QUANTITATIVE MEASURES OF ACHIEVEMENT GAINS AND VALUE-ADDED IN HIGHER EDUCATION: POSSIBILITIES AND LIMITATIONS IN THE STATE OF TEXAS

A Report Commissioned by the Texas Higher Education Coordinating Board

First Draft: October 27, 2008

This Draft: March 25, 2009

Jesse M. Cunha
Department of Economics
Stanford University
579 Serra Mall
Stanford, CA 94107
jcunha@stanford.edu

Darwin W. Miller, III
Department of Economics
Stanford University
579 Serra Mall
Stanford, CA 94107
millerdw@stanford.edu

The views expressed in this document are those of the authors, and do not reflect those of the state of Texas or the Texas Higher Education Coordinating Board. David Gardner was particularly helpful in framing the salient issues surrounding the application of value-added measures in higher education and securing data access. We also thank Eric Hanushek, Caroline Hoxby, Giacomo DiGiorgi, Doug Bernheim, Paco Martorell, Lee Holcombe, and Isaac McFarlin for their helpful comments and suggestions.

EXECUTIVE SUMMARY

This report has three goals. First, summarize issues surrounding the measurement of achievement gains in college. Second, discuss the possibilities and limitations of measuring these gains in Texas colleges using Texas Higher Education Coordinating Board data. Third, empirically estimate achievement gains in the most precise way possible using available data.

Discussion and Literature Review

We begin with a discussion of achievement gains in Higher Education, and emphasize its inherent multidimensionality. Colleges strive to impart different skills and knowledge to different students, and social as well as academic experiences are deemed valuable and desirable. As such, the entire package of knowledge and skills acquired in college is inherently unobservable to an empirical researcher and is not easily comparable across students and schools. We therefore study proxies for educational achievement that are available in the existing data.

We devote considerable attention to the merits of various proxies for achievement. Test scores, grades, and earnings can be used to measure what has been referred to in the literature as "value-added", while college completion and persistence rates inform us of other aspects of educational achievement. Test scores are an appealing proxy but pose significant methodological hurdles; furthermore they are not currently available. Earnings are particularly attractive due to their theoretical link to labor market productivity and are readily available. Persistence and completion rates are currently available and are indicative of certain aspects of achievement.

We turn to review several econometric models that have been used to estimate value-added in the academic literature. The choice of model depends crucially on the number of times the achievement measure is observed. If it is observed both before and after college, as is often the case with test scores, the calculation is easy: simply difference the two measures. If it is only observed once, as with earnings, we are faced with the dilemma of not knowing what level of educational attainment students had upon entering college, and hence how much

value was *added.* In this case, which unfortunately applies to higher education in Texas, one must resort to more sophisticated statistical techniques to estimate achievement gains.

As such, we discuss several statistical techniques that can be used when there is just one observation of achievement. Broadly speaking, these techniques fall into two categories: 1) techniques that attempt to control for all factors jointly determining college choice and the measure of attainment (e.g. controlling for SAT scores and high school achievement), and 2) those that employ additional behavioral assumptions to make causal inference. Methods of the former type suffer from the criticism that the researcher can never control for all confounding factors, such as parenting skills and innate student ability. Hence, we recommend that the state use methods falling under the latter category.

Empirical Analysis

We conclude that the model developed by Dale and Krueger (2002) is the most appropriate for determining the causal impact of Texas colleges on their students. Their framework explicitly models the college application and enrollment decision, making comparisons only between students who apply to and are accepted at the same set of colleges. If we believe that similar students have similar application and acceptance profiles, then this method provides us ideal comparison groups to identify the causal impact of colleges on achievement gains.

Using THECB data, we estimate the Dale and Krueger (2002) model with three different proxies for achievement: earnings, college persistence rates, and college completion rates. Overall, the results indicate that once we control for selection into colleges *there is very little difference in achievement gains across institutions in Texas*. That is, differences in the average earnings of students who enroll at UT Austin and Sul Ross State, for example, are mostly due to pre-existing differences in the types of students who enroll at those schools and not to differences the institutions themselves.

Conclusions and Policy Recommendations

We conclude with a general discussion of quantitative measures of achievement gains in higher education, addressing how the empirical analysis is relevant for educational policy and how these methods could be adapted for use in Texas. Particular attention is paid to the issue of incentive compatibility – the idea that for an achievement gain measure to be used for policymaking purposes, it must not induce institutions to take perverse actions that simply inflate their scores without concurrent achievement gains.

In the abstract, we conclude that indicators measuring achievement gains by individual colleges could be used meaningfully for policy making purposes. Published annually, they could allow individual colleges to monitor their own progress and compare themselves to peer institutions. Funding could also be tied to either the level of achievement gains within a school or to changes in that level to create incentives for schools to improve. In practice, however, currently available value-added and achievement gain measures are not sufficient for these purposes – including the ones proposed in this report. In short, they are simply too imprecise to make meaningful comparisons of achievement gains across institutions or over time. We therefore make the following policy recommendations:

1. **At best, the methods used in this report for measuring achievement gains in institutes of higher education should be a starting point for further research. We do not recommend using these methods to institute an annual value-added ranking of colleges in Texas for monitoring purposes and we urge particular caution against using these methods to develop a "pay-for-performance" type policy.** We present some evidence that achievement gain differences across Texas colleges are small. This, in and of itself, is important information that could - and likely should - be made public. However, as the estimates are imprecise, we cannot in general distinguish the quality of one college over the other and we therefore believe it would be inappropriate to form policy using these methods. Moreover, tying funding to such measures is tantamount to randomly allocating funds across colleges. In future research, we plan to use more refined econometric techniques (i.e.,

Instrumental Variables estimation) in an attempt to make achievement gain estimates more precise.

2. **We do not recommend the wholesale implementation of testing at the higher education level.** At first glance, standardized tests appear to be an ideal way to measure the entirety of the set of skills and knowledge acquired in higher education. But, upon closer inspection, it is unclear that a standardized test could be designed that would test both general and specific knowledge. There are well designed exams of general skills, like the CLA, and there are appropriate exams for testing specific skills for some majors, like the ETS subject tests. However, what is needed is a way to combine these two distinct aspects of educational achievement that is meaningfully comparable across students with vastly different goals and career plans.

3. **We do recommend the Coordinating Board explore other, more qualitative, methods to measure and increase achievement gains by schools.** One such method would be a program of institutional peer review. Under such a program, the Board could set up an exchange of faculty members across colleges to sit in on classes, discuss teaching techniques with faculty in their field, and evaluate the academic goals and standards of the peer school. Emphasis could be placed on sharing ideas for academic advancement and increasing the achievement gains of students while they are enrolled in higher education.

TABLE OF CONTENTS

EXECUTIVE SUMMARY . . . . . p. 1

SECTION 1 Introduction . . . . . p. 6

SECTION 2 Achievement Gain and Value-Added Measures in Higher Education p. 8

SECTION 3 Defining Achievement Gains – Specific vs. General Skills. . . . . p. 10

SECTION 4 Potential Proxies for Achievement
- 4.1 Tests . . . . . p. 12
- 4.2 Grades. . . . . p. 15
- 4.3 Wages. . . . . p. 15
- 4.4 Persistence and Completion Rates. . . . . p. 17

SECTION 5 Methods for Measuring Achievement gains and Value Added. . . . . p. 18
- 5.1 Using "before" and "after" observations of attainment . . . . . p. 19
- 5.2 Using only one observation of the attainment . . . . . p. 20
  - 5.2.1 Simple methods to control for observables . . . . . p. 21
  - 5.2.2 Advanced methods for causal inference . . . . . p. 23
    - 5.2.2.1 Explicitly modeling the selection process . . . . . p. 23
    - 5.2.2.2 Explicitly modeling the production process . . . . . p. 24
    - 5.2.2.3 Explicitly modeling the application process . . . . . p. 25

SECTION 6 Empirical Analysis of the Value-Added by Texas Public Universities
- 6.1 Data . . . . . p. 27
- 6.2 Construction of Sample . . . . . p. 28
- 6.3 Methodology . . . . . p. 29
- 6.4 Results . . . . . p. 32

SECTION 7 Policy Discussion . . . . . p. 35

SECTION 8 Conclusion and Policy Recommendations . . . . . p. 38

REFERENCES . . . . . p. 41

TABLE 1 – Summary Statistics . . . . . p. 43
TABLE 2 – Earnings Models Results . . . . . p. 44
TABLE 3 – Persistence Rate Models Results . . . . . p. 45
TABLE 3 – Graduation Rate Models Results . . . . . p. 46
TABLE 5 – Rankings Comparison across Models . . . . . p. 47

"Student achievement, which is inextricably connected to institutional success, must be measured by institutions on a "value-added" basis that takes into account students' academic baseline when assessing their results. This information should be made available to students, and reported publicly in aggregate form to provide consumers and policymakers an accessible, understandable way to measure the relative effectiveness of different colleges and universities."

– Quote from "A Test of Leadership," the Report of the Spellings Commission on Higher Education

1 Introduction

This report, commissioned by the Texas Higher Education Coordinating Board (THECB), has three main goals. The first is to summarize key issues surrounding the measurement of the achievement gains of students in colleges, and the use of these measures to improve educational outcomes. The second goal is to discuss the possibilities and limitations of measuring achievement gains in Texas colleges using data currently available at the THECB. The final goal is to estimate achievement gains in the most precise way possible using available data. Throughout, we strive to relate the theoretical discussion of the measurement and use of value-added and achievement gain measures to the practicalities of implementing these policies in Texas.

We begin by outlining the two main reasons why achievement gain measures are of interest to practitioners of higher education. First, there is a need to quantify the educational impact that colleges have on their students in a meaningful, unbiased, and comparable manner. Such information is useful both to administrators as a feedback tool and to the general public to inform decisions about which college to attend. Second, achievement gain measures can be used to institute incentivizing policies that aim to increase the achievement gains of their students[1]. Such policies hold schools *accountable* for the increase in performance of their students, and are currently being discussed extensively in K-12 education.

[1] We do not attempt to answer the normative question of *how much* knowledge or skill colleges should strive to impart on their students. However, there is a strong academic debate as to the "optimal" level of knowledge a school should add – see, for example, Dwyer, et al. (2006).

