ig(t-k)}\delta_1 + \widetilde{M}_{ig(t-k)}\delta_2 + X_{idgt}\delta_3 + \widetilde{D}_{dt}\delta_4 + \hat{G}_{idgt}\delta_5 + T_t\delta_6 + \upsilon_{idgt} \qquad (4)$$

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-group g and zero otherwise. Based on preliminary analysis of the course-timing effects, we divide students into three grade-groups based on EOC timing for the first-stage: (1) grades 7-8 (early), (2) grades 9-10 (on-time) and (3) grades 11-12 (late). 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 as in equation (1) with a few exceptions. First, while $Z_{ig(t-k)}$ 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), $\tilde{Z}_{ig(t-k)}$ 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 the students in the analytic sample.[11] By relying on lagged test scores that are realized prior to the grade range of algebra-I course taking, we avoid the possibility of controlling for concurrent outcomes in the course-timing equations. Given the change in the lagged score vector, the vectors $\widetilde{M}_{ig(t-k)}$ and $\widetilde{D}_{dt}$ are correspondingly re-defined. X_{idgt} and T_t are defined as in equation (1).[12]

P_{dt} is the vector of instruments in equation (3). It 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. The

[11] Again, recall that students who take the EOC prior to grade-7 are excluded from our analysis (≈0.1 percent of the students in Missouri – see Table 2).

[12] Given that students who have previously taken the EOC are not included in the estimation of equations (3) and (4), X_{idgt} excludes the indicator for re-taking the exam.

instruments are conceptually similar to those used by Clotfelter, Ladd and Vigdor (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 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, Ladd and Vigdor (2012a).[13]

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 grades 9 and 10 (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 – and concerns related to their failure – 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. The point estimate for taking the exam after grade-10 also indicates a sizeable, negative effect on performance, but it is imprecisely estimated. For accelerated algebra, our estimates in Table 3 are consistent in sign, although not necessarily in magnitude, with similar estimates from Clotfelter, Ladd and Vigdor (2012a, 2012b). 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, likely due to selection issues.

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.

[13] All standard errors in Table 3 are clustered at the district level and calculated to be robust in the presence of heteroskedasticity.

In general terms, 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. Based on our analysis, we use equation (5) to impose performance penalties on districts for students who take the exam in grades 7-8. Districts are not penalized for students who take the exam on-time (grades 9-10) or late (grades 11-12). Although the point estimate for late test taking is large, we carry through our procedure without any late-taking penalty given the imprecision with which the effect is estimated. We return to this issue in Section 5.3.

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 course-timing effects:[14]

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

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 on-time (in grades 9-10). The low correlation (not statistically significant) is a result of the fact that the unadjusted estimates (a) remove

[14] 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.

the effects of cross-grade student sorting on district performance (via the grade-by-grade estimation procedure) but (b) do not account for the effects of course-timing policies. 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 on time. The black circles and squares indicate cases where districts change status in terms of whether they are identified as being statistically different from average, with black circles indicating a decline in status and black squares indicating an improvement. Districts that pursue more effective course-timing policies improve relative to their peers after the adjustment.[15]

Finally, we briefly note an operational issue with regard to implementing the course-timing adjustment. Our preferred approach is to use adjustment parameters estimated with data that pre-dates the evaluation system. 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.[16]

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 up to this point does not account for differences in student aptitude, etc., that might justify different course-taking patterns for some students. For example, high ability students who are 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.

[15] There are alternative ways to illustrate this information. For example, in unreported results we consider a scenario where the state would like to identify the top and bottom 10 percent of districts in terms of EOC performance. Moving from the case where we do not account for course timing to the case where we do account for course timing (from the left to right panel in Figure 1) results in 5 of the 51 districts in the original top 10 percent and 7 of the 50 districts in the original bottom 10 percent being replaced.

[16] An alternative concern is that the fixed course-timing adjustments could become biased over time, as they would not account for changes in the testing instrument, demographics, instructional quality, etc. at different grade-levels. If this is a concern these parameters could be periodically updated, perhaps with some smoothing, with the tradeoff that the updated parameters would potentially be influenced by districts' behavioral responses to the evaluation system.

Clotfelter, Ladd and Vigdor (2012b) provide direct evidence on the effects of accelerated algebra-I course-taking on future math course-taking across the achievement distribution. They show that while all students 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 take algebra-I early. But top-quintile students are the only students for whom early algebra-I course-taking positively affects future course-taking behavior – students in the bottom three achievement quintiles are less likely to pass geometry by grade-11 if they take algebra-I early, and students in the fourth quintile are no more or less likely to pass geometry by grade-11.

Based on the evidence from Clotfelter, Ladd and Vigdor (2012b), we build flexibility into our approach by allowing for "penalty forgiveness" for some students. Specifically, 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. Hence, districts receive no penalty for letting some high-performing students take the exam early.

