

**LEARNING AND EARNING:
AN APPROXIMATION TO COLLEGE VALUE ADDED IN TWO
DIMENSIONS**

EVAN RIEHL*
JUAN E. SAAVEDRA+
MIGUEL URQUIOLA#

MAY 12, 2016

ABSTRACT. This paper explores the implications of measuring college productivity in two different dimensions: earning and learning. We compute system-wide measures using administrative data from the country of Colombia that link social security records to students' performance on a national college graduation exam. In each case we can control for individuals' college entrance exam scores in an approach akin to teacher value added models. We present three main findings: 1) colleges' earning and learning productivities are far from perfectly correlated, with private institutions receiving relatively higher rankings under earning measures than under learning measures; 2) earning measures are significantly more correlated with student socioeconomic status than learning measures; and 3) in terms of rankings, earning measures tend to favor colleges with engineering and business majors, while colleges offering programs in the arts and sciences fare better under learning measures.

For useful comments we thank Caroline Hoxby and Kevin Stange. All remaining errors are our own.

*Columbia University, +University of Southern California and NBER, #Columbia University and NBER.

1. INTRODUCTION

Colleges produce a variety of outputs for their constituencies. Parents and students, for instance, are often interested in colleges' ability to place graduates on good career tracks. As a result, the U.S. and other countries have started to provide information on the labor market *earnings* of graduates from college/major combinations.¹ A drawback of such measures is that they typically do not adjust for ability—some colleges might perform better, for instance, simply because they attract more able students.

In addition, earnings may be only imperfectly correlated with another college output that parents, students, and especially policymakers might care about: colleges' ability to enhance students' human capital or skill. We will label outputs in this general category as *learning*. Despite interest, measures of learning are less common, in part because most countries do not have nationwide college graduation exams.

This study uses data from the country of Colombia to make two contributions regarding colleges' production of earning and learning. First, we arguably improve upon the measures that have been used in the literature. We use detailed administrative records that provide the earnings of nearly all graduates in the country upon labor market entry. With these data we can control for a measure of ability—performance on a national standardized admission exam—and for characteristics related to students' socioeconomic backgrounds. Further, the Colombian setting allows us to propose and implement measures of college productivity in the learning dimension, as all graduates are required to take a national college *exit* exam. In measuring learning performance we can similarly control for individual characteristics and pre-college ability. In particular, some components of the college exit exam are also assessed in the entrance exam, enabling us to implement an approach akin to those commonly used in the teacher value added literature.² In short, our earning and learning measures may not fully isolate college value added, but they have advantages relative to measures previously used in the context of measuring college productivity.

Our second contribution is to show how these measures of college productivity relate to each other and to characteristics of colleges' entering classes. This yields three findings. First, measures of college productivity on earning and learning are far from perfectly correlated. This implies that college rankings based on earnings differ from those based on learning; in other words, the colleges that seem to add most to students' post-graduation earnings are

¹ Other countries, such as Chile and Colombia, have similar initiatives. These are relevant in view of evidence that, at least in some cases, college identity can have a causal impact on graduates' earnings (e.g., Hoekstra, 2009, Saavedra, 2009, Dale and Krueger, 2014, and MacLeod et al., 2015). This finding is not universal; see Stange (2012) for contrasting findings among community colleges.

² See for instance Chetty et al. (2014). Our empirical approach is also closely related to the one in Saavedra and Saavedra (2011), discussed below.

not necessarily the ones that add most to their measured learning.³ For instance, we find that on average the top private schools seem to do relatively better on earning, whereas the top public institutions perform better on learning.

Second, the measures of earnings productivity are significantly more correlated with student socioeconomic status than the learning measures; not surprisingly, earnings are also more correlated with colleges' tuition levels. This leaves open the possibility that learning measures do a better job of isolating a college's contribution to students' human capital. For example, learning may be more easily influenced by factors that colleges can control directly, such as teaching, as opposed to factors such as parental connections and signaling. Consistent with this, we show that a college's measured performance can vary substantially depending on whether earnings are measured right after graduation or later in workers' careers. This illustrates that colleges have only partial control over the earnings paths of their graduates.

Our third finding is that a college's ranking under the earning and learning measures can differ depending on its composition of majors. We show that the earning measures tend to favor majors related to engineering, business, and law; more specialized majors, such as those in fine arts, education, and social/natural sciences, are relatively higher ranked under learning metrics. Thus if measures like the ones we calculate became salient, they could lead colleges to make strategic choices on which majors they offer.

Our findings relate to multiple areas of work, in part because the desire to measure colleges' inputs and outputs is not new. For instance, the Carnegie Foundation Report (Cooke, 1910) tackled this question over a century ago by developing a time use accounting formula to quantify teaching and research-related activities in higher education (Sullivan et al., 2012). The Carnegie Report proposed the "student hour" metric, which represents "an hour of lecture, of lab work, or of recitation room work, for a single pupil" (Silva, White and Toch, 2015). Analogous metrics are still used at higher education institutions today. One can criticize such metrics because they typically do not account for learning outcomes (e.g., Pascarella and Terenzini, 2005). Beyond this, it is the productivity of educational institutions—their effectiveness at turning inputs into outputs—that is ultimately the key question from an economic point of view (e.g., Hoxby, 2002).

