

CALDER



NATIONAL
CENTER for ANALYSIS of LONGITUDINAL DATA in EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

Urban Institute



A program of research by the Urban Institute with Duke University, Stanford University, University of Florida, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington

*Cramming:
The Effects of School
Accountability
on College-Bound
Students*

COLLEEN DONOVAN

DAVID FIGLIO

MARK RUSH

**Cramming:
The Effects of School Accountability on College-Bound Students**

Colleen Donovan
University of California-Berkeley

David Figlio
University of Florida and National Bureau of Economic Research

Mark Rush
University of Florida

Current draft: September 2006

Corresponding author: David Figlio, Department of Economics, University of Florida, Gainesville, FL 32611-7140, figlio@ufl.edu. The authors are grateful to the National Institutes of Child Health and Human Development and the National Science Foundation for financial support, and to an anonymous university for access to its admission, registrar and course download records. This research is also part of the activities of the National Center for the Analysis of Longitudinal Data in Education Research (CALDER). CALDER is supported by Grant R305A060018 to the Urban Institute from the Institute of Education Sciences, U.S. Department of Education. Thanks to Chris Avery, Sunny Ladd, Kim Rueben, and seminar participants at the London School of Economics, Russell Sage Foundation, Texas A&M University, University of California-Berkeley, University of Florida, University of Kentucky, University of Miami, University of Oxford, University of Virginia, NBER Education Program Meeting, and (UK) Institute of Fiscal Studies as well as the SEA and NTA meetings for helpful comments. CALDER working papers have not gone through final formal review and should be cited as working papers. They are intended to encourage discussion and suggestions for revision before final publication. Any opinions, findings, and conclusions expressed in these papers are those of the authors and do not necessarily reflect the views of the Urban Institute, the Institute of Education Sciences, the U.S. Department of Education, or any other sponsors of the research. All errors or omissions are our own. This paper is also available as an NBER Working Paper at <http://papers.nber.org/papers/w12628>.

Cramming: The Effects of School Accountability on College-Bound Students

Introduction

School accountability – the practice of evaluating schools on the basis of the observed performance of students, and rewarding and punishing schools according to these evaluations – is ubiquitous in the world today, with nations on every continent experimenting with such policies. In the United States, many states had implemented school accountability policies well in advance of the federal No Child Left Behind Act of 2001, and some states' experiments with accountability are currently in their second decade. The stated purpose behind school accountability plans is to encourage schools to more effectively educate students, and especially to ensure that all students are gaining ground toward subject-matter proficiency.

Numerous studies, such as Figlio and Rouse (2006) and Jacob (2005), indicate that low-performing schools facing accountability pressure have improved their students' test scores, not just on the high-stakes tests but also on tests that are directly comparable to schools in other states. Aggregate studies, such as Carnoy and Loeb (2002) and Hanushek and Raymond (2005), find that states that adopted high-stakes school accountability plans earlier than others have experienced greater growth in average test scores as compared with late-adopting states. The general conclusion reached by these studies is that school accountability has had a moderately large positive effect on average student outcomes in lower level schools.¹ This research, however, focuses on the effects

¹ Other researchers have studied the incentives embedded within school accountability plans. See, for example, Boyd et al. (2002); Cullen and Reback (2002); Deere and Strayer (2001); Figlio (2006); Figlio and Getzler, forthcoming; Figlio and Winicki (2005); or Jacob (2005).

of school accountability plans on the standardized test scores of average students or low-performing students. To date there has been no published research investigating the effects of these plans on the other end of the academic distribution – high-performing students who would almost surely have attained proficiency in the absence of school accountability plans. Even though these students are not the immediate focus of accountability plans, the effect these plans have on these students is vital to the nation because it is these students who will become the next generation of scientists and leaders.

It is not obvious how school accountability should affect these college-bound students. There are at least two avenues through which accountability rules and tests can influence students. First, accountability rules can affect students' study skills and study methods. Second, these rules also can change students' knowledge and what they learn:

- If schools pay less attention to higher-performing students, these students might learn different study methods, such as cramming for exams instead of practicing deliberate and continuous study habits. In addition, if the rewards and sanctions for school accountability plans give schools the incentive to substitute attention from more successful students to less successful students, the more successful students may learn less or learn differently as a consequence. However, school accountability plans might lead to improved study habits if the accountability systems induce schools to concentrate on material throughout the school year.²

² We suspect that schools may face different incentives to substitute attention from more successful students to less successful students depending on their socio-economic status. In high socio-economic status schools, college-bound students are prevalent and schools are unlikely to shift attention in meaningful ways from these students. But in low socio-economic status schools, college-bound students may comprise a sufficiently small fraction of the population that they can receive less attention without controversy. Given that high-achieving students in low-income schools are likely to receive very high

- If school accountability plans and the tests they employ to evaluate schools are challenging and based on rigorous standards, it could be expected that school accountability plans would have a positive impact on students' learning.

However, if schools adhere strictly to the material most likely to be on the high-stakes test and the focus becomes more fully on high-stakes tests as ends in themselves rather than indicators of subject-matter knowledge, school accountability plans could have the effect of dulling student and teacher creativity and limiting what students learn.

Both of these potential consequences may manifest themselves in student performance in college, where study habits and course performance might be improved or diminished due to school accountability plans. So, ultimately the important effect of accountability plans on college-bound students is an empirical issue, one which has yet to be addressed.