Throughout this discussion, we are careful to distinguish between the concepts of assessment, achievement gain, and value-added. *Assessment* is the measurement of some component of knowledge and skills at a single point in time. Achievement gain is the increase in some component of skills over time. Value-added is the increase in *the entire* skill set over a set time period. The most transparent method of measuring achievement gains involves using a pre- and post-assessment of achievement and then simply differencing in the two measures. Using this method, a measurement of achievement gains is also a value-added measure when the assessment used appropriately measures a student's entire skill set. Thus all value-added measures are achievement gain measures, but there are achievement gain measures that do not pass the bar for being measures of value-added. Similarly, while value-added and assessment are intimately linked, they are fundamentally distinct concepts.

The themes of assessment and accountability have salient parallels in the recent history of K-12 education, and research has shown that accountability policies can indeed increase value-added in these settings (Hanushek & Raymond, 2005; Carnoy and Loeb, 2002). We believe that assessment and accountability are no less important in higher education than they are in K-12 education; however, there are several important differences that render the wholesale importation of what has worked in primary education to higher education an impractical endeavor.

First, colleges *intend* to impart different skills and knowledge on different students; nursing students learn vastly different skills than engineering students and this makes it difficult to measure achievement in a consistent manner across students and schools. The difference between *general skills*, such as the ability to communicate and reason, and *specific skills*, such as nursing or engineering knowledge, plays an important role here. Thus, in the context of higher education, achievement itself is student specific, and encompasses all of the skills and knowledge a student acquires during her college career. It is an unobserved theoretical construct that is difficult, if not impossible, to measure in practice.

Therefore, we proceed to explore a series of quantitative metrics that one could use to assess achievement – standardized tests, grades (GPA), wages, persistence rates, and graduation rates. We discuss the merits and limitations of each measure, both in

general and in the specific case of Texas. While at first glance standardized tests may seem to be the most appropriate proxy for achievement, their use poses significant methodological and practical barriers stemming from the difference between general and specific skills. Therefore, in general, they may be appropriate for measuring achievement gains, but not value-added. Grades, on the other hand, are not standardized across institutions[2] and hence are not a viable proxy for achievement. Wages, however, are a particularly attractive proxy due to their direct link to labor market productivity and their ready availability at THECB. Persistence and completion rates are also currently available. However, as they do not fully reflect the knowledge and skills obtained during college, they are only appropriate for measuring achievement gains.

We then turn to a discussion of the methods that can be used to measure achievement gains for a given achievement proxy. The statistical techniques used critically depend on 1) which proxy measure is used, 2) when and how often the measure is observed, and 3) which students are observed. If we can observe the proxy twice, once before entering college and again some time in the future, measuring achievement gains is a simple task; simply take the difference between the two measures. When such data exists (e.g. if students are tested before and after entering college), this technique is ideal; we can estimate achievement gains at the individual level and this can be compared with other students within or across colleges. A crucial caveat with using an imperfect proxy for achievement is that the quality of the achievement gain measure is directly related to the quality of the proxy. Again, an achievement gain measure is only a measure of value-added if the achievement proxy captures all components of knowledge and skills imparted by colleges.

Unfortunately, proxies that can be measured at two time-points are rare, and there are no such proxies available at this time in Texas. When we observe a proxy measure at just a single time point, as in the case of wages, persistence rates, and graduation rates, we are forced to deal with two additional issues: 1) we cannot observe the level of achievement upon entering college, and 2) students are not randomly assigned to colleges, but rather systematically choose which college to attend based on private information. We devote considerable attention to these issues in the body of the report.

[2] Some may argue grades are not standardized *within* an institution, across majors, either.

We proceed to discuss several econometric models designed to deal explicitly with the issues raised above, paying particular attention to the practicalities of implementing these models with existing data in Texas. Finally, we estimate one particularly attractive model using actual THECB data. A broad conclusion is that, on average, colleges vary little in their ability to produce achievement gains once selection into college is controlled for. We conclude that measures of achievement gains or value-added cannot be reliably used to create accountability systems in higher education along the lines of the programs in place in some K-12 schools.

2 The Need for Achievement Gain and Value-Added Measures in Higher Education

There are two main reasons why practitioners of higher education should be interested in the development of indicators of the achievement that is gained by their students. First, such measures provide useful information about the quality of education provided by individual institutions. By correcting for the difference in types of students choosing to attend particular colleges, achievement gain measures provide an unbiased college-ranking not on *absolute* achievement, but on *relative* achievement gains. Such information can be used by administrators to monitor their progress relative to peer institutions and to evaluate the effectiveness of interventions aimed at increasing educational quality. For example, if Texas A&M institutes a policy to increase completion rates of under-prepared matriculants, but does not see a concomitant rise in completion rates relative to either itself or to the University of Texas, then administrators at Texas A&M have reason to believe that the program is not succeeding. Such information about educational quality is also immensely useful to the general public. Students and parents deciding which college to attend would know that the higher the achievement-gain indicator of an individual institution, the more they can hope to gain by attending that college. Value-added and achievement gain scores could compliment, and in many ways clarify, the existing signals of college quality such as popular national rankings, research output of faculty, class-sizes, and physical inputs such as books, buildings, and computers.

Second, state level value-added and achievement gain measures can be used to institute policies that increase achievement gains amongst their students. For example, THECB might consider tying funding to an institution's value-added score. By holding institutions directly accountable for increasing the educational performance of their students, such policies incentivize effective teaching and force colleges to direct resources towards that goal.

Much can be learned about value-added measures from other educational settings, and the past decade of research suggested that K-12 educators respond to the same work incentives as workers in the private sector. Academic research on state-level accountability systems and the No Child Left Behind Act at the national level has also suggested that students may learn more when teachers and schools are held accountable for the performance of their students (Carnoy and Loeb, 2002; Hanushek & Raymond, 2005). Unfortunately for practitioners of higher education, a valid accountability system relies on a consistent, unbiased, and comparable method for assessing a students' educational attainment at any point in time. Given an adequate assessment of achievement at two points in time, the difference between them tells us the knowledge that was added to a student by her teachers and school. If achievement gains are found to be too low, policies may then be implemented to give educators incentives to increase them, and provide consequences if they fail to do so.

Several key differences, however, bar the wholesale importation of popular K-12 accountability and assessment policies into higher education. Foremost is the fact that colleges have a very broad educational agenda with different students learning vastly different skills and knowledge. While the vast majority of third graders are on a similar academic track, learning long division and improving writing skills, there is wide variance in what knowledge is learned in the third year of college. A student in the School of Education might be preparing for student teaching while a Chemistry student is spending the majority of her time in a laboratory. While a single standardized test may come close to capturing achievement for third graders, it simply cannot do so in the context of higher education. Thus, despite the ever-present need to assess student achievement and to hold professors and colleges accountable for their students'

performance, we face a formidable task in developing an appropriate, quantitative value-added measure suitable for that setting.

3 Defining Achievement Gains and Value-Added in Higher Education – Specific vs. General Skills

We define achievement gains as the increase in some component of academic skills and knowledge that colleges impart on their students. For example, suppose that colleges administer a standardized test of general knowledge, like the CLA, before and after students enter college. The difference in the two scores represents an achievement gain in general knowledge. Only when the assessment used to measure achievement gains encompasses *all components* of academic skills and knowledge can we call the resulting achievement gain measure a measure of value-added. More formally, we define value-added as the increase in all social and academic skills and knowledge that colleges impart on their students. As such, value-added is student-specific and inherently difficult to measure. For nursing students, value-added includes what is learned in the core liberal arts curriculum as well as practical nursing knowledge, like how to inject a vaccine and accurately measure a patient's blood pressure. For a math major, value-added is quite different; while it still encompasses the same core liberal arts curriculum, we do not care if a math major knows how to clean a bedpan; we do, however, hope that she completes college with a thorough understanding of proof-based logic.

More generally, there are two main types of knowledge that an educational institution may seek to impart upon students - specific knowledge and general knowledge. Specific knowledge is what one would learn in a major course of study. In the previous example, nursing students learn how to inject a vaccine and math majors learn how to prove theorems. These practical, major-specific skills are what allow students to specialize, increase their productivity, and ultimately increase their income. However, there is a strong consensus in the academic community that increased productivity not only stems from job-specific skills, but also from the more general ability to function in a complex and rapidly evolving economy (Dwyer, 2006). These "general skills" include the ability to think critically, to communicate effectively, to

reason, and to interact with others. The ubiquitous tradition in higher education of combining a core curriculum with a skill-specific major stands as a testament to the fact that both general and specific skills are important aspects of a students' education.

Given that colleges strive to impart different skills on different students, our task of measuring the value-added by specific schools is now much harder. How are we to compare the human capital acquired by an engineering major to that acquired by an English major? If the engineer has acquired more specific economic skills that will lead to a higher wage, does that mean that there was more value added to her than to the English major who goes on to head an NGO? One can argue that every recipient of a bachelor's degree should have the same set of basic general skills, much like the traditional view of the required knowledge of a high school graduate. But does that mean that the engineer has less value added because she has fewer general skills than the English major who has fewer specific skills?

These questions, and the distinction between general and specific skills, must be kept in mind when choosing any metric to measure achievement gains in college. This is the task we turn to in the next section.

4 Potential Proxies for Achievement

We now discuss the advantages and disadvantages of various proxies that could be used to measure achievement gains in higher education. In doing so, we discuss the availability of each measure in the state of Texas, and note the temporal and monetary costs of obtaining those measures that are not currently available.

4.1 Tests

The historical roots of value-added measures lie in testing. First applied in the case of elementary schools, researchers used the year-to-year difference in test scores to proxy for the knowledge and skills gained by individual students. This measure became known as value-added, and is the basis for numerous state-level accountability programs across the country. Given the ubiquity of testing in primary and secondary schools, and the success of using annual tests to develop value-added measures in that setting, it is

natural to consider the use of test scores as a proxy for achievement in higher education. But given the preceding discussion, can this tradition of testing be extended to the realm of higher education? In general, the answer to this important question depends on two factors: 1) whether the test is norm- or criterion-referenced, and 2) whether the skills measured are general skills, specific skills, or both.

Norm-referenced standardized tests are expected to have a bell-shaped distribution of scores and no student is expected to be able to answer all questions on the exam. The SAT is an example of a standardized, norm-referenced test and it is referenced to a random sample of 1994 test takers. A criterion-referenced test is designed to test whether or not students have achieved a certain level of mastery over a given set of material; the Texas TAKS exam is an example of a criterion-referenced test. While many students fail to achieve the set standard on the TAKS, many others answer all questions correctly.