Our study also relates to two strands of work on college productivity: those related to learning and to earning. In terms of learning, a variety of standardized tests exist in the U.S. that could in principle be used to measure student-learning outcomes. These tests include the Measure of Academic Proficiency and Progress (MAPP), the Collegiate Assessment of

³ With learning measures, a concern often arises regarding whether these capture anything that the market and therefore students actually value. In the Colombian setting, student performance on the field-specific component of the exit exam is predictive of student wages, even after controlling for students' performance on the admission exam, college reputation, and socioeconomic status.

Academic Proficiency (CAAP), the Collegiate Learning Assessment (CLA), the California Critical Thinking Skills Test (CCST), the Watson-Glaser Critical Thinking Appraisal, and the Cornell Critical Thinking Tests (Pascarella and Terenzini, 2005; Sullivan et al., 2012). However, these tests are not systematically used across the country.

A few studies use these tests to measure learning outcomes during college, though their focus is typically on learning gains rather than the ranking of institutions. One is Facione and Facione (1997), which uses a cross-sectional application of the CCST in a sample of over 6,000 nursing students across 150 undergraduate institutions. The authors find that the difference in critical thinking skills between seniors and freshman is 0.45 standard deviations. The limitation of this study is that it focuses on students from a single professional program (Pascarella and Terenzini, 2005). Klein et al. (2008) use cross-sectional data from an administration of the CLA test to a sample of freshmen and seniors across 93 U.S. colleges in the Integrated Postsecondary Education Data System (IPEDS).⁴

Saavedra and Saavedra (2011) use a similar cross-sectional design from an administration of Australia's Graduate Skills Assessment (GSA) to estimate educational value added in a nationally representative sample of Colombian college students in their freshman and senior years. The GSA, which is most similar to the CLA in the U.S., measures four general skill domains: critical thinking, problem solving, writing and interpersonal skills. After controlling for incoming student characteristics and selection bias due to attrition, Saavedra and Saavedra find that the average gain across all domains between seniors and freshmen is about 0.3 standard deviations.

To date, only two studies use longitudinal data to assess gains in general skills during college. Mentkowski et al. (1991) use a sample of 135 students from a small liberal arts college and find a freshmen-senior gain of 0.25 standard deviations in the Watson-Glaser Critical Thinking Appraisal. Klein, Steedle and Kugelmas (2010) use a longitudinal administration of the CLA test to students at 50 nationally representative colleges. Only 26 colleges, however, have complete follow-up test administrations and, of those, only 20 had at least 25 students taking the test. They find that the longitudinal gain between freshmen and seniors was between 0.4 and 0.5 standard deviations.

Few studies investigate the extent to which variation in the development of general cognitive skills is related to institutional characteristics. Predergast (1998), Hagedorn et al. (1999), and Saavedra and Saavedra (2011) find that, once controlling for incoming student ability, measures of college quality—like resources, selectivity, and reputation—are not systematically related to value-added. Saavedra and Saavedra, however, find that private ownership

⁴ It is hard to interpret the magnitude of the average difference between freshmen and seniors in Klein et al. (2008) since the institution-level difference is normalized to zero at the typical school. There is, however, wide dispersion in mean institutional differences across colleges in the sample.

is predictive of greater learning. A limitation of each of these studies is the small number of colleges in the sample.

In terms of analyses of earning productivity, our work relates to an increasing number of studies considering labor market outcomes in the U.S. For example, Hoxby and Bulman (2015) use data from the Social Security Administration to estimate college earnings value-added. They leverage quasi-random variation in admission to colleges of different selectivity by focusing on students near expected admission cutoffs. While prior studies have also used regression discontinuity approaches, Hoxby and Bulman expand this estimation strategy to a much wider set of institutions.

Rothwell and Kulkarni (2015) combine administrative and website (e.g., PayScale and LinkedIn) data to compute quality-adjusted value-added for U.S. colleges. The measure they propose is a college-level earnings residual after accounting for college traits such as faculty salaries, graduation rates, and financial aid levels. The main outcome measure is median total earnings for full-time workers with at least ten years of experience. As in Hoxby and Bulman (2015), Rothwell and Kulkarni provide a ranking based on earning productivity, although they lack the ability to leverage individual-level data to focus on marginal applicants.

Cunha and Miller (2014) develop a methodology for measuring college value added which accounts for observable differences in pre-enrollment student characteristics. In addition, they control for potential unobserved differences in students' preferences for schools, and schools' preferences for students, by including fixed effects for a student's application and acceptance profile.⁵ Cunha and Miller implement this approach using administrative data from Texas and find that colleges' value-added measures change considerably upon controlling for student characteristics and preferences.

No single study to date, however, simultaneously combines system-wide longitudinal data on learning and earning—our aim is to contribute by addressing that gap. The remainder of the paper is structured as follows. Section 2 presents background on the Colombian higher education sector, and Section 3 describes our data and sample. Section 4 discusses the computation of our productivity measures, and Section 5 presents results. Section 6 concludes with broader implications.