This paper starts to fill this void in the literature. We exploit data from a state that changed the basis of its accountability system in 1999. This change directly influenced a large number of schools that immediately either transitioned from being threatened with sanctions to not being threatened at all, or vice versa. Using this identification strategy, we can measure the impact on students of the school they attend either becoming threatened or becoming less threatened. Because the bases of these two accountability systems were so different, we can employ a variety of falsification tests to distinguish the effects of school accountability sanctions from the effects of the variables underlying these sanctions. We compare students in the high school classes of 2000 and 2001 – the

grades anyway, a reduction of attention might not provoke substantial parental awareness and protest in these schools.

last high school class affected directly by the old accountability regime (because the high-stakes test grade is the tenth grade) and the first high school class affected directly by the new accountability regime. Because of the possibility for spillovers – even though ninth grade is not a high-stakes grade, for instance, schools may have begun gearing up for the tenth grade test, and curricula and teaching styles might not immediately change even when accountability pressures change – this empirical approach represents a strong test of the effects of accountability sanction threats on student university performance.

In order to implement this identification strategy, we utilize a remarkable dataset from a large selective public university in the state in question. (The identity of the university and the state in question are suppressed to preserve the anonymity of the institution that provided us with the data.) The university in question provided us with data that allow us to investigate the impact of attending a high school threatened with accountability sanctions on student study habits. For six large classes in which the university’s students watch lectures online rather than in a traditional classroom setting, for the Spring 2003 semester the university provided us with the precise timing of when students first downloaded each lecture.³ We can therefore measure the degree to which students “cram” for exams – delaying studying for a course until the time prior to the test, rather than keeping up with the course throughout the semester – and relate this cramming behavior to the threat of accountability sanctions faced by the student’s high school during the student’s tenth grade. While our measure of cramming is only one measure of student study habits, it has the advantage that it is directly observed and measured,

³ During the semester in question, the download dates were collected by the university in the normal course of events. The students were unaware that these data were being collected and so had no incentive to alter their behavior.

meaning that we do not need to rely on notoriously unreliable student self-reports for outcome data.

The university in question also provided us with information from admissions records (the student's high school attended, grade point average, college entrance exam scores, race/ethnicity, and sex) as well as transcript data when in college (courses taken and grades earned). These data allow us to identify the effects of attending a high school threatened with accountability sanctions on a student's subsequent collegiate success.

We find that school accountability plans have the potential to substantially affect high-achieving students' study habits and performance in college. First, we observe that school accountability plans apparently alter student study habits in college. Students who attended schools threatened with sanctions under *either* school accountability system displayed a dramatically increased likelihood of cramming once they attended the university in question. Although we lack the ability to explore whether cramming behavior is directly responsible for the relationship between accountability sanction threats and student collegiate performance across a wide range of subjects, we can show that the more students in an online-viewed course crammed, the worse they performed in *those* classes. Second, we find that when looking at a range of college courses, school accountability plans affect student course performance at the university in question – but not in uniform ways. While accountability threats from the first accountability system appears to have reduced student performance in college in general, the new accountability system did not have the same broad negative effect on collegiate performance. Instead, this introduction had a positive estimated effect on mathematics

performance and other technical subjects, and no large negative effects on other subjects. We suspect this is due to the fact that the new system introduced in 1999 was strongly standards-based, while the disbanded system was not. Therefore, the evidence suggests that poor student study habits may be partially to blame for the low course performance of students affected by the first accountability system, and that the second accountability system, while apparently improving students' performance in college courses, may have been more effective in improving student collegiate performance were it not for the poor study habits that accompanied the system's introduction.

Two accountability systems

In 1999, the state in question dramatically changed its school accountability system. Prior to 1999, schools were graded according to aggregate student performance on a low-level test of basic skills, that is, a minimum standard of proficiency that was not linked to instructional or curricular standards. Beginning in 1999, the state changed its system of evaluating schools to be based on meeting proficiency targets on a new standards-based test. The new test discriminated performance at a much higher level, and was aligned to a far greater extent to the subject matter that state educators expected students to master. While the majority of students statewide scored in the highest category of the test used for pre-1999 accountability, only a small fraction of students scored in the highest category of the test used for post-1999 accountability. The result is that the newer accountability system was likely to affect directly the measured learning of *all* students, including four-year college-bound students, in a high school, while the older accountability system was less likely to affect directly the high-performing students' measured academic skills. Put differently, because high-achieving students were more or

less automatically going to do well on the low-level test they faced, schools had more of an incentive under the older accountability regime to substitute attention away from college-bound students than they did under the new accountability regime. And since college-bound students' performance mattered more in the new accountability regime, there was a larger incentive for schools to build the human capital of these students under the new regime.

But while the two accountability systems differed considerably in their potential direct effects on measured academic performance in high school, both accountability systems had the potential to influence student study skills and the ways in which students approach learning. Students in schools facing accountability pressure might learn from their school that test performance is the goal in itself, and so might change their approach to education from deliberate study for the sake of learning to concentrated study, that is, cramming, for the sake of test performance.⁴ Both accountability regimes could potentially help to instill this new study ethic in high-performing students. It is not clear which accountability system would do this the most. In the earlier accountability system, schools had a greater incentive to generally ignore high-achieving students and prepare these students for the tests, if at all, in a very concentrated period of time which could well lead these students to learn to cram as a result. On the other hand, in the later accountability system, the stakes for schools were higher and the concentration on test results as a major outcome of interest increased. The increased importance of the test

⁴ Anecdotal evidence abounds regarding these changing attitudes. Schools threatened with accountability sanctions often offer parties to reward students for high aggregate test performance, hold pep rallies to gear students up for the accountability test, and teach test-specific preparation skills to students and their parents.

results could induce schools to prepare the students more intensively as the test drew closer and less intensively when the test was either farther away or over, thereby teaching students to cram.