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 evaluation 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, seven districts see an improvement in their status while another eleven see their status change for the worse. The reason for these changes is apparent in Table 4, which shows the percentage of students receiving accelerated course-taking penalties with and without penalty forgiveness for the seven districts that experience an improvement in status.[17] As can be seen in the table, a large portion

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

of students in these districts receive penalty forgiveness. In fact, the average district in Table 4 went from having 29.1 to 7.5 percent of its students receiving a course-timing penalty, a 74.2 percent decline.[18]

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). Table 5 reports results from the first-stage regressions and establishes instrument relevance. 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-level grade 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 observed 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.

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

[18] 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.

likely to push for earlier algebra-I course taking.[19] Other stories can be told. 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 our application.[20]

On point (1), if we operate under the assumption that there is some bias in the IV estimates, it is worthwhile to consider its likely 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 IV models, it may be that there are similar unobserved differences between districts with different course-timing policies (e.g., see Altonji, Elder and Taber, 2005). If this were the case, high-achieving districts would be more likely to have higher conditional EOC performance and would also be more likely to accelerate algebra-I course taking. Noting that available evidence shows that the causal effect of accelerating algebra-I course taking on achievement is negative (Clotfelter, Ladd and Vigdor, 2012a, 2012b), any such positive bias would imply that the "course-timing penalty" terms that we apply in equation (5) are too small in magnitude (but still signed properly).

[19] 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 vary substantially between districts 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. A related issue is that teacher quality might be systematically higher in some grades relative to others in Missouri – for example, in grades 9 and 10. If this were the case, then differences in teacher quality across grades would be a mechanism for the course-timing effects we estimate. However, the likelihood that our findings are strongly driven by cross-grade differences in teacher quality seems low given our OLS estimates and the corroborative findings from Clotfelter, Ladd and Vigdor (2012a, 2012b), with their 2012a study being particularly compelling because it relies on an abrupt policy change for identification (in the case of an abrupt policy change it is unlikely that there will be a wholesale change in personnel, but rather a change in which teachers teach in which grades).

[20] 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.

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, regarding whether outside evidence is available to support the notion that the instruments are providing useful information, estimates from Clotfelter, Ladd and Vigdor (2012a, 2012b) can be compared to our estimates in Table 3, at least for accelerated algebra-I course taking. Our estimate of the effect of accelerating algebra-I to grades 7 and 8 relative to grades 9 and 10, -0.178 as reported in Table 3, is roughly one-half the size of analogous estimates reported in their studies but still represents a sizeable, negative effect.

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 might be explained by the fact that Clotfelter, Ladd and Vigdor (2012a, 2012b) aim to identify the effects of sharp changes in course-taking policies within school districts that occur over short periods of time, while our model is designed to capture the effects of "steady-state" differences in algebra-I course-timing policies across districts. This is important because a sharp policy change to accelerate algebra-I course taking may not have the same effect as a long-term accelerated algebra-I policy. In the latter case, districts may be better able to tailor lead-in courses to accommodate students taking algebra-I in grade-8, whereas a sharp policy change will be less accommodating in this regard (this caveat to their findings is noted by Clotfelter, Ladd and Vigdor). Although we cannot precisely resolve the discrepancy in our estimates and those from Clotfelter, Ladd and Vigdor (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.[21]

5. Diagnostic Value of the Model and Other Concerns

5.1 Diagnostic Value of our Approach

Although accounting for the effects that district-level grade placement policies have on student achievement is our primary motivation in this work, the multi-part structure of the approach we outline above also provides valuable diagnostic information that can be used by both policymakers and practitioners to improve student outcomes. To illustrate, consider a district that has implemented a policy whereby most of its students take algebra-I in grade-8. Suppose that instructional quality in the 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 or 10, all else equal). The high quality of instruction delivered by the district is captured by the unadjusted district-effect estimates from equation (2). Districts that promote effective instructional strategies (e.g. better teachers, improved curricula, enhanced tutoring services) can be identified using the output from equation (2) and serve as models for other districts in the state in this regard.

But despite its strength in instruction, this hypothetical district's grade-8 policy is 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

[21] 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 (although re-takers are excluded from the estimation of equations (3) and (4)). This is in contrast to the method used in Clotfelter, Ladd and Vigdor (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 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.

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

5.2 *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. Because the first part of the model estimates the instructional quality measures separately by grade (equations (1) and (2)), it forces the comparisons to be between students who are taking the course in the same grade. This removes bias caused by course-timing issues that would be present in a model that pools algebra-I EOC test takers across grades. Thus, estimated teacher effects from the grade-specific first step of our approach are the natural choice to use for the foundation of teacher-level performance measures.[22]

Turning to the grade-placement policies that the second part of the model is designed to address, these are largely out of the control of individual teachers, and as such, teacher-level value-added measures should not be subject to the course-timing penalties. 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 might 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 the larger district policies.