2. BACKGROUND

This section provides background on Colombia's higher education system.

2.1. Access to college. In the past decades, Latin American countries have seen a marked expansion in access to secondary and tertiary education. Access to the latter has actually risen faster, although from a lower base. As Figure 1 shows, the gap between secondary

⁵ This is analogous to the approach in Dale and Krueger (2002).

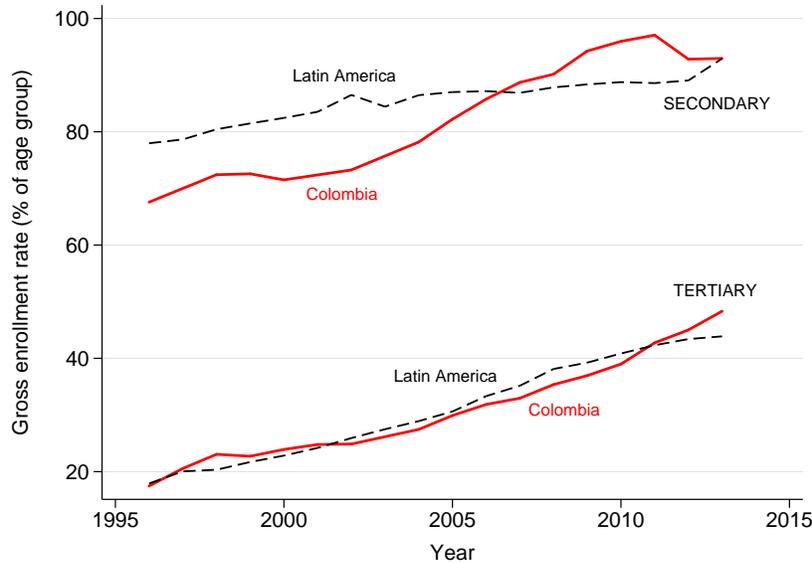


FIGURE 1. Enrollment trends in Colombia and Latin America

Notes: The data come from the World Bank indicators (<http://databank.worldbank.org>, consulted on April 7, 2016). The figure plots gross secondary and tertiary enrollment rates for Colombia and the corresponding aggregate for the Latin America as a whole. Gross secondary enrollment rate is the number of individuals enrolled in secondary school as a fraction of the total number of individuals 12 to 17 years of age. Gross tertiary enrollment rate is the number of individuals enrolled in tertiary education as a fraction of the total number of individuals 18 to 24 years of age.

and tertiary enrollment in the region narrowed from 60 percentage points in 1996 to 50 percentage points by 2013. By this year, about 43 percent of the population had enrolled in some type of tertiary education. The evolution in Colombia has generally mirrored that in the rest of the region, although the gap between both types of enrollment has remained stable at about 45 percentage points.⁶

Throughout the region, there are constraints for further tertiary expansion. In the case of Colombia these partially reflect market structure. Private and public providers co-exist, and while public colleges are significantly subsidized, their capacity is strained. Table 1 shows that public colleges account for 23 percent of institutions but 52 percent of total tertiary enrollments.⁷

⁶ The salient difference between Colombia and the rest of the region is that secondary rose faster initially and then stagnated. Tertiary enrollment trends are essentially identical in Colombia and the region as a whole.

⁷ Throughout this paper we use the term “colleges” to refer to both universities and technical institutions, as depicted in Table 1.

TABLE 1. Colombian higher education market structure

	Institutions			Enrollment		
	Public	Private	Total	Public	Private	Total
Universities	47 0.17	142 0.53	189 0.70	495,855 0.25	799,673 0.40	1,295,528 0.65
Technical schools	15 0.06	65 0.24	80 0.30	524,007 0.27	163,886 0.08	687,893 0.35
Total	62 0.23	207 0.77	269 1.00	659,142 0.52	601,744 0.48	1,983,421 1.00

Notes: Calculations based on the Colombian national higher education information system (SNIES) for 2013, the last year with data available. Enrollment data only includes undergraduate students. The category “universities” combines universities and university institutes. Technical schools combines technical institutes, technological institutes, and the National Job Training Agency (SENA).

There is little regulation on the entry of tuition-charging, unsubsidized private providers, and these generally offer few financial aid opportunities.⁸ As a result, private colleges represent 77 percent of all institutions but only 48 percent of total enrollment.

Colleges and universities are also geographically concentrated: 50 percent are in Colombia’s three largest cities, which account for 26 percent of the population. Bogotá, the capital, is home to 31 percent of all colleges. In addition about 75 percent of tertiary students attend a college in the city of their birth (Saavedra and Saavedra, 2011).