Most important for the purposes of this study is that the changing accountability regimes dramatically affected which schools were threatened with accountability sanctions. Some schools immediately transitioned from being threatened to no longer being threatened, while other schools were affected in the reverse. Table 1 presents evidence of the transitions between accountability systems. Specifically, because we will be employing high school fixed effects in our regression models, we describe the transitions only of high schools for which students in both the high school class of 2000 and the high school class of 2001 are represented in our data. We observe that 22 percent of schools changed accountability status from one accountability regime to the other, with the majority of these schools becoming less threatened once the accountability regime shifted. Nine percent of schools were threatened with sanctions in both accountability regimes, while 69 percent of schools were threatened in neither accountability regime. Our analysis will focus on the bottom-left and top-right cells of Table 1 – the schools that transitioned from threatened to not threatened, and vice versa.

Table 2 describes the attributes of the high-achieving students in our study, based on the accountability status of their high school. We observe that the schools that transitioned from unthreatened to threatened and vice versa were somewhat different from one another, and also different from the schools whose accountability status did not change over this time period. Specifically, the high schools that were not threatened under the

old accountability regime but became threatened under the new accountability regime sent students to the university in question who were more likely to be male, black and Hispanic and who had modestly lower college entrance exam scores and slightly lower adjusted high school grade point averages than did the high schools whose accountability status changed from threatened to unthreatened or whose accountability status did not change. The fact that these differences in student characteristics across schools are apparent underscores the importance of estimating high school fixed effects models, where we compare students from one graduating class in a given high school to students in the next graduating class from the *same* high school.

Table 2 also compares one student attribute – SAT scores – within the same schools across graduating classes. As can be seen, SAT scores improved for enrollees from the high school class of 2001 vis-à-vis enrollees from the high school class of 2000 for all types of schools. However, we observe that the group of schools that experienced a transition from unthreatened to threatened as the accountability regime changed saw the greatest improvement in students sent to the university, at least according to this metric. This finding underscores the necessity of controlling for student background characteristics in the school fixed effects analyses that follow.

Measuring study skills in college

The first purpose of this paper is to measure the effects of school accountability sanction threats on the development of study skills. While it is impossible to directly measure study skills, we can measure an indicator of these skills. The university that provided us with the data had six classes for which the professors delivered lectures and the students

watched the lectures on-line. All of these classes are required for at least one major and several of them are required for many different majors across the university. The students in these majors must take these on-line classes. For these six on-line classes, for the Spring 2003 semester, the university recorded not only the date of download for each online-viewed class, but also the specific lecture downloaded. We can therefore identify the degree to which students delay watching lectures, which is one measure of cramming for the exam and one potentially important study skill.

Our data consist of observations on 2,064 students who attended 249 public high schools in the state for which students in both the high school graduating classes of 2000 and 2001 are present in the data. Using these data, we first study the extent of cramming, its persistence across courses, and its effect on students' grades. Then we explore whether accountability sanctions can lead to cramming behavior.

Figure 1 illustrates definitively that cramming for exams occurs and is widespread in the online-viewed classes. Each of the six graphs presents time-series plots of the fraction of first-time downloads of lectures per day, per course. It is immediately apparent that the pattern of downloads exhibits large cycles, and the dates of the exams in each course are obvious by inspection. Table 3 presents similar data in tabular form. The second column shows that 16 percent of the lectures were first downloaded by students on the day they were delivered and another 20 percent of the lectures were first downloaded over the next two days. But one quarter of lectures delivered more than one week before the exam were first downloaded in the week prior to the exam and the same fraction is true for those lectures delivered more than two weeks before the exam. Looking more strictly at

last minute cramming, three percent of the lectures were downloaded for the first time on the day of the test, and another nine percent of the lectures were downloaded for the first time one or two days before the test (this is not counting lectures delivered close to the time of the test in question.) For empirical purposes, we define cramming as the fraction of lectures delivered seven or more days before an exam that were first downloaded within two days of an exam. Table 3 shows that by this definition, over 13 percent of lectures were crammed.

The argument that accountability sanction threatens teach cramming behavior requires that, for some students, cramming is a “trait” – an attribute of the student that persists and not simply the random occurrence within a specific class. Tables 4 and 5 present some evidence that this is the case. We observe (see Table 4) that some students consistently kept up with lectures while others consistently procrastinated. The second column shows that 31 percent of students were never observed downloading a lecture within two days of the lecture delivery and only 2.5 percent of students downloaded every lecture within two days of delivery. 57 percent of students downloaded fewer than one fifth of the lectures within two days of delivery. While few students waited until the last moment to download *every* lecture, a non-trivial fraction of students delayed downloads of many lectures. Column three shows that over 2 percent of students first viewed more than half of their lectures for the first time (!) within two days of an exam and over 10 percent first downloaded at least one fifth of their lectures within two days of an exam. The presence

of these patterns suggests that students tend either to keep up with the course or to delay viewing the lectures.⁵

Table 5 illustrates that students who tend to cram in one class tend to cram in all of their classes. This table presents a cross-tabulation of students' cramming behavior in two courses taken simultaneously, for the set of students for whom behavior in two courses is observed. The top panel investigates the percentage of lectures downloaded within two days of delivery. We observe that if a student in course #1 downloaded less than one third of lectures within two days of delivery, he or she was very likely to do the same (69 percent) in his or her other class, and only had a 15 percent chance of downloading more than two thirds of the lectures in the other class within two days of delivery. On the other hand, if the student in course #1 downloaded more than two thirds of lectures within two days of delivery, he or she had a 43 percent chance of downloading less than one third of lectures in his or her other class within two days, but had a 41 percent chance of downloading two thirds or more of lectures in his or her other class within two days of delivery. The bottom panel presents similar evidence for downloading classes within two days of an exam. In this panel, if a student downloaded zero percent of the lectures for the first time within two days of an exam for the first class, there was an 82 percent chance that the student also downloaded zero percent of lectures in the second course within two days of an exam and only a 5 percent chance of downloading more than one fifth of the lectures within two days of an exam. The lesson from these tables is that while

⁵ Table 4 shows the persistence of cramming for all students for all classes. There is also considerable within-student persistence within any given course. Students who tend to cram for the first exam in a course also tend to cram for the second exam in the course.

cramming in courses does not follow a one-to-one correspondence, there is a very high degree of within-student persistence in cramming across classes.