5.3 *Late Course Taking and Incentivizing Enrollment in Courses Linked to EOCs*

[22] That said, substantial challenges remain in developing teacher-level performance measures based on student EOC exam scores beyond simply accounting for course-timing effects. A central concern is how to deal with more complex student tracking (particularly within-grade), an issue discussed in recent studies by Anderson and Harris (2013) and Jackson (forthcoming).

As discussed previously, the estimated course-timing effects in Table 3 are suggestive of an educationally-meaningful negative impact of taking algebra-I after grade-10. However, the lack of statistical power resulting from the clustering structure in the data is such that our estimate for late-takers is imprecise and cannot be statistically distinguished from zero. We have elected not to assign a late-taking penalty to districts in the above-described evaluation procedure for this reason.

Given that the variation in course-timing policies used for identification in our models occurs entirely at the district level, we are skeptical that a precise estimate of the late-taking effect can be obtained with data from a single state. But our findings, while only suggestive, provide ample motivation for future research aimed at providing a more precise estimate of the effect of late course-taking on algebra-I EOC performance. If our suggestive result is ultimately confirmed, late-taking penalties could be constructed analogously to the early-taking penalties we describe above in order to dissuade districts from allowing large fractions of students to take the algebra-I EOC in later grades.[23]

A related issue is that, unlike standardized exams in grades 3-8, students need not take courses that are tied to EOCs. Whether this is of concern depends on the specific course and students' educational plans, but by making EOC performance a part of the larger evaluation system, it is important to be cognizant of the potential to create incentives that inadvertently encourage districts to keep some students from taking specific courses. This issue is particularly important if penalties are imposed on students taking courses later than it is empirically determined to be optimal for most students.

Fortunately, the model presented above is flexible enough to directly incorporate students who never take the EOC. Generally, the first instinct in such situations is to use a predictive model 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

[23] While policymakers await stronger evidence on this issue, they may still choose to develop incentives for school districts to discourage late algebra-I course taking. Kane (2013) provides a rationale for why this might occur. In short, the issue is that the standard hypothesis testing framework is not well-suited for some policy decisions.

scores from their predictions. Hence, by definition, any student with a score imputed in this manner would have a zero residual, and the inclusion of these students 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.[24] For EOCs that are required for all students (like algebra-I in Missouri), the penalty would be assigned to any student who never takes the test. For non-required courses, this method of dealing with students who never take the exam is conceptually more difficult. One possibility would be to empirically determine a likelihood of success in the course for each student based on prior achievement, and then exclude students below some threshold value from the model without penalty.

6. Conclusion

The increased availability of EOC assessments in higher grades 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, using models that have been 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, Ladd and Vigdor, 2012a, 2012b). The procedure developed in this paper attempts to deal with these issues within an evaluation framework.

The first step in our approach tackles the cross-grade student sorting issue – ignoring the course-timing policy issue – and produces district performance measures of “instructional effectiveness” that are conditional on the grade levels in which students take EOCs. The second step explicitly incorporates the

[24] A similar strategy is applied by Clotfelter, Ladd and Vigdor (2012a, 2012b) in assigning exam scores for students who never take the EOC.

effects of course-timing policies to provide a direct accounting for the role that these policies play in determining student achievement. In the third step, we introduce flexibility to facilitate district discretion in terms of allowing some students to take courses in grade levels that our models indicate are suboptimal for most students.

The end result is a district performance measure that is informative about efficacy and 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. Forthcoming. Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*.

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 8^{th} 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. Forthcoming. Selecting Growth Measures for Use in School Evaluation Systems: Should Proportionality Matter? *Educational Policy*.

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. 2013. Presumed Averageness: The Mis-Application of Classical Hypothesis Testing in Education. The Brown Center Chalkboard at Brookings (12.04.2013).

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).

Konstantopoulos, Spyros. 2006. Trends of School Effects on Student Achievement: Evidence from NLS: 72, HSB: 82, NELS: 92. *Teachers College Record* 108(12), 2550-2581.

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.

Tables

Table 1. Summary Statistics.

Analytic Sample Size	
Number of Districts	505
Number of Schools	874
Number of Student/Year Observations	138,142
Student Characteristics	
Percent Female	49.3%
Percent Free/Reduced Price Lunch Eligible	43.6
Percent Minority	22.2
Percent English as a Second Language	2.1
Percent with an Individualized Education Plan	10.2
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 Algebra-I EOC in 2012 and 2013.