2.2. College entrance exam. To apply to college, Colombian students must take a standardized entrance exam called the Icfes, which is administered by a government agency.⁹ The Icfes is generally analogous to the SAT in the U.S., but it is taken by the vast majority of high school seniors regardless of whether they intend to apply to college.¹⁰ The Icfes also plays a larger role in admissions in Colombia than the SAT does in the U.S. In addition to using it as an application requirement, many schools extend admission offers based

⁸ Technically there are no for-profit colleges in Colombia. It is widely perceived, however, that many non-selective private colleges are *de facto* for-profit, as their owners are the residual claimants of excess revenue typically distributed through wages, rental charges, investments, etc. In this sense the situation resembles that which has existed during certain periods in other countries with large private college sectors, such as Chile.

⁹ Icfes stands for Institute for the Promotion of Higher Education, the former acronym for the agency that administers the exam. The Colombian Institute for Educational Evaluation, as it is now called, was created in 1968 and is a State agency under the authority of the national Ministry of Education. The Icfes exam is now known as Saber 11°, reflecting the fact that students usually take it in the 11th grade. We use the name Icfes to match the designation during the period covered by our data.

¹⁰ Angrist et al. (2006) and our personal communications with the Colombian Institute for Educational Evaluation suggest that more than 90 percent of high school seniors take the exam. The test-taking rate is high in part because the government uses Icfes exam results to evaluate high schools.

solely on students' entrance exam performance. Others consider additional factors like high school grades while heavily weighting the Icfes, and a handful administer their own exams. Applications and admissions are major-specific; students apply to a college/major pair.

The Icfes tests multiple subject areas including biology, chemistry, English, math, reading/language arts, social science, philosophy, and physics.

2.3. College exit exam. In 2004 the agency that administers the Icfes introduced, with considerable publicity, new field-specific college *graduation* exams. These exit exams are standardized and administered at every institution that offers a related program.¹¹ The exams are intended to assess senior students' competencies in fields ranging from relatively academic in orientation (e.g., economics and physics) to relatively professional (e.g., nursing and occupational therapy).

The creation of the exit exams was a major undertaking, as it required coordination among departments in multiple colleges. The stated intent of this effort was to improve quality, transparency, and accountability in the higher education sector. Consistent with this, school-level aggregate scores were made available and have been used by news outlets as part of college rankings.

Field-specific exams became available for most majors in 2004, with several majors receiving field exams in subsequent years. A few fields such as political science, anthropology, history, and philosophy never received a corresponding field-specific exam. In part because of this, for the first few years taking the exit exam was optional, although the majority of students in tested fields took the exam. This changed in 2009, when the exit exam became a graduation requirement for all students. A generic test was introduced for majors that did not previously have a field-specific exam. In addition, from 2009 onward the exam included several common components in subjects such as English and reading comprehension, which were taken by all students regardless of their field.

Increasingly, colleges and students use results on the college exit exam as a signal of ability. For example, students may report whether they obtained a top score nationally, or their score in comparison to the university or the national average. Some universities use exit exam results in admissions to graduate programs, and the Colombian Student Loan Institute offers a postgraduate study credit line (of up to 16,000 dollars) exclusively to the best ten nationwide scorers. In addition, every year the Colombian President and Education Minister publicly recognize the individuals with the top ten scores in each field. Anecdotally, the best scorers receive job offers based on public knowledge of their test scores, and MacLeod et al. (2015) provide evidence that the exit exams affect graduates' labor market earnings.

¹¹ These tests were initially labeled Ecaes, which stands for *Exámenes de Calidad de Educación Superior*, i.e., higher education quality exams. They are now called *Saber Pro*.

3. DATA AND SAMPLE

This section describes our sources of data and the sample we use for our analysis.

3.1. **Data.** We use individual-level administrative datasets from three sources:

- (1) The Colombian Institute for Educational Evaluation, which administers the college entrance and exit exams, provided records for both tests. This includes scores for all high school seniors who took the entrance exam between 1998 and 2012, as well as college exit exam scores for all exam takers in 2004–2011.
- (2) The Ministry of Education provided enrollment and graduation records for students entering college between 1998 and 2012. These include each individual’s college, program of study, and enrollment and graduation dates. These data cover roughly 90 percent of all college enrollees; the Ministry omits a number of smaller colleges due to poor and inconsistent reporting.
- (3) The Ministry of Social Protection provided monthly earnings records for formal sector workers during 2008–2012. These come from data on contributions to pension and health insurance funds.

We link these data sources using student names, birthdates, and national ID numbers. The resulting dataset includes students from nearly all colleges in Colombia, with information on their entrance exam scores and, if applicable, their exit exam performance and formal labor market earnings.

3.2. **Sample.** We select a sample that allows us to cleanly compare measures of college performance on earning and learning. Thus we set aside other relevant outcomes, such as graduation, and focus on college graduates for whom we observe both exit exam scores and formal labor market earnings.

Specifically, we restrict our sample to graduates who satisfy two important criteria. First, we include only students who took the college exit exam in 2009–2011. As noted above, the exit exam was voluntary prior to 2009, so we exclude pre-2009 exam takers to limit selection into taking the exam. Second, we include only graduates for whom we observe initial labor market earnings. Since students typically take the exit exam one year before graduating, this means that we include only 2010–2012 graduates with earnings observed in their graduation year.