Table 6 shows that students who cram are somewhat more likely to be male, somewhat more likely to be Hispanic and Asian, and tend to have higher SAT scores than do students who don't cram. While high school grade point averages are slightly higher for cramming students as for those who do not cram, the differences are not great.

Given that cramming tends to persist for individual students, the next logical question is whether cramming has a deleterious effect on student outcomes in the class in which cramming occurs. Table 7 presents separate regressions of the relationship between cramming behavior (as measured by the fraction of classes delivered more than a week in advance of the test and first downloaded within two days of the exam) and a student's grade in the class, for each of the six online classes. We observe a consistent pattern between cramming and course performance: holding constant student attributes, students who cram more tend to perform at a lower level in the class. This correlation is significant at conventional levels in five of the six regressions. Table 7 also presents separate regressions for students who attended threatened schools in either accountability regime, and those who attended schools that were not threatened in either accountability regime. Though not all the coefficients attain conventional levels of statistical significance, the signs of all the estimated coefficients are negative so we again observe consistent evidence of a negative relationship between cramming behavior and student course grades, regardless of the nature of the high school that the student attended, and across courses. In sum, student cramming persists across time within a course, and

across courses, and is associated with reduced student course performance. Recall that if anything, cramming students tend to look better on observables (in terms of SAT scores and perhaps in terms of high school grade point average) than do students who cram less. So it is important to explore whether accountability sanction threats lead high schools to “teach” their high-performing, college-bound students to cram. We turn to this possibility next.

Do accountability sanction threats induce student cramming?

Table 8 presents estimates of the relationship between accountability sanction threats and student cramming behavior. Two measures of cramming are presented – the fraction of total lectures first downloaded within one week of the exam and the fraction of total lectures delivered a week or more before the exam and first downloaded within two days of the exam. We estimate the model:

$$(\text{Fraction of lectures downloaded})_{isc} = \nu_s + \alpha_c + \gamma X_{isc} + \beta(\text{Accountability threat})_{sc} + \mu_{isc},$$

where X is a vector of student background characteristics (race/ethnicity, sex, high school adjusted grade point average, and SAT scores), s represents the high school attended and c represents the student’s graduating class. Because outcomes might differ depending on which accountability regime is active at the relevant time for a given student i , we estimate two β coefficients – one for the estimated effect of the old accountability sanctions and one for the estimated effect of the new accountability sanctions. If the

accountability system induces increased cramming behavior, the estimated β coefficients on the presence of threatened sanctions will be positive.⁶

The results in Table 8 show that cramming behavior is associated with accountability sanction threats. When such a threat is removed, students attending a formerly-sanctioned school subsequently cram less in college than do their predecessors who attended the high school when the school was threatened. Conversely, when a new threat is introduced, students attending a newly-sanctioned school subsequently cram more in college than did their predecessors who attended the high school before it was threatened. The estimated effects of the new accountability system are greater in magnitude and statistical significance than are the estimated effects of the older accountability system. This finding is consistent with the suggestion that the new accountability system, with its increased emphasis on all students' test scores, may teach higher-achieving students to cram more than did the older accountability system, under which the higher-performing students could basically have been neglected.

To help insure the robustness of our results, we conducted two falsification exercises. The first such exercise, reported in the second column of Table 9, utilizes data from the graduation classes of 1999 and 2000, rather than 2000 and 2001, and assumes that the accountability system switch occurred a year before it actually did.⁷ We observe that there is no apparent relationship between cramming and changes in accountability regimes when the accountability regime did not actually change.

⁶ A full set of coefficient estimates is presented in Appendix Table 1.

⁷ The number of observations differs from the first and third column because the number of relevant schools and students is different in this exercise. The results of the falsification exercise are virtually unchanged were we to limit the sample of schools to be the same as in the first and third columns.

Our second falsification exercise involves using the same data as in the actual analysis, but applying the 1998 standards to test scores realized in 1999 and 1999 standards to test scores realized in 1998, for the purposes of calculating which schools would have been sanctioned if the systems, but not the test scores used to determine a school's accountability grades, had been swapped. Again, we find no apparent relationship between measured cramming and this other characterization of school test scores. These results suggest that it is the potential sanctions from the accountability system, and not the actual school performance underlying the accountability system, that apparently led to teaching cramming behavior to an affected school's students. Our results also indicate that schools apparently responded rapidly to the change in accountability regimes. The first year after a school is newly threatened with sanctions or newly relieved of potential sanctions, the empirical results demonstrate an immediate change in their college-bound students' study habits.