For our purposes, the main distinction between norm- and criterion-referenced tests is that a norm-referenced test distinguishes student skills much more accurately than a criterion-referenced test near the tails of the distribution of test-takers. This poses a significant problem for achievement gain and value-added measures. Since the typical student enrolling at UT Austin is near the top of the test score distribution for college matriculants, it is nearly impossible to use a criterion-referenced test to gauge how much UT students learn during their college careers. Thus, if a test-based proxy for achievement was to be implemented, we recommend using a norm-referenced test.

Now, consider the set of skills we desire to measure. If policy makers wish to measure general skills, there are several nationally recognized norm-referenced tests to choose from – the Collegiate Learning Assessment (CLA), ACT's Collegiate Assessment of Academic Proficiency (CAPP), and ETS's Measure of Academic Proficiency and Progress (MAPP). Each exam was developed and approved by well-trained educators and psychometricians and appear to be a valid proxy for the attainment of general knowledge and skills.

However, if one wishes to test specific knowledge and skills, a fundamental problem is that specific knowledge is discipline-specific. To accurately measure specific skills, the state must adopt an exam for each college major offered in the state. There

must be a nursing test for nursing students and an engineering test for engineering students. It is important to note that discipline-specific exams currently exist for some college majors. ETS has developed a series of Major Specific Tests, there are GRE subject tests for some majors, and some disciplines require licensure exams[3]. Unfortunately, none of these options comprehensively covers all college majors offered by schools in the state. We believe the development of exams for every major would be impractical, requiring an excessive in investment of scarce resources.

Each of the above types of exams is appropriate for measuring achievement gains. One can use a test of general skills, like the CLA, to produce a measure of achievement gains in general skills. Similarly, one can use a test of specific knowledge and skills, like the GRE subject tests, to measure achievement gains in a particular discipline. However, neither test alone can measure *all components* of knowledge and skills imparted by students, so neither type of exam can be used to measure value-added.

Even if the state were to initiate a campaign to develop major-specific tests for all college majors, there are still considerable hurdles to overcome if the state hopes to use these exams to measure value-added. First and foremost, the state would have to come up with an objective method to combine the scores from a general test with the scores from a specific test, and this is quite troublesome[4]. In particular, different majors place different emphases on the relative importance of specific versus general skills. Engineering majors spend the majority of their academic career learning practical knowledge related specifically to their field; English majors learn how to think creatively and write persuasively – skills useful in a wide range of jobs and specializations. How do we place an objective weight on the general and specific exam components when different majors place different weights on their relative importance? Should we penalize an institution for instructing a teacher how to work with children while rewarding another for teaching a philosophy major how to reason to a logical conclusion?

A further problem is that many students who enroll in college do not graduate. This poses a problem for testing as we hope to assess achievement before and after college, and it is unlikely that we would be able to test dropouts. If those students who

[3] See Yunker (2005) for an example related to CPA's.
[4] Also, since not all colleges offer all majors, how do we compare one institution to another? The state would have to come up with an objective way to overcome this issue as well.

would score lower on standardized exams are more likely to drop out, school-average value-added measures would be biased towards colleges with higher dropout rates. In high drop-out schools, we only observe value-added for the best students, whereas in schools with lower dropout rates, we observe value-added for all students. Given the above criticisms, we conclude that, at best, standardized tests can be used to measure achievement gains in general skills in higher education – they cannot be used to measure value-added.

4.2 Grades

Another natural measure one might consider using as a proxy for achievement is college grades or GPA. Unfortunately, grades depend on the courses a student takes and are set by individual instructors, often based on a bell-curve of students in the class. As such, they are not comparable within or across schools. Is a "C" in Spanish Literature comparable to a "C" in Quantum Mechanics, even if both courses are taken at UT Dallas? Is an "A" in Calculus at UT Austin the same as an "A" in the Calculus at TAMU-Commerce? In both cases, the answer is most likely "No". Therefore, there is no way to ensure that differences in grades reflect meaningful differences in educational achievement and this renders student GPA a poor proxy for achievement in higher education.

4.3 Wages

There are several compelling reasons why wages might serve as an appealing proxy for achievement. First and foremost, economic theory teaches us that in a free market, workers are paid in accordance with their contribution to the value of the good they produce. The more a worker adds to the value of a final product, the higher her wage will be. For example, an engineer that designs a car will be paid more than the factory worker that builds the car if the additional value added by the engineer is greater than that of the factory worker. Therefore, there is a theoretical argument for the use of wages as a proxy for human capital: wages signal productivity and productivity is the economic byproduct of the knowledge and skills learned in school.

Second, unlike tests, wages effectively combine the effects of both general and specific skills in a coherent and meaningful way - they are weighted in the manner that best reflects worker productivity. A nursing major's wages reflect his productivity in taking blood pressure and caring for his patients, while a political science major's wages reflect her writing and reasoning skills. Therefore, in a perfectly competitive labor market, any two workers with the same wage are equally productive; our nurse should be equally as productive as the political scientist provided the two are paid the same.

Labor markets, however, are not perfectly competitive and the above conclusion will not be valid to the extent that frictions in labor markets exist. For example, if workers are intrinsically motivated, as in the non-profit sector, then wages do not fully reflect productivity[5]. Furthermore, wages are only observed for those who work and the social, non-academic aspects of achievement such as civic-mindedness are likely not reflected in wages. Any empirical analysis must take this these caveats seriously, carefully controlling for avoidable biases and noting the effects of unavoidable biases. However, despite the drawbacks, we feel that the strengths outweigh the weaknesses and advocate the use of wages as a proxy for achievement. Since wages measure both general and specific skills, they can be used to measure value-added. Furthermore, quarterly wage data is currently available at the Coordinating Board for all college enrollees.

4.4 Completion rates and persistence rates.

A final set of proxies for achievement in higher education are college persistence and completion rates. Whether or not a student persists in college is certainly of interest to policy makers. What is unclear, however, is the extent to which eventually obtaining a college degree informs us of the general and specific skills that the student acquired while in college. Furthermore, we cannot be sure that the standard for college completion is the same across institutions. If courses are not as rigorous at UT Tyler as they are at UT Austin, should not the value of a degree from UT Austin be higher than that signaled by a degree from UT Tyler? Using only persistence or completion rates ignores this

[5] For the particular econometric method we use to estimate value-added, this is only a problem if these benefits or "warm glow" jobs vary systematically across colleges.

important distinction. Thus, these measures are a useful, albeit weak proxy for achievement. As such, they may be used to measure achievement gains, but not value-added. However, as they are readily available at the individual level at the THECB, we include them in our analysis.

To summarize, many potential proxies for achievement in higher education exist, but none is perfect. We limit our attention to those measures that can attribute some observable component of student achievement to the college they attend, and are that comparable across time and institutions. Despite their ubiquitous use in K-12 education, the methodological issues surrounding the use of standardized tests in higher education are many. Grades, or GPA, are not useful in creating achievement gain measures as they are not comparable across students, even within schools. Wages are an attractive proxy for achievement. Given their theoretical link to labor market productivity, they are particularly attractive for value-added models. However, as wages are not observed prior to entering college, any value-added measure derived from wages must carefully control for the "value" students had upon enrolling (we discuss this below). Finally, completion and persistence rates can be useful, using the same methodology as used for wages, but it is unclear what component of achievement they capture.

5 Statistical Methods for Measuring Achievement Gains and Value-added

There are several well known statistical methods that can be used to measure the achievement gains of students in individual colleges, based on the proxies for achievement we have outlined above. We discuss these below and throughout assume that data is available for all students in the state at the individual level, as it is at the THECB.[6] First, however, we offer a hypothetical example to help illustrate the statistical problem we face.

[6] If data is only available for a sub-sample of the population, or in aggregated form, any value-added analysis will, in general, be less precise. Other methodological concerns, such as selection bias, must then also then be addressed. We do not discuss these issues as they are not relevant for the THECB at this time, despite the fact that they have been significant barriers to other higher education value-added studies in the academic literature (Tracy and Waldfogel, 1995; Kreutzer and Wood, 2007; Berg and Kreuger, 2002; Brewer, Eide, and Ehrenberg, 1998).

Imagine students were randomly assigned to colleges in Texas, and we do not know the initial achievement level of students. This would imply that within every college the *average* achievement level of the freshman class would be the same – even though we do not know initial achievement. Imagine also that at some time in the future each student's true level of educational achievement could be measured, and this measure was averaged over all students in the school. Due to the random assignment into college, any differences in average achievement across schools could only be explained by the effect that the specific schools had on their students, as on average every school had a similar mix of student achievement in their freshman classes. These average differences in achievement levels across schools are exactly how one would define achievement gains by the college when only the "after" measure of achievement is available. For example, if the cohort of students who were randomly assigned to UT Austin had a higher average achievement level upon graduating than the cohort assigned to Sam Houston State, we could conclude with confidence that UT Austin did a better job of increasing the achievement of its students.

The real world, however, departs from this hypothetical ideal in two ways. First, we can not observe true achievement for every student, for all of the reasons discussed above. Second, and more importantly, students are not randomly assigned to colleges. In fact, the opposite is true in practice as students *deliberately and systematically* choose which school to attend, often for reasons unobserved by the researcher. The statistical methods we discuss and employ below were specifically designed to circumvent this second problem – *the selection problem* – and approximate the hypothetical ideal of random assignment to colleges.

While the scenario described above - when there is only an "after" measure of achievement - is the relevant case for higher education in Texas, we also discuss the measurement of value-added when both "before" and "after" observations are available for the sake of methodological completeness.

5.1 Achievement gains with "before" and "after" observations of achievement

If "before" and "after" measures of achievement are available, we could simply difference the measures as an estimate of achievement gains. For example, if the

achievement measure is a standardized test of general knowledge, one could administer the exam before students enroll and again upon graduating. In terms of intra-personal assessment, this technique is ideal. Schools would know how much knowledge was gained over the course of a student's tenure in college, and this could be compared with other students in the school.[7] Individual measures can then be aggregated to form institution-level achievement gain measures. Again, these achievement gain measures are only value-added measures when the exam used for assessment can adequately measure all components of knowledge and skills that students learn in college.

A serious issue, however, arises when college drop-out rates are high. When a college enrollee fails to persist to graduation, her achievement level is only observed once, implying that she is omitted from institution-level averages. Left unaccounted for, this type of *selection bias* tends to skew achievement gain measures in favor of those institutions with higher dropout rates. Students who drop out would have likely had lower achievement upon graduating, so dropping these students from the achievement gain calculation would increase the school-level achievement gain estimate[8].