In addition to these restrictions, we drop individuals with missing values on any of the other variables we use, including entrance exam scores, high school of origin, mother’s education, and previous year’s tuition.¹² This ensures that all performance measures calculated below

¹² The entrance exam underwent a major overhaul in 2000, and so we also exclude the small number of students who graduated in 2010–2012 but took the entrance exam prior to 2000. Since one of our learning outcomes below is a student’s English exit exam score, we additionally drop the fewer than one percent of

TABLE 2. Sample and college types

College type	No. of colleges	No. of grads	Admit rate	Annual tuition	Mother went to college	Entrance exam pctile
Public (most selective)	12	15,642	0.20	\$369	0.42	0.82
Public (medium selective)	24	13,228	0.55	\$509	0.29	0.67
Public (least selective)	12	6,063	0.87	\$535	0.23	0.59
Top private	8	9,653	0.64	\$2,584	0.90	0.90
Other private (high cost)	51	19,229	0.82	\$1,696	0.59	0.72
Other private (low cost)	50	17,489	0.86	\$1,079	0.31	0.63
Total	157	81,304	0.65	\$1,134	0.46	0.72

Notes: Admission rate data are from Colombian national higher education information system (SNIES) and average over 2007–2012. Tuition data are from the exit exam records, which report each exam takers’ tuition in the previous year in six categories. We compute the average across all students using the midpoint of each category and convert to U.S. dollars using 2012 exchange rates. Entrance exam percentiles are relative to all exam takers in each year, including those who did not attend college.

are based on the same set of individuals. Lastly, to obtain reasonable precision for each of our performance measures, we restrict our analysis to colleges that have at least 50 graduates satisfying the above criteria.

The resulting sample includes approximately 81,000 graduates from 157 colleges. This is much larger than samples available in previous studies that use longitudinal data to compute college performance measures (e.g., Klein, Steedle and Kugelmas, 2010). The last row in Table 2 presents summary statistics on our sample.

3.3. College categorization. Table 2 additionally categorizes colleges into six types with the aim of providing a useful portrayal of the college market in Colombia. The top three rows separate public colleges into three groups based on quartiles of their admission rates. We define the most selective public colleges as those in the quartile with the lowest admission rates, and the least selective colleges as those in the highest admission rate quartile. Medium selective colleges are those in the middle two quartiles.¹³ Table 2 shows that the most selective public colleges admit 20 percent of their applicants on average, while the least selective are essentially open enrollment.¹⁴

students who took the French or German entrance exams, which were offered until 2006, rather than the English exam.

¹³ We use quartiles rather than terciles to define these three groups to provide more detail on colleges at the extremes of the distribution.

¹⁴ Note that non-selective colleges often have admission rates that are slightly less than one in Table 2. This reflects that students may fail to follow all application procedures or may withdraw their applications before admission.

Selectivity defined by admission rates has limited usefulness in categorizing private colleges in Colombia, as most private colleges admit nearly all of their applicants. Instead, sorting into private colleges is defined more strongly by the tuition rates they charge. We therefore define “top private” colleges as those few that are actually selective—i.e., they reject some of their applicants—and in which average annual graduate tuition exceeds the equivalent of about twenty-five hundred dollars.¹⁵ This definition picks out eight colleges which represent the most elite private schools in the country. We divide the remaining private institutions—which we label “other private”—into two types based on the average tuition payments reported by their graduates. We define high cost private colleges as those above the median tuition, and low cost colleges as those below.¹⁶

Average annual tuition varies significantly across private college types, with a mean of roughly one thousand dollars at low cost private colleges. Average tuition is significantly lower at all public college types, as they offer substantial discounts to low SES students.

The last two columns of Table 2 summarize the socioeconomic and academic backgrounds of graduates from each college type. Graduates from private colleges are much more likely to have mothers with a college education. For instance, 90 percent of students at top private colleges do so. Academic preparation, as defined by each student’s entrance exam percentile in the full distribution of test takers, also varies starkly across college types. Average entrance exam performance is at the 82nd percentile at the most selective public colleges and the 90th percentile at top private schools. Graduates from the lowest college types, both public and private, have average entrance exam scores near the 60th percentile.

We use the sample and college categorization in Table 2 for our analysis of college performance measures below.

4. MEASURES

This section describes the outcome variables we use, and the measures we employ to approximate college earning and learning productivity.

4.1. Earning and learning variables. Our earnings variable is log average daily formal labor market earnings, which we calculate by dividing base monthly earnings for pension contributions by the number of employment days in each month and averaging across the

¹⁵ Specifically, we use a four million peso cutoff for top private colleges, and we define their selectivity using a 2002 report from the Colombian Institute for Educational Evaluation entitled *Estadísticas de la Educación Superior*. Selective private colleges as those for which the number of applicants exceeded the number of offered slots, according to this report.

¹⁶ We note that we do not use an institution’s level of training (university or technical, as in Table 1) to define these six college categories. We find that this distinction provides little additional information on average college characteristics conditional on the categories defined by financing, selectivity, and tuition.

year. We use earnings in the year of each student’s graduation (2010–2012) and demean earnings in each year.