Effects of accountability on collegiate grades

Given that school accountability systems of both types apparently induced a persistent change in study behavior among college-bound students, it is interesting to determine the degree to which school accountability has affected these students' performances in college. The same forces that led to changes in outcomes could have made students more effective or less effective class-takers because, in addition to teaching college-bound students to cram, school accountability pressures may have either improved or harmed these students' human capital directly by increasing or decreasing their knowledge in particular fields. Given the differences in the accountability systems, we expect that the threat of sanctions with old system, under which high-performing students could be

neglected, would have more adverse net consequences for high-performing students than similar threats under the new system, which is based on meeting higher-level proficiency targets linked to the curriculum. To examine this issue, we focus on the introductory classes taken at the university. These courses, with the first digit being a “1” or a “2”, are typically taken by freshmen and sophomores at the university. This set of courses is much more inclusive than the six courses considered in the cramming portion of this paper. The university in question provided us with a broader dataset that we can use for this analysis. Between Fall 2002 and Fall 2004, for every student who ever took one of the six Internet courses we focused upon above, the university provided us with transcript data giving us the grades in all their courses and admission data.⁸ These students account for over three quarters of the students at the university, so selection into the sample does not seem to be at issue. We focus on all the introductory courses taken by these students, regardless of when they took the course. Our approach is to estimate variants of the following model:

$$(\text{Grade})_{isc} = \forall_s + \alpha_c + \gamma X_{isc} + \beta (\text{Accountability threat})_{sc} + \mu_{isc} ,$$

where all notation is as before and “Grade” is the student’s grade in the *i*th introductory course. In addition, we include course-specific fixed effects to account for the fact that student grades differ dramatically from one class to another at the university.

Table 9 presents evidence of the relationship between student grades and accountability sanction threats for students in introductory courses at the university in question. Each row represents a different set of courses, and the two estimated β coefficients for each

⁸ Unfortunately we had the download data we used for the analysis above for only the one semester.

regression are reported in the table.⁹ The first column of the table shows the estimated effects of the old accountability sanction threats, while the second column shows the estimated effects of the new accountability system. The outcome is a student's course grade (on a four-point scale, with a grade of "A" being 4 points).

The first row of Table 9 illustrates the estimated effects of school accountability regimes on student grades across all introductory subjects at the university. The results are starkly different from the cramming results earlier: While the old school accountability system, based on low-level tests of basic skills, appears to have a negative impact on the collegiate performance of college-bound students who attended potentially-sanctioned high schools, the new accountability system improved the performance of these students. Apparently the beneficial effects of the new system's increased emphasis on higher level skills overcame its tendency to teach students to cram so that the net effect on students' grades in introductory courses was positive. Moreover, the estimated effects are relatively large in addition to being statistically significant: Given that the mean college grade is a "B", an increase or decrease of 0.1 grade points is a large change.

Both accountability systems placed emphasis on student performance in mathematics, reading, and writing. While it is difficult to measure courses in reading comprehension in college, we can directly test for estimated effects of school accountability systems on mathematics and English composition. The second and third rows present the estimated effects of accountability on math classes (e.g., calculus, college algebra, introductory

⁹ The coefficient estimates on race/ethnicity, sex, graduation class, SAT scores, and high school grade point average are available upon request from the authors.

statistics, quantitative methods) and English composition classes. We observe that the estimated effects of the old accountability system are negative for these classes as well, but the estimated effects of the new accountability system are particularly strong and positive for the mathematics classes. If schools were responding to higher standards embedded in the new accountability system, it follows that mathematics training may be improved, an effect manifested in higher performance once at the university. On the other hand, there is no apparent positive effect of the new accountability system in English composition classes. This might be due to the possibility that the new accountability standards for writing focused on formulaic writing skills that reflect clarity of expression and exposition, but not higher-order writing skills.

The fourth row of Table 9 considers the relatively quantitative non-mathematics courses at the university in question. We define “relatively quantitative” as the courses where the average mathematics SAT scores of students taking the class exceeded the university average; these courses include introductory chemistry, physics, economics, and other similar classes. We find evidence of positive effects of the new accountability system – though half the size of the estimated effects on mathematics courses – as well as negative estimated effects of the older accountability system. This implies that the higher mathematics standards in the new accountability system appear to have positive spillovers to the other quantitative courses at the university.

Finally, the fifth row of Table 9 presents estimated effects of the two accountability standards on the relatively unquantitative courses at the university. We continue to find negative, significant estimated effects of the old accountability system. However, the

estimated effect of the new accountability system on the less quantitative system is statistically and economically around zero. This result indicates that the potential positive benefit of the new accountability system is concentrated in the more mathematical courses, suggesting a positive spillover from increased pressure of schools to perform, especially in mathematics.

While we believe that an accountability system based on poorly-defined standards and low-level skills, such as the system in place before 1999, leads to reduced student performance in college, an alternative explanation for the findings that the old accountability system had a generally negative consequence is that the schools that were threatened with sanctions have been improving over time – either due to the accountability system or for other reasons. There are at least three reasons to believe that this is not the case. First, we note that we are comparing two adjacent cohorts of students from the same school. Because these students came from the same schools, it is unlikely that there were massive changes in student composition from one year to the next. In addition, these students shared three of the four years of high school together. It therefore seems more likely that the findings presented herein are due to shocks to a school from one year to the next, rather than temporal trends. More compelling is the finding that schools which were threatened with sanctions under the former regime but no longer threatened with sanctions under the new regime did not improve their aggregate test scores (as reported by the state’s department of education) from one year to the next. This fact indicates that it was the regime change, rather than improvements due to incentives put in place under the old accountability regime, that led to the changes in observed behavior by students once they came to the university. Finally, we conducted

falsification exercises in which we feed each cohort's aggregate test scores through the other cohort's accountability regime; we find no evidence that changes in student attributes or test scores are responsible for the patterns described herein. Rather, the patterns appear due to the design of the accountability regimes themselves.