Given the paucity of testing data available in higher education, relatively few studies have actually employed this type of differencing methodology. One that stands out is Yunker (2005), which measures the institution-level value-added of Certified Public Accountant (CPA) programs[9]. He concludes that there is little difference across institutions in their ability to add value to accounting majors.

5.2 Value-added with only one observation of achievement

When only one achievement measure is available, as with wages, we face the problem of not knowing the achievement level of a student was upon enrolling in college. For example, with wages we do not know how much a student would have earned if he

[7] Note, however, the limitations that major-specific skills would place on the usefulness of such inter-personal comparisons.
[8] For example, suppose Texas A&M and UT Brownsville had dropout rates of 8% and 25%, respectively. If a statistical analysis failed to correct for this difference, the A&M value-added score would be based on the top 92% of college enrollees while the UT Brownsville score will be based on the top 75% (assuming the lowest achieving students are the ones who drop out) and this inaccurately favors UT Brownsville.
[9] Using institution level averages, Yunker regressed school mean CPA exam scores on the mean SAT score for incoming accounting majors. The residual of this regression is an indicator of value-added. Effectively, this method is equivalent to the differencing methodology outlined above.

had entered the work force directly after high school. In this case, we cannot use the standard differencing technique; rather we must rely on sophisticated econometric techniques to account for selection into particular colleges[10]. Moreover, this setting is not ideal since it does not permit a within-school assessment of achievement gains; we can only compare achievement gains across institutions.

Fortunately, there exists an array of methods to deal with these problems. One class of methods attempts to explicitly control for all factors that jointly determine college choice and educational achievement – these encompass Ordinary Least Squares (OLS) and matching techniques. The other methods can be loosely categorized as those that impose further theoretical assumptions on the relationship between college selection and achievement. None of these methods is perfect, although they can be ranked in order of their usefulness. OLS and matching type methods are inadequate as one can never observe all factors jointly responsible for college choice and the achievement, thereby leading to biased estimates (this is the *omitted variable problem*). The second class of methods suffers from the caveat that they are only as good as the theoretical assumptions used to derive them.

5.2.1 Simple methods to control for observables

The first general class of methods used to control for selection into college are those that attempt to control for all factors correlated with both achievement and the decision of which school to attend. Left unaccounted for, such confounders can bias achievement gain measures. Two methods of this type have been employed to derive value-added measures in higher education: ordinary least squares (Estelle, 1989; Tracy & Waldfogel, 1997; Kreutzer & Wood, 2007) and matching (Black & Smith, 2004)[11]. Generally speaking, if we as researchers are able to observe and control for all factors jointly determining college choice and achievement, then either of these methods is a perfectly valid method for determining achievement gains.

[10] In practice, even most good studies using the differencing method also employ one or more of the techniques we discuss below. In fact, Yunker (2005), which we discussed above, employs an OLS regression to correct for pre-existing differences in the types of students enrolling in particular CPA programs.

[11] The key difference between matching and OLS is that matching is a non-parametric method, while OLS imposes a specific functional form assumption. Thus, matching is more general than OLS. However, the generality comes at the price of less transparency in the exposition of estimation results.

Unfortunately, this is almost never the case. Since we cannot observe these confounding factors, we are unable to control for them explicitly and hence, methods of this type are usually biased. For example, suppose that students with educated parents are more likely to attend UT Austin over Texas A&M, and we do not observe parental education (as is the case with THECB data). Furthermore, assume that students with educated parents are more likely to complete college. If completion rates are used as the measure of achievement, then both OLS and matching would produce biased achievement gain estimates due to the influence of the unobserved, yet influential factor of parental education. In particular, this scenario would overestimate the achievement gain of UT Austin because part of the estimated impact of enrolling at UT would be attributable to the unobserved influence of parental education.

Despite this limitation, many researchers in higher education have used OLS and matching in the past. The earliest such study is Estelle et al. (1989), which uses future earnings as an achievement measure and employs OLS to estimate the value-added by colleges. They find that private colleges on the East Coast have higher value-added than other colleges. Two more recent studies replicate this methodology in different higher educational settings - Tracy and Waldfogel (1995) in the context of MBA programs, and Kreutzer and Wood (2007) for undergraduate business schools. Both studies find significant differences in their estimated value-added rankings relative to an alternative ranking based on raw earnings of graduates. In a unique study, Black and Smith (2004) compare the results of OLS value-added models to matching models, and recommend the use of matching due to a key assumption of the OLS model not being satisfied.

We believe that measures based on either OLS or matching should be viewed, at best, as an intermediate step between a college ranking system based on raw average of the achievement measure (i.e., a pure ranking of average wages of graduates) and the hypothetical ideal of an achievement gain measure based on random assignment to colleges. Thus, while these methods can be implemented in the state of Texas, we do not recommend doing so for policymaking purposes, and offer a superior alternative below.

5.2.2 Advanced methods for causal inference

A set of more advanced methods for causal inference employ added behavioral assumptions to implicitly control for factors that jointly determine college choice and achievement. These methods fall into three categories[12]. First, selection models (Brewer et al., 1998) and instrumental variables (IV) models (Behrman et al., 1996) control for selection by positing a specific model governing college choice. Second, data envelope techniques (Cegniz & Yuki, 1998; Fisher et al., 2008) assume optimizing behavior by institutions. Finally, Dale and Kreuger (2002) develop a theoretical model of college choice which implies that all selection can be captured by a student's college application and acceptance profile. We discuss each of these techniques in some detail, describing the possibilities and limitations of implementing each in the state of Texas. However, we conclude that the Dale and Kreuger (2002) model is the most appropriate and use it in the empirical section below.

5.2.2.1 Explicitly modeling the selection process

Both selection models and IV techniques control for selection by positing a particular functional form for the selection process. In the present context, this simply means that the researcher must specify a formal relationship describing the college choice process. Selection models require the researcher to make a specific functional form assumption for the error term in the selection equation, while IV models require that at least one determinant of college choice – called the *instrument* - is independent of the measure of achievement except through its influence on college choice. For example, if we believed that students with educated parents were more likely to attend UT Austin, but that parental education does not otherwise influence future wages, then parental education is a valid instrument for college choice, and we can estimate the causal impact of college choice on wages, our proxy for the value-added by UT Austin.

[12] The vast majority of past studies have used wages or earnings as the value proxy. Due mostly to data quality issues, these studies have limited the scope of analysis to the return to "college quality," as opposed to individual colleges per se. Colleges are typically grouped together and researchers ask whether there is any causal difference in earnings of graduates of colleges across groups. For example, do graduates of Ivy League schools earn more than they would have if they had attended a flagship state college instead? Due to the tremendous data resources available in Texas, we are able to generalize this methodology to estimate the return to each particular college. As far as we know, we are the first to attempt this type of exercise.

An advantage of selection models over IV models is that they do not require the use of an instrument. Instruments are difficult to justify in practice; it is hard to find a variable that has no impact on the achievement measure, and there is no way to test whether the necessary relationship holds or not. On the other hand, the functional form assumption required in a selection model is often chosen in an ad hoc manner, thereby reducing its credibility.

Brewer et al. (1998) use Lee's (1983) selection model to estimate the impact of college quality on earnings, using a conditional logit for the selection equation. Overall, they find that college quality has a positive impact on earnings and that this impact is not as large as the raw difference in earnings that does not account for selection. While this method could be applied to colleges in Texas, the Brewer model is difficult to understand and computationally burdensome. We believe that any achievement gain model should be straightforward - if institutions do not understand how to improve their achievement gain score, then they tend not to respond to incentives to do so. We therefore do not recommend the use of this model.

Behrman et al. (1996) use an IV model to isolate the causal impact of college quality on earnings. Their findings are similar to Brewer et al. (1998) – college quality has a positive impact on earnings, but a smaller impact than what is implied by raw earnings difference across colleges of different quality. Again, we do not recommend using an IV strategy in Texas because 1) no plausible instrument for college choice is readily available, and 2) like selection models, IV models are not transparent to institutions, rendering any achievement gain rankings based on them less useful.

5.2.2.2 Explicitly modeling the production process

Alternatively, one could control for selection by explicitly modeling the educational production process, a technique commonly referred to as Data Envelope Analysis. Modeling the production process requires specifying all of the goals – educational and research based alike – that colleges seek to achieve, as well as the factors such as salaries and tuition levels that constrain institutional decision making. Furthermore, the researcher must specify the process by which colleges make decisions; typically the assumption is that colleges optimize an objective function. An obvious

drawback to this approach, however, is that there is no single common objective for colleges and universities.

Two studies - Cegniz & Yuki, (1998) and Fisher et al. (2008) - use data envelope analysis to estimate the value-added of MBA programs. Both studies find that the value-added rankings change considerably after controlling for selection. However, it is important to note the differences between MBA programs and traditional liberal arts colleges.

While data is available to employ this method for Texas colleges, we do not recommend its use. Data envelope analysis is only applicable to settings where the production process is known – such as in a factory, or to some extent, a business school. In higher education more generally, it is unclear what factors enter an institution's objective function, or even if institutions seek to maximize some objective function at all. On top of this strong criticism, data envelope analysis is not very transparent and it is unclear what factors actually drive the rankings, so institutions would again not know how to improve their performance.

5.2.2.3 Explicitly modeling the college application process

A final method that has been used to mitigate the selection problem is to explicitly model the college application and acceptance process. Dale and Krueger (2002) develop a theoretical model of the decision faced by a college applicant. When this student decides where to apply to college, she considers all of the information she knows about herself and chooses the optimal set of colleges to apply to. Some of the information she used in her decision, like SAT scores and income, is observable to the researcher. However, much of it, like motivation and desired career path, is not observed. Dale and Kreuger's insight is to recognize that as researchers we often observe students' application and acceptance profiles and that this data contains all of the personal information the student and colleges used to make their choices, *even the information that the researcher can not observe directly*. Thus, by comparing only students with identical application profiles, we can go a long way towards correcting for selection into college.[13]

[13] This does not fully correct for selection, however, as amongst students with identical application and acceptance profiles we do not know why different students ultimately chose different schools.

Dale and Krueger (2002) use this methodology to determine the labor market return to college quality, where college quality is determined by the mean SAT scores of incoming freshmen. While they find significant returns to college quality, the returns are, as expected, not as large as the raw differences in earnings across colleges of different quality.