Our learning variables are based on students’ scores on the college exit exam. During the exam years we analyze (2009–2011), this test included a field-specific component related to a student’s major (e.g., economics or mechanical engineering) as well as several components taken by all students. We focus on three of these: i) the field-specific score, ii) a reading common component score, and iii) an English common component score.

These components have different strengths and weaknesses in measuring college productivity. The field exit score, because it typically reflects each student’s college major, provides arguably the best measure of the material studied in college. However, in general there is no direct analog on the entrance exam. The English component of the exit and entrance exams are very similar and thus well placed to measure progress, but English proficiency may be less directly related to college productivity. Since the exit and entrance exams include a similar but not identical reading/language arts component, the reading component lies arguably in the middle of the comparability and relevance spectrums.

Using these three exit exam scores, we calculate each student’s percentile relative to all other students in our sample in the same exam field and cohort. We use exam score *percentiles* because the entrance and exit exams are not on a common scale and thus cannot measure growth in human capital. As a result, our learning measures will capture a college’s relative rather than absolute performance. The same caveat applies to our earning measures since we do not observe a pre-college measure of earnings.

4.2. Calculation of productivity measures. We use four procedures to measure learning and earning performance. Some of these procedures are simple and require less-detailed information, and thus they correspond to measures that may be more commonly reported in the media or easier for policymakers to compute. Other procedures use comprehensive information on students’ backgrounds and align more closely with “value-added” methods employed in other areas of economic research. These four procedures, which we describe in the following subsections, allow us to explore the sensitivity of our results to different data requirements and methodologies.

4.2.1. Raw means. Our first performance measure is the average log earnings, or the average exit exam percentile, at each college:

$$(1) \quad \bar{\theta}_c = E\{y_{ic} | i \in c\},$$

where y_{ic} is either outcome for individual i who graduated from college c . We label $\bar{\theta}_c$ the *raw means* measure, as it implements the simplest and least data-intensive of our four

procedures. Note that it does not adjust for differences across colleges in incoming student characteristics—i.e., in the student “inputs” to college production.

4.2.2. *Entrance exam residuals.* Our second performance measure adjusts for differences in college inputs by controlling for students’ entrance exam performance. We do this through an individual-level regression of the following form:

$$(2) \quad y_{ic} = \beta' t_i + \tilde{\theta}_c + \tilde{\epsilon}_{ic},$$

where t_i is a vector of student i ’s entrance exam percentiles on eight components, which include reading/language arts and English.¹⁷ We decompose the residual from this regression into a school-specific term, $\tilde{\theta}_c$, and an idiosyncratic component, $\tilde{\epsilon}_{ic}$. Our second college productivity measure, which we call *entrance exam residuals*, is the $\tilde{\theta}_c$ coefficient from equation (2).

4.2.3. *Entrance exam + SES residuals.* Our third performance measure is closely related to the second, but we include additional controls for students’ socioeconomic background in regression (3):

$$(3) \quad y_{ic} = \beta' t_i + \gamma' x_i + \hat{\theta}_c + \hat{\epsilon}_{ic},$$

where x_i represents dummies for four categories of mother’s education (primary, secondary, vocational, university), which are fully interacted with dummies for each of the approximately six thousand high schools in our sample. The *entrance exam + SES residuals* measure for each college is the $\hat{\theta}_c$ coefficient from this regression. This coefficient is identified from variation in college attendance across students with the same high school and mother’s education combination. This measure is most analogous to benchmark “value added” models in other work in economics, which control for a broad array of initial individual characteristics.

4.2.4. *College-level residuals.* Our fourth performance measure is the residual from a college-level rather than an individual-level regression. Specifically:

$$(4) \quad y_c = \beta' t_c + \theta_c,$$

where y_c is the average exit exam percentile or log earnings at college c , and t_c is the vector of college mean percentiles for each of the eight entrance exam components.¹⁸ The *college-level residuals* measure is the residual from regression (4), θ_c . As we discuss below, this measure has properties that differ from those of measures based on individual residuals because it is uncorrelated with college mean entrance scores by construction.

¹⁷ The other components are biology, chemistry, math, social sciences, philosophy, and physics. As with the exit exam scores, we convert entrance exam scores into percentiles within each exit exam field and cohort.

¹⁸ Observations in regression (4) are weighted by the number of graduates from each college. All college-level computations in this paper use these same weights.

4.3. Correlations of productivity measures with inputs. For our earning and each of our three learning variables, the above procedures yield four separate productivity measures—in short, 16 measures for each college in our sample. We normalize each of these to have mean zero and standard deviation one across the 157 colleges. This normalization is convenient because it makes the coefficient from a linear regression of one measure on another equal to their pairwise correlation coefficient.

To provide context on these measures, we show how they relate to a college characteristic that is in principle easily observable to many agents: colleges’ mean entrance exam score. We begin with a graphical exposition using only one learning outcome: the field-specific exam score. The four panels of Figure 2 depict our four measures for this outcome. The grey circles are the 157 colleges in our sample. The vertical axis in each panel represents the learning performance under each measure, while the horizontal axis depicts the raw mean entrance exam score at each college.¹⁹ The solid line depicts the linear relationship between these two measures, with the slope indicated on the graph.