Conclusion

This paper presents the first evidence of the effects of school accountability systems on the long-term human capital development of high-performing, college-bound students. The results are mixed. On the one hand, the evidence is consistent that school accountability sanction threats are associated with changes in student study habits. Students who attended high schools that were threatened with accountability sanctions systematically study differently than did their school-mates who graduated from high school in an adjacent year but who were not directly affected by the school's threat of sanctions. Students from high schools threatened with sanctions postpone studying and cram more than students from the same high schools when the schools were not threatened with sanctions. The tendency toward increased cramming seems likely to result when threatened schools place a heavy emphasis on the outcomes of a specific test, rather than on overall learning outcomes over the entire year. We also find evidence that the induced increase in student cramming, *ceteris paribus*, leads to reduced course performance in the classes in which the student crams.

On the other hand, whether school accountability sanction threats lead to lower overall student course performance than would have occurred in the absence of such threats is not as clear. We observe that students attending high schools that were threatened with

sanctions under the state's old accountability regime tended to perform at a lower level in their coursework at the university, all else equal. But we do not observe this same pattern with regard to the new accountability regime. In fact, students attending high schools threatened with the new sanctions performed at a higher level in mathematics and quantitative courses than their school-mates had previously done, all else equal, but the differential performance is not observed in less quantitative courses. We suspect that the difference in outcomes across these accountability regimes – and across course subjects in the new accountability regime – is due to the standards-based nature of the new accountability regime and to the subjects emphasized in the new accountability system.

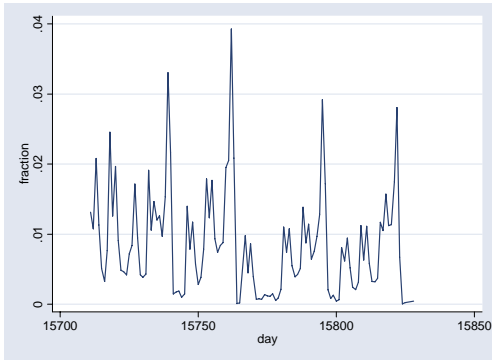
These results provide potential lessons for the design of accountability systems. The finding that students are apparently better prepared for college coursework in the specific subjects that are emphasized by the accountability system suggests that accountability systems that cover more subject areas may in turn produce better-prepared students. At the same time, the finding that students performed better when their schools were threatened with sanctions under a standards-based accountability system than under an accountability system with lower standards indicates that a system with high-quality tests closely aligned with rigorous standards is likely to yield better student preparation for higher education. Indeed, one result that seems unambiguous is that an accountability system based on poorly-defined standards and low-level skills unambiguously harms college-bound students from high schools that are threatened with sanctions. Students from these high schools not only cram but also do significantly worse in their introductory college classes.

To the extent to which school accountability systems encourages ignoring college-bound students, the systems may lead to the unintended consequence of adversely changing the ways in which high-performing students approach learning and scholarship. Hence the design of the accountability system is perhaps more important than previously thought. Accountability systems, such as Florida's system implemented in 2002, that reward and sanction schools based on gains in students' scores rather than levels of students' performance might present schools with less of an incentive to substitute effort away from higher-performing students toward lower-performing students. But this suggestion must remain speculation at the present because determining the impact of this new accountability system will require additional research.

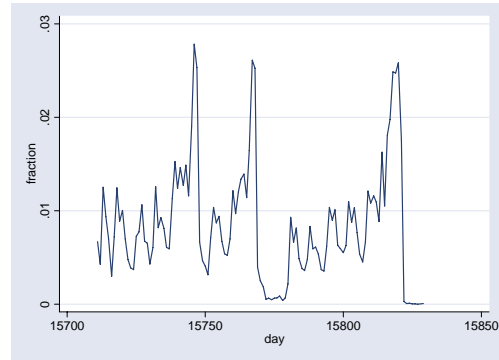
References

- Boyd, Don, Lankford, Hamilton, Loeb, Susanna, and James Wyckoff. 2002. "Do High-Stakes Tests Affect Teachers' Exit and Transfer Decisions? The Case of the 4th Grade Test in New York State." Unpublished Manuscript, Stanford Graduate School of Education.
- Carnoy, Martin and Susanna Loeb. "Does External Accountability Affect Student Outcomes? A Cross-State Analysis." *Educational Evaluation and Policy Analysis*. 24(4). Winter 2002. 305-331.
- Cullen, Julie Berry and Randall Reback. 2006. "Tinkering Toward Accolades: School Gaming under a Performance Accountability System." in Tim Gronberg and Dennis Jansen (eds.), *Advances in Microeconomics*. Amsterdam: Elsevier Science. Forthcoming, 2006.
- Deere, Donald and Wayne Strayer. 2001. "Putting Schools to the Test: School Accountability, Incentives, and Behavior." Unpublished Manuscript. Texas A&M University. March 2001.
- Figlio, David. 2006. "Testing, Crime and Punishment." National Bureau of Economic Research. *Journal of Public Economics*. Cambridge, MA.
- Figlio, David N. and Lawrence S. Getzler. 2006. "Accountability, Ability, and Disability: Gaming the System," in Tim Gronberg and Dennis Jansen (eds.), *Advances in Microeconomics*. Amsterdam: Elsevier Science. Forthcoming 2006.
- Figlio, David N. and Cecilia Elena Rouse. 2006. "Do Accountability and Voucher Threats Improve Low-Performing Schools?" *Journal of Public Economics*. 90(1-2). January 2006. 239-255.
- Figlio, David N. and Joshua Winicki. 2005. "Food for Thought: The Effects of School Accountability Plans on School Nutrition." *Journal of Public Economics*. Vol 89(2-3). February 2005. 381-394.
- Hanushek, Eric A. and Margaret E. Raymond. 2005. "The Effect of School Accountability Systems on the Level and Distribution of Student Achievement." *Journal of the European Economic Association*. Vol. 2(2-3). April-May 2005. 406-415.
- Jacob, Brian A. "Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools." *Journal of Public Economics*. Vol. 89(5-6). June 2005. 761-796.