In many ways, the Dale and Krueger methodology is ideal for the empirical question posed in this paper. The largest criticism of this method is that the identification is only valid if *all* students apply to college in a rational way. Otherwise, a student's application decision does not fully capture the unobservable variables that also impact the attainment measure in question. However, the restrictions placed on behavior are certainly weaker than those required in the other models we have considered above. Moreover, the state already collects data on college application and acceptance behavior for all students, so no extra data collection is necessary. Finally, the methodology is straightforward, easy to understand, and relatively easy to implement. For these reasons, we recommend that the state use the Dale and Krueger (2002) methodology if it decides to develop achievement gain and value-added measures for Texas public colleges, and we use this technique in the following empirical exercise.

6 Empirical Analysis of Achievement Gains and Value-Added by Texas Public Universities

We turn now to an empirical analysis of the achievement gains and value-added by Texas four-year public Universities. Of primary interest is whether individual colleges *vary* in their ability to increase the achievement of their students. Towards this end, we estimate econometric achievement gain models and produce rankings of schools based on the outcomes of these models. Despite the wealth of individual-level data housed at THECB, the scope of the analysis is limited by the measures of achievement that are currently available. We thus concentrate on three such measures: earnings, persistence rates, and graduation rates. Based on the previous discussion, all three of these measures can be considered measures of achievement gains, but only those based

on earnings can be considered value-added measures. Our analysis shows that certain schools do vary in ability to increase the achievement of their students, but that the variation is slight.

We begin with a description of the data, giving considerable attention to the use and availability of specific THECB databases. A description of the intuition behind each econometric model follows. The empirical results are then presented, along with the implied achievement gain rankings, and we conclude with a general discussion of the results and policy implications.

6.1 Data

Several secure, individual-level databases housed at the THECB offices in Austin, Texas are used in this analysis. Combined, they track the universe of Texas public high school graduates through Texas colleges and into the workforce, and can be linked to each other and across time using Social Security Numbers. For this empirical exercise, we construct a longitudinal panel of the 1998 and 1999 cohorts of Texas public high school graduates.[14]

While college-level data is reported directly to the Coordinating Board by Texas colleges and universities, high school level data, wage data, and testing data is reported first to other agencies, and then shared with THECB. The Student Report, the Graduation Report, and the Application Report are reported by all Texas colleges directly to THECB for the years 1998-2007. Along with several demographic variables, the Student Report includes semester credit hours attempted and declared major for every institution a student attended. The Application Report identifies which public institutions a student applied to and was accepted at during the report year, while the Graduation Report indicates any degrees earned during the report year.

The Texas Education Association (TEA) requires Texas public high schools to report data on all their students. They have shared some of this data with the THECB for the years 1991-2007, including a yearly report containing, for each graduate, demographic variables, an indicator for the high school of graduation, and a list of all academic courses taken during

[14] We use the 1998 and 1999 cohorts so that there is ample time (at least 8 years) for students to both complete college and enter into the workforce, at which time we believe their wages will be indicative of value-added in college. Data for earlier cohorts is not available.

grades 9-12. The Texas Workforce Commission (TWC) collects earnings data on all employed persons in the state, and have shared it with the THECB from the first quarter of 1998 through the second quarter of 2007. These reports include total quarterly nominal earnings for each job a person held.

SAT and ACT test scores are purchased annually by the THECB from the corresponding testing agencies. The Board currently has scores from SAT tests taken between 1998 and 2004, and ACT scores for the years 1998 and 1999. In addition to test scores, the SAT database includes demographic information self-reported by the test taker on the day of the exam – household income, the educational attainment of each of the test takers parents, planned college major, and planned level of educational attainment. Unfortunately, the ACT databases do not have this additional demographic information.

Lastly, we use the 1997 County and City Data Book from the US Census to assign several indicators of local labor market conditions to each student[15]. This data is not housed at the THECB, but can be found at the Geospatial and Statistical Data Center at the University of Virginia Library (http://fisher.lib.virginia.edu/).

6.2 Construction of Sample

We begin by identifying all 1998 and 1999 graduates of Texas public high schools who enroll in the Texas Public University System in the year following graduation – approximately 82,000 students.[16] From this group, the sample is further restricted in several ways in order to identify achievement gain differences across institutions. First, college enrollees that do not have a valid SAT or ACT score are excluded. Since most Texas public colleges require one of these tests as a condition for enrolling, this restriction only excludes 5,773 of the original sample. Second, 1501 graduates of high schools located in 13 largely rural counties are excluded as there is no data on local labor markets for these areas. The next restriction stems from the preferred empirical strategy employed in the analysis (that of Dale and Kreuger, 2002) – we exclude 10,613 students who applied to more than 3 Texas public colleges and whose application/acceptance profile do not exactly match at least 9 other enrollees in the sample. Finally, when using earnings as an outcome measure, the

[15] Students are assigned the labor market variables of the county in which he or she went to high school.
[16] As any student may be enrolled in more than one college in a given year, we assign each student to the college or university at which he or she attempted the most credit hours during that year.

sample is restricted to include only college enrollees who 1) have completed their schooling, 2) are in the Texas labor force, and 3) earned at least $2000 in the fiscal year beginning 8 years after graduating high school – 19,648 observations.

Therefore, two separate samples are used, with 45,657 observations in the earnings models and 65,305 observations in the persistence and graduation rate models. With sample sizes this large, there is ample statistical leverage to expose differences in achievement gains across colleges, if these differences exist. Summary statistics for both of these samples are presented in Table 1. Note that for the "earnings" sample, earnings, graduation, and completion rates are higher than in the "persistence and graduation rate" sample. This, however, is by design – only students earning more than $2000 a year are included in the "earnings" sample. It also seems reasonable that students earning less than $2000 per year would also be the least likely to persist though and complete college, driving the observed differentials across samples. Furthermore, students in both samples tend to be more academically prepared and have higher SAT scores than the state as a whole (not shown). For example, the mean SAT score in the earnings sample is 1026, considerably higher than the state average of 994. This, again, is by design. Our samples include only enrollees at four year colleges as opposed to SAT takers at large, and we therefore expect them to have higher SAT scores and be more academically prepared, on average, than students who take the SAT and fail to enroll.

6.3 Methodology

Three distinct sets of achievement gain rankings are produced, one for each measure of achievement identified above: earnings, college completion rates, and college persistence rates. While none of these is an ideal measure, each identifies different, and important, aspects of the college-going process. Identification of the *variance* across colleges in the ability to produce achievement gains is difficult, mainly due to the selection problem discussed above. Our preferred methodology controls for selection of students into schools in the most comprehensive manner possible; however, it is instructive to see how this process works. For each outcome, we therefore produce *three* sets of college rankings, sequentially adding more controls for student selection into college. (These rankings can be seen in tables 2 through 4.)

The first ranking – the "Unconditional Ranking" – simply lists the 30 Texas colleges according to the average of the outcome measure.[17] This naïve approach ignores the selection problem completely. As an example, if using this procedure UT Austin is ranked above Angelo State because, we would conclude that 8 years after enrolling students at the former earn more on average than those at the latter. Clearly, this method is fraught with problems, chief among them being the selection issue – different colleges attract different types of students who come to college with different levels of preparation and academic skills. This "selection problem" must be overcome if we are to accurately rank colleges in their ability to produce achievement gains.

The second ranking – or the "Conditional Ranking" – uses Ordinary Least Squares (OLS) to explicitly control for student-level, observable characteristics that may influence collegiate choice. We control explicitly for a large set of observable student-level characteristics that may impact the outcome of interest and are also correlated with college choice. However, OLS ignores the *unobservable* characteristics (like parental influence and ability) that likely play an important role in college choice. Continuing with our example, imagine the only difference between freshmen at UT and Angelo State is the fact that UT students have higher SAT scores. In this case, the Conditional Ranking would fully account for selection as we control explicitly for SAT score in the model. However, if UT students are also more likely to come from families with more involved parents, then the Conditional Ranking will be biased as we do not observe parenting skills and thus can not control for them. As such, the Conditional Ranking represents an intermediate step that controls for many, but not all, of the factors that simultaneously determine college choice and achievement measures.

The final, and preferred, "Matched Applicant" ranking goes a long way towards controlling for the unobservable confounds that can bias the Conditional Rankings model by employing the Dale and Kreuger (2002) methodology. This procedure is identical to the one used in the Conditional Ranking except that it further controls for the full application-acceptance profile of each college enrollee. In essence, this procedure only compares students who applied and were accepted to the exact same set of colleges. For example,

[17] The only minor exception is that the rankings based on earnings also condition upon labor market experience, as is common practice in the labor economics literature. See Mincer (1973) for more details. These conditional averages are therefore obtained using an OLS regression.

suppose Student A applied only to Texas Tech, Texas State, and Texas A&M, and she was accepted only at Texas Tech and Texas State. In this case, Student A would only be compared to other students who applied to Texas Tech, Texas State and Texas A&M, and were accepted only at Texas Tech and Texas State.

6.4 Results

Earnings

Column 1 of Table 2 presents the Unconditional Ranking based on earnings as a measure of achievement. The coefficient for any school represents the percentage difference in annual earnings for a typical student from that school relative to a typical student from Texas A&M[18]. For example, the typical enrollee at UT Austin (#2 in the Unconditional Ranking) earns 8.3% less than the typical Texas A&M enrollee, 8 years after high school graduation. The double asterisk on this coefficient indicates that the estimate is statistically significant (from the implied "zero" coefficient on Texas A&M) and thus precise enough that it is likely not attributable to chance.

There are several points to note. First, there is considerable variance in raw earnings across colleges. Students from the top-ranked school, Texas A&M, earn 63.4% more on average than students with similar labor market experience who instead enrolled at the bottom-ranked school, Texas Southern University. Second, this method ranks schools similarly to the "popular" rankings published in the U.S. News and World Report and Barron's Magazine. This similarity should not be surprising as these popular rankings do not control for the selection problem either. The highest ranked schools in the Unconditional Ranking – UT Austin, Texas A&M and UT Dallas – are the three most selective public universities in the state.

Conditional Rankings for earnings are listed in column 2 of table 2. In this model we control for the following observable factors: demographic characteristics, SAT scores, high school curriculum, initial college major, local labor market conditions, and high school level characteristics. Each school's coefficient estimate represents the percentage difference in

[18] The choice of reference school – here, we chose Texas A&M - is arbitrary and does not affect the rankings or the interpretation of the coefficient estimates.

annual earnings for a student of that school relative to a student *who is identical in terms of the observable characteristics* but instead initially enrolled at Texas A&M. As such, the typical student who enrolled at UT Austin earned 6.7% less than an *observationally equivalent* student who enrolled at Texas A&M. Note that while these rankings are quite similar to the Unconditional Rankings, estimated differences in earnings across schools are now much smaller – the correlation coefficient between the Unconditional and Conditional rankings is 0.88, and the top- and bottom-ranked colleges remain the same. However, the magnitude of the coefficient estimate for the bottom school, Texas Southern University, drops from 0.634 to 0.37 - a 42 percent decline. This indicates that approximately 42% of the difference in earnings for Texas Southern University and Texas A&M students is accounted for by the observable variables that are included in the Conditional Rankings model.