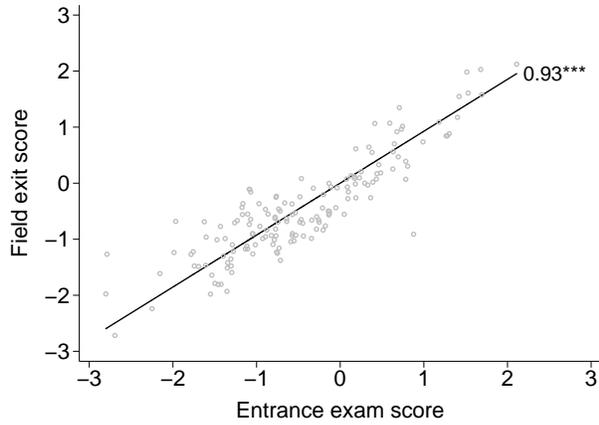
Panel A shows that the correlation between a college’s raw mean field exit score ($\bar{\theta}_c$ from equation (1)) and its mean entrance exam score is 0.93. Panel B shows that controlling for individual entrance exam scores (using $\tilde{\theta}_c$ from equation (2)) reduces this correlation only slightly. Note that while $\tilde{\theta}_c$ ensures that *individual* exit residuals are uncorrelated with *individual* entrance exam scores, it allows *college-level* exit scores to be correlated with *college-level* entrance exam performance. This can arise if other individual characteristics that affect exit exam performance, such as socioeconomic background, also affect the colleges students choose to attend.

Panel C partially addresses this issue by using the entrance exam + SES residual measure ($\hat{\theta}_c$ from equation (3)), which controls for students’ observable background. Panel C shows that these controls have little effect on the correlation of the exit field score with college mean entrance exam performance; in fact, the correlation coefficient increases slightly. This illustrates that our individual learning productivity measures may still be correlated with unobservable student characteristics that affect both college choice and exit exam performance.

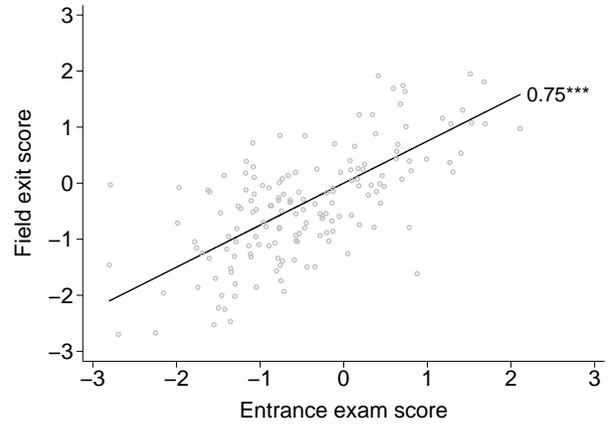
Panel D illustrates that our last productivity measure, the college-level residual (θ_c from equation (4)), is uncorrelated with college mean entrance exam performance by construction.²⁰ This addresses the issue that individual characteristics may be correlated with college mean entrance scores, but allows these individual characteristics to be correlated with other

¹⁹ Raw mean entrance score is the average percentile across the same eight components included in regressions (2)-(4), also normalized to mean zero and standard deviation one.

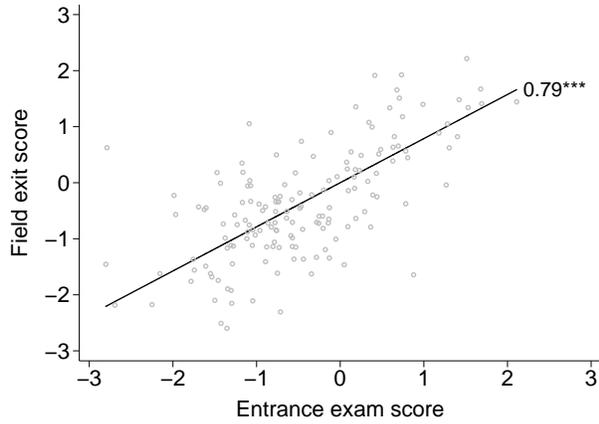
²⁰ The correlation between the two measures in Panel D is not strictly zero because the horizontal axis is the average of the eight entrance exam components, not any individual component from regression (4).