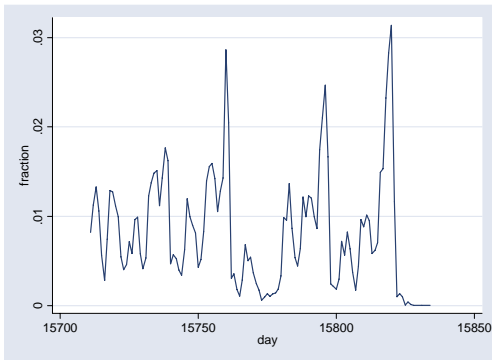
Figure 1: Time-series patterns of first time student downloads of lectures, six Internet courses, Spring 2003



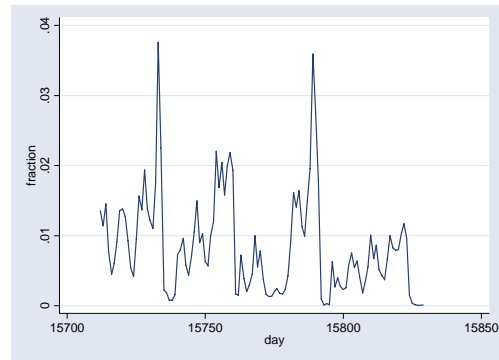
Course A with four midterm exams



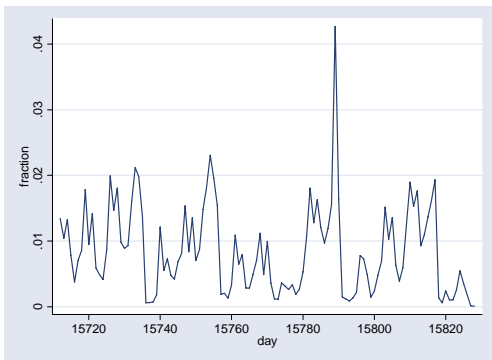
Course B with three midterm exams



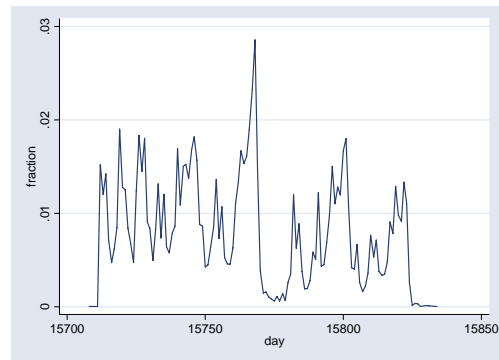
Course C with one quiz and three midterm exams



Course D with three midterms and optional final



Course E with four midterm exams



Course F with three midterm exams

Table 1: Transitions from one accountability system to another, high school classes of 2000 and 2001

		High school threatened with sanctions in high-stakes test year, class of 2001	
		No	Yes
High school threatened with sanctions in high-stakes test year, class of 2000	No	0.689	0.021
	Yes	0.198	0.091

Table 2: Attributes of students in study, by high school transition, high school classes of 2000 and 2001

In-state public school graduates only

	High school transitioned from unthreatened to threatened	High school transitioned from threatened to unthreatened	High school did not change status
Male	0.56	0.48	0.49
Black	0.13	0.08	0.06
Hispanic	0.21	0.08	0.09
Asian	0.12	0.13	0.08
SAT score	1196	1218	1239
Adjusted high school GPA	3.8	3.9	3.9
Comparing attributes of students in each of the two high school graduation cohorts			
SAT scores: high school class of 2000	1174	1208	1233
SAT scores: high school class of 2001	1208	1228	1245

Table 3: Patterns of first time student downloads of lectures, six Internet courses, Spring 2003

Timing of download	Fraction of first downloads	Fraction of first downloads, lecture delivered ≥ 7 days before test	Fraction of first downloads, lecture delivered ≥ 14 days before test
Day of lecture	0.159	0.140	0.137
1-2 days later	0.203	0.177	0.167
3-6 days later	0.253	0.235	0.199
≥ 7 days later, ≥ 7 days before test	0.168	0.196	0.271
≥ 7 days later, 3-6 days before test	0.099	0.116	0.123
≥ 7 days later, 1-2 days before test	0.086	0.100	0.084
≥ 7 days later, Day of test	0.032	0.037	0.019

Table 4: Cramming as a trait:
 Percentage of lectures downloaded immediately or at the last moment, by student

Frequency	Fraction of students: Percentage of lectures downloaded within 2 days of delivery	Fraction of students: Percentage of lectures downloaded 2 or fewer days before the exam (for lectures delivered ≥ 7 days before exam)
Never	0.311	0.533
0.1-9.9%	0.132	0.263
10-19.9%	0.131	0.102
20-29.9%	0.098	0.045
30-39.9%	0.067	0.021
40-49.9%	0.052	0.015
50-59.9%	0.051	0.010
60-69.9%	0.043	0.006
70-79.9%	0.037	0.002
80-89.9%	0.036	0.002
90-99.9%	0.017	0.001
Always	0.025	0.001

Table 5: Cramming as a trait:
Persistence of cramming behavior across classes, by student (row percentages)