Lastly, column 3 of Table 2 presents the Matched Applicant Rankings. Coefficient estimates represent the average increase or decrease in earnings students would have had if they had instead attended Texas A&M. Using our familiar example, this implies that a typical student enrolling at UT Austin would earn 3.8% less than if she had instead chosen to enroll at Texas A&M. Importantly, however, this estimate is not statistically different from zero, and therefore is not precise enough to determine if the observed earnings differential is actually due to differences between UT and A&M or simply due to chance. To the extent that earnings fully capture college quality, these Matched Applicant Rankings are our best estimate of achievement gains of individual colleges in Texas. Moreover, since earnings come close to fully capturing the knowledge and skills imparted by colleges to their students, these measures can be considered measures of value-added.

Several points are again important to note. First, the Matched Applicant coefficients are even smaller than the Conditional Ranking coefficients, with the range dropping from 0.37 to 0.29. Second, coefficient estimates are rarely significantly different from zero, which implies that differences across colleges in their ability to add value are slight or nonexistent. Therefore, most of the difference in earnings across Texas colleges is due to the types of students that attend those schools, and not due to the quality of education they receive. Lastly, the ordering of schools according to the Matched Applicant Rankings are very different from either the Unconditional or Conditional Rankings. However, as coefficient

estimates are most likely not significant from one another, rankings based on these imprecise estimates do not have substantive meaning.[19] For example, one can not be sure that the top-ranked Texas A&M at Galveston actually adds more value than most of the other schools, except possibly for those at the bottom of the ranking list, as the estimates are statistically indistinguishable from one another.

Persistence and Graduation Rates

The above analysis is repeated using 1-year persistence rates and graduation rates as a measure of achievement and the resulting rankings are presented in Tables 3 and 4. For the most part, we find similar patterns to when earnings were used as a measure of attainment. As we move rightwards across the columns, adding successively more controls for both observable and unobservable characteristics, absolute differences in outcomes across schools diminish. Also as with earnings, when moving from the conditional rankings to the matched applicant rankings, the ordering of schools changes drastically and overall differences between schools become less precise. It is important, however, to keep in mind that persistence and graduation rates measure very different things than do earnings and our interpretation of the resulting rankings should take these differences into account. More importantly, since these measures do not fully capture the growth in all components of knowledge and skills, they can only be considered measures of achievement gains; they should not be interpreted as measures of value-added.

7 Policy Discussion

We proceed to a general policy discussion of achievement gains and value-added in higher education. We address the ways in which the empirical analysis above is relevant to educational policy, and note how elements of these methods would have to change if the Coordinating Board were to use them for specific policy purposes.

Uses for achievement gain and value-added measures

[19] As coefficient estimates are statistically insignificant from zero, they are likely to be insignificant from each other as well. However, we would have to perform pair-wise hypothesis tests to prove this statement conclusively.

Two reasons can be identified as to why achievement gain and value-added measures may be desirable in higher education: *external, cross-institution comparisons at any point in time* and *internal institutional evaluation over time*. In terms of external comparisons, common knowledge of the distribution of achievement gain and value-added measures across colleges may increase competition within the higher education sector. For example, consider two schools with similar resources and incoming students – Texas Tech and the University of Houston. If Texas Tech is less able to increase achievement than is the University of Houston, and this is common knowledge amongst students, parents, policy makers, then Texas Tech should feel pressure to increase its performance to match, or exceed, that of UH.

In relation to internal evaluation, it is instructive for institutions to know the size of the achievement gains of their students and how these gains are changing over time. Such information would help Texas Tech, for example, to monitor progress towards its goal of increasing achievement gains. Certain policies Texas Tech employs will work and others might not, and changes in achievement gains would measure the effectiveness of the policy in accomplishing its goal.

Caveats in using achievement gain and value-added measures for Policy Formation

Despite the obvious desirability of achievement gain and value-added measures, three issues must be understood and addressed before they are used to form policy. First, assumptions made in estimating achievement gains and value-added may not, in fact, be valid. Of particular concern is the assumption that the achievement gain methodology controls for all unobserved factors that simultaneously influence both the attainment measure of study and the choice of school to attend. The Dale and Krueger (2002) model, by making comparisons only between students who applied to and were accepted at the same set of colleges, mitigates this concern somewhat, but there are likely confounding factors that will produce biased estimates of achievement gains and we can only hope that this bias is indeed small.

A second caveat is that for the matched-applicant rankings to represent true value-added, the outcome considered must fully capture the entire set of social and educational skills and knowledge that colleges impart on their students. Otherwise, specific outcome

measures will only isolate the causal impact of college on a subset of the components of achievement. As seen in the empirical results, achievement gain rankings across colleges differ widely when different outcomes are used (wages, completion rates, or persistence rates). These rankings are re-displayed in Table 5 for easier comparison. As is evident, some schools are better at increasing earnings while others are better at getting students to complete college. However, since earnings is the only achievement measure that captures all components of knowledge and skills, only the results based on earnings should be considered true value-added measures.

The final caveat is that even if we believe that the achievement gain rankings actually measure what they purport to do, the resulting estimates must be estimated precisely. Otherwise, we cannot tell whether there is simply no difference in achievement gains across colleges, or whether there actually is a difference, but our estimates are too noisy to detect it. In the above analysis controlling for selection into college, we nevertheless find that differences in achievement gains across colleges are small. Further detailed statistical analysis is needed to determine whether these small differences are statistically different from one another, and therefore can have substantive meaning for policy formation.[20]

Using Achievement gains and Value-added measures for Accountability and Monitoring

Caution must also be exercised if achievement gains or value-added measures are to be used for accountability and monitoring purposes. For one, the chosen achievement gain measure must respond to policies in a timely manner; otherwise, institutions will not be able to productively increase achievement gains in the short term. This issue is especially salient when earnings are used as an achievement measure, since we believe that only wages of students sufficiently absorbed in the labor market are accurately indicative of the skills and knowledge acquired in college. In our analysis, we use wages 8 years after initial college enrollment, and we caution against using a shorter time period. Using an achievement gain measure with such a long time-lag in responsiveness to institutional interventions would significantly reduce incentives for individual schools to increase achievement. Thus, achievement gain measures employing earnings as the

[20] We plan to conduct these statistical hypothesis tests in future work.

assessment measure are less useful for the purposes of short-term accountability and monitoring.

Some have suggested using achievement gain or value-added measures as a basis for an incentive pay policy, whereby institutional funding is tied directly to the achievement gains of their students. This mechanism is currently being used in some K-12 school districts, and is an ubiquitous feature of any for-profit business model. Pay for performance policies are intended to incentivize institutions to maximize achievement gains; however, tying funding to achievement gains may also create perverse incentives to "game the system." That is, if schools know how the state is measuring achievement gains, they may find it in their best budgetary interest to act in a manner that increases their achievement gain ranking *even if their actions do nothing to actually increase the true achievement gains of their students*. For example, if the achievement gain measure is based on college completion, institutions should realize that by lowering graduation standards to allow more students to graduate, they would increase their funding. Such actions would improve the schools ranking – hence, increasing their funding - but would also quite likely decrease true achievement gains of students.

To prevent such perverse behavior, a pay-for-performance policy must use an achievement gain measure that satisfies two conditions, and hence be deemed "incentive compatible." The measure must 1) be a value-added measure and 2) fully account for selection into colleges. If condition 1 were met, the incentive problem in the example above would be mitigated and "gaming the system" would not be feasible. Condition 2 assures that schools cannot alter their achievement gain score by changing their admissions policies. If the achievement gain measure does not fully account for selection, then institutions would have the incentive to be more selective and only admit those students that inflate their value-added scores.

Are the empirical achievement measures we derived above "incentive compatible?" As evidenced by the preceding example, those based on completion and persistence rates surely violate condition 1. The extent to which achievement gain measures based on earnings violate condition 1 is less obvious. Earnings reflect labor market productivity, so incentivizing policies designed to increase graduate earnings should be tantamount to rewarding policies that make students more productive workers.

While we certainly desire college graduates to be productive workers, we also would like for them to be productive citizens – and earnings do not capture this component of knowledge and skills. The extent to which our achievement gain measures violate condition 2 hinges crucially on whether or not we believe students apply to college in an optimal manner. If so, then our matched applicant model fully captures selection into college, and condition 2 holds for all three outcomes.

Finally, it is a separate question as to whether or not schools will actually take advantage of the opportunity to game the system. We would hope that social norms and pro-social behavior would mitigate the problem, but given that this type of behavior has been observed in response to poorly designed accountability measures in K-12 education (see, for example, Figlio & Getzler (2002)), it is crucial to be cognizant of the problem and eliminate perverse incentives to begin with.

8 Conclusion and Policy Recommendations

This paper has addressed three main goals. The first was to summarize the issues involved in the measurement of achievement gains in college. The second was to discuss the possibilities and limitations of measuring this value-added in Texas colleges using data currently available at the THECB. The third was to carry out empirical estimates in the most precise way possible using the data available at the THECB.

We began by distinguishing between the concepts of achievement gains and value-added: achievement gains are the increase in some component of skills and knowledge over a given period of time, while value-added is the increase in the entire set of skills and knowledge over a given period of time. From a methodological standpoint, this definition poses two key problems for measuring value-added. First, true value-added is inherently unobservable to the researcher, and we thus only observe imperfect proxies. Second, value-added in higher education is individual specific and difficult to compare inter-personally. This observation stems from the differences between specific and general skills that are learned across different college majors and institutions. We continued by discussing the relative merits of several potential proxies for achievement, including test scores, grades, wages, persistence rates, and completion rate. We

concluded that wages are the best observable proxy available due to their direct link to labor market productivity, and hence knowledge and skills.

Several econometric techniques that have been used to estimate achievement gains in higher education were then reviewed. The key issues involved in the choice of technique were 1) the extent to which the method adequately controls for selection into college, 2) the method's transparency to institutions and the general public, and 3) the practicality of implement the method in Texas using THECB data. In light of these considerations, our recommendation was to use the matched applicant model of Dale and Kreuger (2002).