	Grade Level of EOC							
All Students	Missing	< 7	7	8	9	10	11	12
No. of Students	58	123	1499	28919	65142	23654	9951	8977
Percent of Students	0.0	0.1	1.1	20.9	47.1	17.1	7.2	6.5
First-time Test Takers	Missing	< 7	7	8	9	10	11	12
No. of Students	54	123	1498	28884	63461	21865	7966	7974
Percent of Students	0.0	0.1	1.1	21.9	48.1	16.6	6.0	6.1

Note: The distribution is reported for test takers in 2012 and 2013 combined. The year-specific distributions are substantively similar.

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

	OLS	IV
Grades 7 and 8	0.180**	-0.178*
	(0.025)	(0.085)
Grades 11 and 12	-0.370**	-0.147
	(0.028)	(0.137)
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: Standard errors are clustered at the district level.
** represents statistical significance at the 0.01 level.
* represents statistical significance at the 0.05 level.

Table 4. The Effect of Accelerated-Algebra Penalty Forgiveness on Student Residuals (Districts with Significantly Improved Effect Estimates).

	Percentage of Student Residuals Receiving an Accelerated Course-Taking Penalty	
	Before Forgiveness	After Forgiveness
District 1	19.6	1.0
District 2	27.5	5.4
District 3	36.2	12.1
District 4	26.3	2.1
District 5	34.2	12.1
District 6	35.4	11.3
District 7	24.6	8.8
Simple Average	29.1	7.5

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

	Dependent Variables – Student took the EOC in:	
	Grades 7/8	Grades 11/12
Instruments		
Share in District Taking Exam in Grades 7 and 8	**0.010****	0.000
	(0.000)	(0.000)
Share in District Taking Exam in Grades 11 and 12	0.000	**0.008****
	(0.000)	**(0.000)**
F-Statistic for Instruments	1223**	884**
Other 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: Standard errors are clustered at the district 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 = 74)					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	0.428	0.439	0.157	0.337	0.636	0.046	0.256	-0.106	0.054	0.204
Avg. Grade-6 MAP Com Arts Score	0.317	0.407	0.008	0.264	0.641	0.019	0.218	-0.094	0.032	0.158
Percent Female	49.0	18.6	44.4	50.0	56.6	48.8	6.6	46.2	49.0	52.1
Percent F/RL	43.9	25.1	25.0	40.8	58.3	50.0	16.6	39.6	50.0	59.5
Percent Minority	14.0	27.5	0.0	1.8	11.8	11.1	19.8	1.6	4.4	10.0
Percent of Students with an IEP	5.7	7.0	0.0	2.4	11.8	10.5	7.2	6.4	9.8	13.1
Percent ESL	0.8	4.0	0.0	0.0	0.0	1.2	3.7	0.0	0.0	0.8
Percent Mobile	2.6	3.9	0.0	0.0	4.3	4.8	4.5	2.4	4.0	6.1

Appendix A - Controlling for Incomplete MAP Score Histories

The inclusion of three years of lagged scores in two subjects in the models used 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. The general method used to control for this issue parallels that in Ehlert et al. (forthcoming) – that is, missing exam scores are set to zero (the standardized mean) and indicator variables are initialized for the missing scores. However, the length of the lagged score vector along with the fact that some algebra-I EOC takers have no prior MAP records presents complications.

By way of comparison, in the model presented in Ehlert et al. (forthcoming), the lagged-score vector is shorter and students are *required* to have a same-subject lagged exam score to be included in the analytic sample. Hence, there is at most one missing lagged score per student. But in our application, to control for every possible combination of missing lagged scores would require $2^6 = 64$ indicator variables, many of which could not be included in every grade-specific regression because no students in the given grade would have that missing score combination. In addition, some students have no prior MAP scores at all.

To simplify and improve the tractability of our models, we create only four indicator variables for missing lagged-score data (included in the vector $M_{ig(t-k)}$) – one to indicate that 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 she was missing the lag 2 *and* lag 3 exam scores (both subjects), and a fourth to indicate any other missing lagged-score combination. The first three of these indicator 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. In contrast, the last 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 distributions of the indicator variables that we create, by grade, are presented in Table A.1.

Table A.1. Missing Test Score Percentages by Grade and Indicator-Group.

	Grade					
	7	8	9	10	11	12
No Missing Scores	88.1%	91.5%	86.5%	83.5%	78.9%	61.7%
Missing MAP Lag 1, 2, 3	2.9	2.0	5.7	6.9	9.6	17.5
Missing MAP Lag 2, 3	4.5	2.4	2.5	2.6	2.4	4.0
Missing MAP Lag 3	2.8	3.1	2.9	3.2	3.7	7.9
Missing MAP Lag - Other	1.7	1.0	2.5	3.8	5.3	8.9
Total N	1499	28919	65142	23654	9951	8977