Panel 1: Lectures downloaded 2 or fewer days after delivery (mean: 25.1%)

		Percentage of lectures downloaded 2 or fewer days after delivery, class #2		
		0-33.3%	33.4-66.6%	66.7-100%
Percentage of lectures downloaded 2 or fewer days after delivery, class #1	0-33.3%	0.691	0.159	0.151
	33.4-66.6%	0.565	0.272	0.164
	66.7-100%	0.428	0.162	0.410

Panel 2: Lectures downloaded 2 or fewer days prior to exam, for lectures delivered ≥ 7 days before exam (mean: 6.5%)

		Percentage of lectures downloaded 2 or fewer days prior to exam, class #2		
		0%	0.1-20%	>20%
Percentage of lectures downloaded 2 or fewer days prior to exam, class #1	0%	0.819	0.128	0.054
	0.1-20%	0.622	0.300	0.078
	>20%	0.516	0.258	0.226

Table 6: Attributes of students with different levels of cramming behavior

	Percentage of lectures downloaded 2 or fewer days before exam		
	0-33.3%	33.4-66.6%	66.7-100%
Male	0.54	0.55	0.62
Black	0.06	0.09	0.08
Hispanic	0.11	0.12	0.14
Asian	0.08	0.08	0.11
SAT score	1189	1210	1227
Adjusted high school GPA	3.84	3.83	3.89

Table 7: Cramming and student course performance:
 Student-level relationship between course grade and cramming behavior

	Coefficient on fraction of classes first downloaded 2 or fewer days before exam		
	All students in the class	Students who attended threatened schools in either accountability regime	Students who attended schools not threatened in either accountability regime
Course A	-1.875** (0.649)	-2.079** (1.065)	-1.441 (1.098)
Course B	-0.371** (0.141)	-0.084 (0.268)	-0.535** (0.165)
Course C	-1.558** (0.462)	-2.032* (0.766)	-1.090* (0.621)
Course D	-0.446 (0.458)	-1.919 (2.564)	-0.440 (0.429)
Course E	-0.810** (0.239)	-0.726* (0.474)	-0.943** (0.280)
Course F	-1.046** (0.258)	-0.868* (0.509)	-1.151** (0.294)

Note: Each cell represents a separate regression. Regressions also control for student sex, race/ethnicity, SAT scores, and high school grade point average. Coefficients marked * (**) are statistically significant at the ten (five) percent level.

Table 8: Estimated effects of changing accountability regimes on student cramming behavior:

High school fixed effects regressions; clustered standard errors in parentheses

Dependent variable: fraction of total classes first downloaded within week (two days) of exam

In-state public school graduates only

	Specification 1: Actual accountability sanctions included in model (graduation years 2000,2001)	Falsification test 1: Assume that accountability switch occurred one year before it did (graduation years 1999, 2000)	Falsification test 2:Apply 1998 standards to 1999 scores, and 1999 standards to 1998 scores (graduation years 2000, 2001)
Number of students	2064	2394	2064
Number of schools present in both years	249	265	249
Cramming measure: Fraction of total lectures downloaded within the week before the exam			
Estimated effect of old accountability sanction threat	0.030* (0.017)	0.007 (0.019)	-0.010 (0.027)
Estimated effect of new accountability sanction threat	0.078** (0.035)	-0.008 (0.026)	0.016 (0.032)
Cramming measure: Fraction of total lectures downloaded 2 or fewer days before exam			
Estimated effect of old accountability sanction threat	0.027* (0.015)	0.014 (0.017)	-0.016 (0.028)
Estimated effect of new accountability sanction threat	0.062** (0.032)	0.008 (0.025)	0.015 (0.026)

Note: Coefficients marked * (**) are statistically significant at the ten (five) percent level. Clustered standard errors are in parentheses. Regressions also include school fixed effects, course fixed effects, time effects, student race/ethnicity variables, student sex, student high school grade point average, and student SAT scores.

Table 9: Estimated effects of changing accountability regimes on student course outcomes:

High school fixed effects regressions; clustered standard errors in parentheses

Dependent variable: grade in introductory course

Introductory course	Estimated effect of old accountability sanction threat	Estimated effect of new accountability sanction threat
All introductory courses	-0.097** (0.030)	0.100** (0.051)
Mathematics courses	-0.149** (0.078)	0.277** (0.115)
English composition courses	-0.173** (0.061)	-0.048 (0.088)
Other quantitative courses	-0.139* (0.042)	0.118* (0.066)
Less quantitative courses	-0.064** (0.029)	0.044 (0.051)

Note: Coefficients marked * (**) are statistically significant at the ten (five) percent level. Clustered standard errors are in parentheses. Regressions also include school fixed effects, time effects, student race/ethnicity variables, student sex, student high school grade point average, and student SAT scores, as well as course fixed effects.

Appendix Table 1: Full set of coefficient estimates (except for course fixed effects and high school fixed effects): Table 8, specification 1, top panel

Variable	Coefficient (standard error)
Old accountability sanctions × post-change	0.030* (0.017)
New accountability sanctions × pre-change	0.078** (0.035)
Post-change	-0.007 (0.014)
Male	0.016 (0.012)
Black	0.002 (0.041)
White	-0.004 (0.033)
Asian	0.050 (0.037)
Hispanic	0.027 (0.038)
Have SAT score	-0.093 (0.070)
SAT score	0.00006 (0.00006)
Have high school GPA	-0.075 (0.065)
Adjusted high school GPA	0.021 (0.017)

Note: Coefficients marked * (**) are statistically significant at the ten (five) percent level. Clustered standard errors are in parentheses.