Next, we used the matched applicant model to empirically estimate the achievement gains of students in Texas college, showing for comparison estimates obtained from both a naïve ordinal ranking (the Unmatched Ranking) and a ranking controlling for observables (the Conditional Ranking). Our general finding is that there is little difference in achievement gains across institutions. Most of the difference in the achievement gains of matriculates from Texas A&M vs. UT Pan Am, for example, is due to the types of students attending those schools and not relative institutional differences.

Finally, we turned to a general policy discussion of achievement gain measures in higher education. Here, we argued that for an achievement gain measure to be used for policy-making purposes, it should have two properties: the outcome considered must capture all components of knowledge and skills imparted by colleges, and the methodology must account for selection into colleges. Otherwise colleges might have the incentive to pursue perverse policies to artificially inflate their achievement gain scores. We also argued that for a achievement gain measure to be practically applied in the policy arena, it must be precisely estimated and respond to interventions in a timely manner.

In the abstract, achievement gain measures could be used for meaningful policy-making purposes. Published annually, they could allow individual colleges to monitor their own progress and compare themselves to peer institutions. Funding could also be tied to either the level of achievement gains, or to changes in that level, thereby creating incentives for schools to improve. In practice, however, currently available achievement gain measures are not sufficient for these purposes.

In light of the discoveries made in this report, we make the following policy recommendations for the state of Texas:

1. **At best, the methods used in this report for measuring achievement gains in institutes of higher education should be a starting point for further research. We do not recommend using these methods to institute an annual value-added ranking of colleges in Texas for monitoring purposes and we urge particular caution against using these methods to develop a "pay-for-performance" type policy.** We present some evidence that value-added differences across Texas colleges are small. This, in and of itself, is important information that could - and likely should - be made public. However, as the estimates are imprecise, we cannot in general distinguish the quality of one college over the other and we therefore believe it would be inappropriate to form policies using these methods. Moreover, tying funding to such measures is tantamount to randomly allocating funds across colleges. In future research, we plan to use more refined econometric techniques (i.e., Instrumental Variables estimation) in an attempt to make achievement gains estimates more precise.

2. **We do not recommend the wholesale implementation of testing at the higher education level.** At first glance, standardized tests appear to be an ideal way to measure the entirety of the set of skills and knowledge acquired in higher education. But, upon closer inspection, it is unclear that a standardized test could be designed that would test both general and specific knowledge. There are well designed exams of general skills, like the CLA, and there are appropriate exams for testing specific skills for some majors, like the ETS subject tests. However, what is needed is a way to combine these two distinct aspects of educational achievement that is meaningfully comparable across students with vastly different goals and career plans.

3. **We do recommend the Coordinating Board explore other, more qualitative, methods to measure and increase achievement gains by schools.** One such method would be a program of institutional peer review. Under such a program, the

Board could set up an exchange of faculty members across colleges to sit in on classes, discuss teaching techniques with faculty in their field, and evaluate the academic goals and standards of the peer school. Emphasis could be placed on sharing ideas for academic advancement and increasing the achievement gains of students while they are enrolled in higher education.

REFERENCES

Behrman, Jere, et al. (1996). "College Choice and Wages: Estimates Using Data on Female Twins." *Review of Economics and Statistics.* 78(4): 672-685.

Black, Dan & Jeffrey Smith (2006). "Estimating the Returns to College Quality with Multiple Proxies for Quality." *Journal of Labor Economics.* 24(3): 701-728.

_____ (2004). "How Robust is the Evidence on the Effects of College Quality: Evidence from Matching." *Journal of Econometrics.* 121(1-2): 99-124.

Black, Dan et al. (2005). "College Quality and Wages in the United States." *German Economic Review.* 6(3): 415-443.

Brewer, Dominic, et al. (1999). "Does it Pay to Attend an Elite Private College? Cross-Cohort Evidence on the Effect of College Type on Earnings." *Journal of Human Resources.* 34(1): 104-123.

Carnoy, Martin & Susanna Loeb (2002). "Does External Accountability Affect Student Outcomes? A Cross State Analysis. *Educational Evaluation and Policy Analysis*, 24(4): 305-331

Cengiz, Hakesever & Muragishi Yuki (1998). "Measuring Value in MBA Programs." *Education Economics.* 6(1): 11-25.

Dale, Terry Berg & Alan Krueger (2002). "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables." *Quarterly Journal of Economics*. 117(4): 1491-1527.

Dwyer, Carol, et al. (2006). "Culture of Evidence: Postsecondary Assessment and Learning Outcomes, Recommendations to Policymakers and the Higher Education Community." ETS Policy Paper.

Eide, Eric, et al. (1998). "Does it Pay to Attend an Elite Private College? Evidence on the Effects of College Quality on Graduate School Attendance." *Economics of Education Review.* 17(4): 371-376.

Estelle, James, et al. (1989). "College Quality and Future Earnings: Where Should You Send Your Child to College?" *American Economic Review.* 79(2): 247-252.

Figlio, David N. and Getzler, Lawrence S. (2002) "Accountability, Ability and Disability" *NBER Working Paper* No.W9307.

Fisher, Dorothy, et al. (2007). "A Value-Added Approach to Selecting the Best Master of Business Administration (MBA) Program." *Journal of Education for Business.* 83(2): 72-76.

Kreutzer, David & William Wood (2007). "Value-Added Adjustment in Undergraduate Business School Ranking." *Journal of Education for Business.* 82(6): 357-362.

Lee, Fung-Lei (1983). "Generalized Econometric Models with Selectivity." *Econometrica.* 51(2): 507-512.

Light, Audrey & Wayne Strayer (2000). "Determinants of College Completion: School Quality or Student Ability?" *Journal of Human Resources.* 35(2): 299-332.

Miller, Charles, et. al. (2006) "A Test of Leadership: Charting the Future of U.S. Higher Education." A Report of the Commission Appointed by the Secretary of Education Margaret Spellings. *http://www.ed.gov/about/bdscomm/list/hiedfuture/reports/final-report.pdf*

Rivkin, Steven, Eric Hanushek, & John Kane (2005). "Teachers, Schools, and Academic Achievement." *Econometrica*, 73(2): 417-458.

Rodgers, Timothy (2007). "Measuring Value-Added in Higher Education: A Proposed Methodology for Developing a Performance Indicator Based on the Economic Value Added to Graduates." *Education Economics.* 15(1): 55-74.

Tracy, Joseph & Joel Waldfogel (1997). "The Best Business Schools: A Market-Based Approach." *Journal of Business.* 70(1): 1-31.

Wales, Terrence (1973). "The Effect of College Quality on Earnings: Results from the NBER-Thorndike Data." *Journal of Human Resources.* 8(3): 306-317.

Yunker, James (2005). "The Dubious Utility of the Value-Added Concept in Higher Education: The Case of Accounting." *Economics of Education Review.* 24: 355-367.

Table 1 Summary Statistics - Mean (Standard Deviation)

	Of the sample used in the:	
	Earnings Models	**Persistence and Graduation Rate Models**
Outcome Variables		
Annual Earnings, $	$37,707 ($24,633)	$27515 ($26,365)
% Graduated within 8 years	68.57% (46.43%)	62.18% (48.49%)
% Persisted into 2nd year of college	86.55% (34.12%)	84.59% (36.1%)
Demographic Variables		
SAT (or converted ACT) score	1026.16 (17.8)	1033.41 (18.2)
% Black	10.41%	9.82%
% Hispanic	18.86%	18.84%
% White	64.92%	64.82%
% Male	43.56%	44.87%
% Eligible for Free Lunch	13.39%	13.26%
% At Risk of Not Graduating	10.71%	10.69%
# of AP courses taken	1.01	1.04
H.S. Courses as Senior, % enrolled		
English as a 2nd Language (ESL)	0.14%	0.18%
Gifted and Talented Program	25.51%	25.77%
Calculus	25.19%	25.55%
Pre-Caculus	24.02%	23.86%
Algebra 2	5.80%	6.02%
Biology	7.25%	7.44%
Chemistry	10.17%	10.40%
Physics	27.11%	27.70%
Observations	65,305	45,657

Notes: Earnings are the sum of the 4 quarterly earnings 8 years after graduating high school. Other HS courses not displayed are used as controls.

Table 2 - Earnings Models - Log earnings regressions, including dummy variables for 4-year Public schools in Texas

	Unconditional Ranking		*Conditional Ranking*		*Matched Applicant Ranking*	
	1		**2**		**3**	
Variables	Coefficient	Rank	Coefficient	Rank	Coefficient	Rank
TAMU - Galveston	`-0.221**	6	`-0.144**	10	0.069	1
Sul Ross State University	`-0.386**	21	`-0.18**	15	0.047	2
Texas State University	`-0.253**	10	`-0.105**	7	0.022	3
Stephen F. Austin State Univ.	`-0.241**	9	`-0.119**	8	0.012	4
Texas Women's University	`-0.254**	11	`-0.081*	3	0.011	5
Midwestern State University	`-0.423**	25	`-0.256**	29	0.002	6
Angelo State University	`-0.429**	26	`-0.249**	27	0.0001	7
Texas A&M University	--	1	--	1	--	8
University of Noth Texas	`-0.3**	12	`-0.173**	14	-0.018	9
Texas Tech University	`-0.166**	5	`-0.09**	5	-0.023	10
Univ. of Houston - Downtown	`-0.436**	27	`-0.213**	21	-0.024	11
Tarleton State University	`-0.356**	17	`-0.192**	17	-0.027	12
UT - Tyler	-0.136	4	-0.087	4	-0.028	13
UT - Austin	`-0.083**	2	`-0.067**	2	-0.038	14
UT San Antonio	`-0.397**	22	`-0.202**	18	-0.052	15
Sam Houston State Univ.	`-0.301**	13	`-0.14**	9	-0.061	16
TAMU - International	`-0.319**	14	-0.1	6	-0.065	17
TAMU - Corpus Chirsti	`-0.372**	19	`-0.207**	20	-0.071	18
Prarie View A&M University	`-0.538**	29	`-0.255**	28	-0.071	19
UT - Pan American	`-0.436**	28	`-0.228**	24	-0.073	20
UT - El Paso	`-0.42**	24	`-0.244**	26	-0.108	21
University of Houston	`-0.239**	8	`-0.161**	12	`-0.116**	22
TAMU - Kingsville	`-0.411**	23	`-0.206**	19	-0.131	23
Lamar University	`-0.339**	15	`-0.229**	25	`-0.164*	24
TAMU - Commerce	`-0.348**	16	`-0.191**	16	-0.166	25
UT - Arlington	`-0.224**	7	`-0.165**	13	`-0.168**	26
UT - Permian Basin	`-0.371**	18	`-0.214**	22	-0.17	27
West Texas A&M Univ.	`-0.374**	20	`-0.217**	23	`-0.198*	28
UT - Dallas	`-0.13**	3	`-0.16**	11	`-0.22*	29
Texas Southern University	`-0.634**	30	`-0.37**	30	`-0.227*	30
SAT score / 100	--		0.0004		0.0002	
At Risk of not Graduating HS	--		`-0.045**		`-0.042**	
Low Income in HS	--		`-0.054**		`-0.054**	
Male	--		`0.069**		`0.072**	
Constant	`9.626**		`9.307**		`9.612**	
Student Level HS Controls	No		Yes		Yes	
High School Level Controls	No		Yes		Yes	
College Major Controls	No		Yes		Yes	
Matched App. Group Controls	No		No		Yes	
Experience Controls	Yes		Yes		Yes	
Observations	45657		45657		45657	
Adjusted R-squared	0.09		0.13		0.14	

Notes: Includes all TX high 1998 & 1999 HS graduates enrolled in a TX 4-yr public college, that took the SAT or ACT, and earned over $2000 8 yrs after graduating HS. The excluded school is Texas A&M. Student level controls include HS courses taken and ethnicity. HS level contols include county average income, unemployment rate, population, and crime rate. ** and * indicate significance at the 5% and 10% level, respectively. TAMU = Texas A&M University. UT = University of Texas

Table 3 - Persistence Rate Models - Probits predicting the probablitiy of persisting into the 2nd year, including dummy variables for 4-year Public schools in Texas

	Unconditional Ranking		*Conditional Ranking*		*Matched Applicant Ranking*	
	1		**2**		**3**	
Variables	Coefficient	Rank	Coefficient	Rank	Coefficient	Rank
Prarie View A&M University	`-0.142**	12	`-0.044**	5	0.086	1
UT - Permian Basin	`-0.174**	19	`-0.129**	23	0.075	2
Univ. of Houston - Downtown	`-0.177**	22	`-0.058**	10	0.064	3
TAMU - Corpus Chirsti	`-0.11**	9	`-0.058**	9	0.057	4
TAMU - Kingsville	`-0.183**	23	`-0.099**	18	0.041	5
Texas State University	`-0.087**	6	`-0.041**	4	0.012	6
UT - Austin	0.009	1	-0.006	2	0.008	7
UT - Dallas	`-0.053**	4	`-0.051**	6	0.004	8
Texas A&M University	--	2	--	1	--	9
Texas Tech University	`-0.042**	3	-0.01	3	-0.004	10
University of Houston	`-0.096**	8	`-0.067**	11	-0.007	11
Texas Women's University	`-0.093**	7	`-0.054**	8	-0.008	12
TAMU - International	`-0.177**	21	`-0.1**	19	-0.013	13
University of Noth Texas	`-0.126**	11	`-0.089**	14	-0.024	14
Texas Southern University	`-0.21**	28	`-0.112**	21	-0.032	15
UT - Pan American	`-0.22**	29	`-0.133**	26	-0.033	16
Sam Houston State Univ.	`-0.163**	15	`-0.097**	17	-0.036	17
UT - El Paso	`-0.163**	16	`-0.096**	16	-0.042	18
Lamar University	`-0.126**	10	`-0.07**	12	-0.044	19
Angelo State University	`-0.177**	20	`-0.108**	20	-0.048	20
Midwestern State University	`-0.164**	17	`-0.093**	15	-0.049	21
UT San Antonio	`-0.206**	27	`-0.137**	27	`-0.065**	22
UT - Arlington	`-0.161**	14	`-0.13**	24	`-0.065*	23
TAMU - Galveston	`-0.083**	5	`-0.053*	7	-0.07	24
Stephen F. Austin State Univ.	`-0.196**	26	`-0.143**	28	`-0.081**	25
UT - Tyler	`-0.166**	18	`-0.146**	29	-0.087	26
West Texas A&M Univ.	`-0.15**	13	`-0.084**	13	`-0.091*	27
Tarleton State University	`-0.191**	24	`-0.122**	22	`-0.092*	28
TAMU - Commerce	`-0.193**	25	`-0.13**	25	`-0.096*	29
Sul Ross State University	`-0.258**	30	`-0.158**	30	`-0.209**	30
SAT score / 100	--		`0.001**		`0.001**	
At Risk of not Graduating HS	--		`-0.03**		`-0.029**	
Low Income in HS	--		`-0.045**		`-0.044**	
Male	--		`-0.045**		`-0.044**	
Constant	`0.942**		`0.822**		`0.885**	
Student Level HS Controls	No		Yes		Yes	
High School Level Controls	No		Yes		Yes	
College Major Controls	No		Yes		Yes	
Matched App. Group Controls	No		No		Yes	
Observations	65305		65305		65305	
Adjusted R-squared	0.05		0.08		0.08	

Notes: Includes all Texas high school graduates in the 1998 and 1999 cohorts that enrolled in a Texas 4 year public college and took either the SAT or the ACT. The excluded school is Texas A&M. Student level controls include HS courses taken and ethnicity. High School level contols include county average income, unemployment rate, population, and crime rate. ** and * indicate significance at the 5% and 10% level, respectively. TAMU = Texas A&M University. UT = University of Texas

Table 4 - Graduation Rate Models - Probits predicting the probablitiy of graduating within 8 years of enrolling, including dummy variables for 4-year Public schools in Texas

	Unconditional Ranking		*Conditional Ranking*		*Matched Applicant Ranking*	
	1		**2**		**3**	
Variables	Coefficient	Rank	Coefficient	Rank	Coefficient	Rank
TAMU - International	`-0.364**	22	`-0.19**	11	0.048	1
Sul Ross State University	`-0.479**	28	`-0.286**	27	0.019	2
Prarie View A&M University	`-0.41**	25	`-0.169**	7	0.006	3
Texas A&M University	--	1	--	1	--	4
TAMU - Kingsville	`-0.397**	24	`-0.214**	15	-0.031	5
TAMU - Galveston	`-0.184**	5	`-0.128**	5	-0.027	5
UT - Austin	`-0.044**	2	`-0.062**	2	`-0.033*	7
TAMU - Corpus Chirsti	`-0.274**	8	`-0.167**	6	-0.038	8
Sam Houston State Univ.	`-0.294**	12	`-0.179**	10	-0.055	9
Texas State University	`-0.202**	6	`-0.125**	4	`-0.057**	10
UT - Pan American	`-0.439**	26	`-0.236**	20	-0.063	11
Texas Tech University	`-0.163**	3	`-0.105**	3	`-0.068**	12
Stephen F. Austin State Univ.	`-0.286**	10	`-0.197**	14	`-0.085**	13
Tarleton State University	`-0.295**	13	`-0.192**	13	-0.092	14
UT - El Paso	`-0.461**	27	`-0.303**	28	-0.093	15
Midwestern State University	`-0.352**	19	`-0.224**	18	-0.094	16
Texas Women's University	`-0.3**	14	`-0.219**	16	-0.098	17
University of Houston	`-0.326**	15	`-0.243**	24	`-0.106**	18
Texas Southern University	`-0.598**	30	`-0.377**	30	`-0.111*	19
UT - Dallas	`-0.183**	4	`-0.172**	9	`-0.115*	20
University of Noth Texas	`-0.287**	11	`-0.222**	17	`-0.118**	21
UT San Antonio	`-0.386**	23	`-0.242**	23	`-0.119**	22
UT - Arlington	`-0.327**	16	`-0.256**	25	`-0.13**	23
TAMU - Commerce	`-0.344**	18	`-0.237**	21	`-0.136*	24
Angelo State University	`-0.36**	21	`-0.24**	22	`-0.137**	25
Univ. of Houston - Downtown	`-0.565**	29	`-0.319**	29	`-0.143**	26
West Texas A&M Univ.	`-0.335**	17	`-0.225**	19	`-0.15**	27
Lamar University	`-0.357**	20	`-0.26**	26	`-0.152**	28
UT - Permian Basin	`-0.279**	9	`-0.191**	12	-0.165	29
UT - Tyler	`-0.205**	7	`-0.17**	8	`-0.277*	30
SAT score / 100	--		`0.002**		`0.002**	
At Risk of not Graduating HS	--		`-0.04**		`-0.037**	
Low Income in HS	--		`-0.065**		`-0.064**	
Male	--		`-0.11**		`-0.107**	
Constant	`0.847**		`0.678**		`0.772**	
Student Level HS Controls	No		Yes		Yes	
High School Level Controls	No		Yes		Yes	
College Major Controls	No		Yes		Yes	
Matched App. Group Controls	No		No		Yes	
Observations	65305		65305		65305	
Adjusted R-squared	0.10		0.15		0.16	

Notes: Includes all Texas high school graduates in the 1998 and 1999 cohorts that enrolled in a Texas 4 year public college and took either the SAT or the ACT. The excluded school is Texas A&M. Student level controls include HS courses taken and ethnicity. High School level contols include county average income, unemployment rate, population, and crime rate. ** and * indicate significance at the 5% and 10% level, respectively. TAMU = Texas A&M University. UT = University of Texas

Table 5 - College Rankings, by Value Added Measure
(rankings determined from Matched Applicant Models)

	Value-Added Measure		
School	Earnings	Graduation	Perrsistence
TAMU - Galveston	1	5	24
Sul Ross State University	2	2	30
Texas State University	3	10	6
Stephen F. Austin State Univ.	4	13	25
Texas Women's University	5	17	12
Midwestern State University	6	16	21
Angelo State University	7	25	20
Texas A&M University	8	4	9
University of Noth Texas	9	21	14
Texas Tech University	10	12	10
Univ. of Houston - Downtown	11	26	3
Tarleton State University	12	14	28
UT - Tyler	13	30	26
UT - Austin	14	7	7
UT San Antonio	15	22	22
Sam Houston State Univ.	16	9	17
TAMU - International	17	1	13
TAMU - Corpus Chirsti	18	8	4
Prarie View A&M University	19	3	1
UT - Pan American	20	11	16
UT - El Paso	21	15	18
University of Houston	22	18	11
TAMU - Kingsville	23	5	5
Lamar University	24	28	19
TAMU - Commerce	25	24	29
UT - Arlington	26	23	23
UT - Permian Basin	27	29	2
West Texas A&M Univ.	28	27	27
UT - Dallas	29	20	8
Texas Southern University	30	19	15