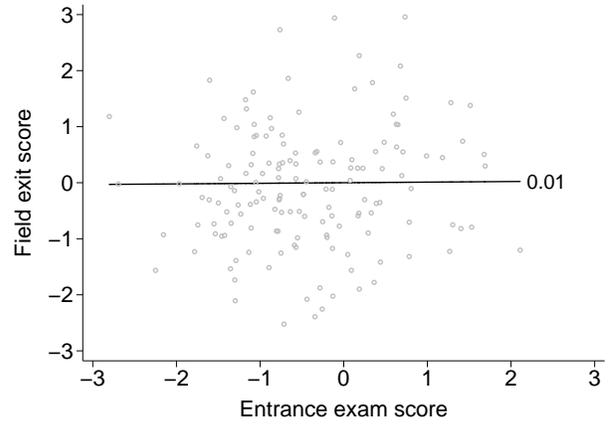
Panel A. Raw means



Panel B. Entrance exam residuals



Panel C. Exam + SES residuals



Panel D. College-level residuals

FIGURE 2. Illustration of field-specific learning measures

Notes: Grey circles represent the 157 colleges in our sample. The solid line depicts the linear relationship between the learning measures and college mean entrance scores, with colleges weighted by their number of graduates. Stars on the slope coefficients indicate statistical significance with robust standard errors.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

college mean traits (e.g., college mean SES). In addition, the college residual measure, θ_c , has a potential disadvantage not present in the individual residual measures. Specifically, it rules out the possibility that colleges with high mean entrance scores systematically produce better learning outcomes than colleges with low average scores. Rather, this measure is better suited for comparing the performance of colleges with similar inputs as defined by mean entrance scores.

As stated we have 16 outcome measures in total (log earnings plus three learning measures, each calculated using the procedures in equations (1)-(4)). Table 3 displays the correlations

TABLE 3. Correlations with college mean entrance scores

	(A)	(B)	(C)	(D)
	Raw	Entrance	Exam	College-
	means	exam	+ SES	level
		residuals	residuals	residuals
Field exit score	0.93***	0.75***	0.79***	0.01
Reading exit score	0.90***	0.59***	0.65***	-0.00
English exit score	0.88***	0.73***	0.71***	0.02
Log earnings	0.70***	0.63***	0.57***	0.13

Notes: This table displays coefficients from linear regressions of college mean entrance exam scores on each of our 16 learning and earning measures. All regressions have 157 observations with weights equal to each college's number of graduates. Stars indicate statistical significance with robust standard errors.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

of each of these measures with college mean entrance scores. The top row refers to the field exit score and replicates the correlation coefficients depicted in Figure 2. The remaining three rows cover the other measures. The manner in which the correlation measures change as one moves across columns is similar across all rows; in other words, the above discussion applies to all our of learning and earning measures. This provides an additional justification for using multiple methods to calculate productivity in examining our key findings below.

5. RESULTS

This section presents empirical results related to three questions: 1) How are the earning and learning measures related to each other? 2) How are they related to other factors that influence students' choice of colleges? 3) How do these measures vary with the majors a college offers?

5.1. Comparing learning and earning measures. Our first empirical task is to explore how the learning and earning measures relate to each other. Table 4 shows the correlation coefficients for each of our three learning measures with our earning measure, where each has been calculated according to the procedure listed in the column.

A simple but important result is that the learning measures are positively related to our earning measure, but far from perfectly so, with correlations ranging from 0.01 to 0.71 across the learning outcomes and the four procedures. The raw mean learning and earning measures are more strongly correlated than those that control for individual characteristics. The college-level residual measures are only weakly related, with correlation coefficients that are not statistically different from zero. It is also notable that the English learning measures

TABLE 4. Correlations with earning measure

	(A)	(B)	(C)	(D)
	Raw	Entrance	Exam	College-
	means	exam	+ SES	level
		residuals	residuals	residuals
Field exit score	0.62***	0.45***	0.45***	0.09
Reading exit score	0.58***	0.29***	0.41***	0.09
English exit score	0.71***	0.62***	0.51***	0.01

Notes: This table displays coefficients from linear regressions of our earning measures on each of our learning measures. All regressions have 157 observations with weights equal to each college’s number of graduates. Stars indicate statistical significance with robust standard errors.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

are generally more correlated with earnings, which may reflect a stronger socioeconomic component to English education relative to the other subjects.

Figure 3 depicts the relation between the earning measures (vertical axis) and the field-specific learning measures (horizontal axis). The imperfect correlations from Table 4 are evident here in the dispersion of the dots, which is most prevalent for the college-level residual method in Panel D. Each panel also contains a 45 degree line that represents the boundary between whether colleges appear more productive on the learning or earning measures. In all four panels, the most selective public colleges (indicated by the light red triangles) typically lie below the diagonal line—these colleges appear in a more favorable light when we define productivity by learning. Conversely, top private colleges (dark blue squares) mostly lie above the 45 degree line; this means that they appear in a more favorable light when performance is defined in terms of earnings. Note that these conclusions hold across all four procedures for calculating productivity despite the different properties discussed above.

Table 5 elaborates on this point by presenting the average institution rank that arises from the use of learning or earning measures. Specifically, we sort colleges according to each measure and calculate their percentile rank among the 157 schools. We then compute the average rank in each of the six college types defined in Table 2. We repeat this calculation for the field-specific learning measures and the earning measures from the entrance exam residual method (Panel B) and the college-level residual procedure (Panel D).²¹ For instance, using the field exit score and individual entrance exam residuals, the most selective public colleges have an average rank at the 88th percentile, while the average rank of a top private college is the 89th percentile.

²¹ The panel labels in Table 5 correspond to the method used, i.e., to the columns in Table 4.

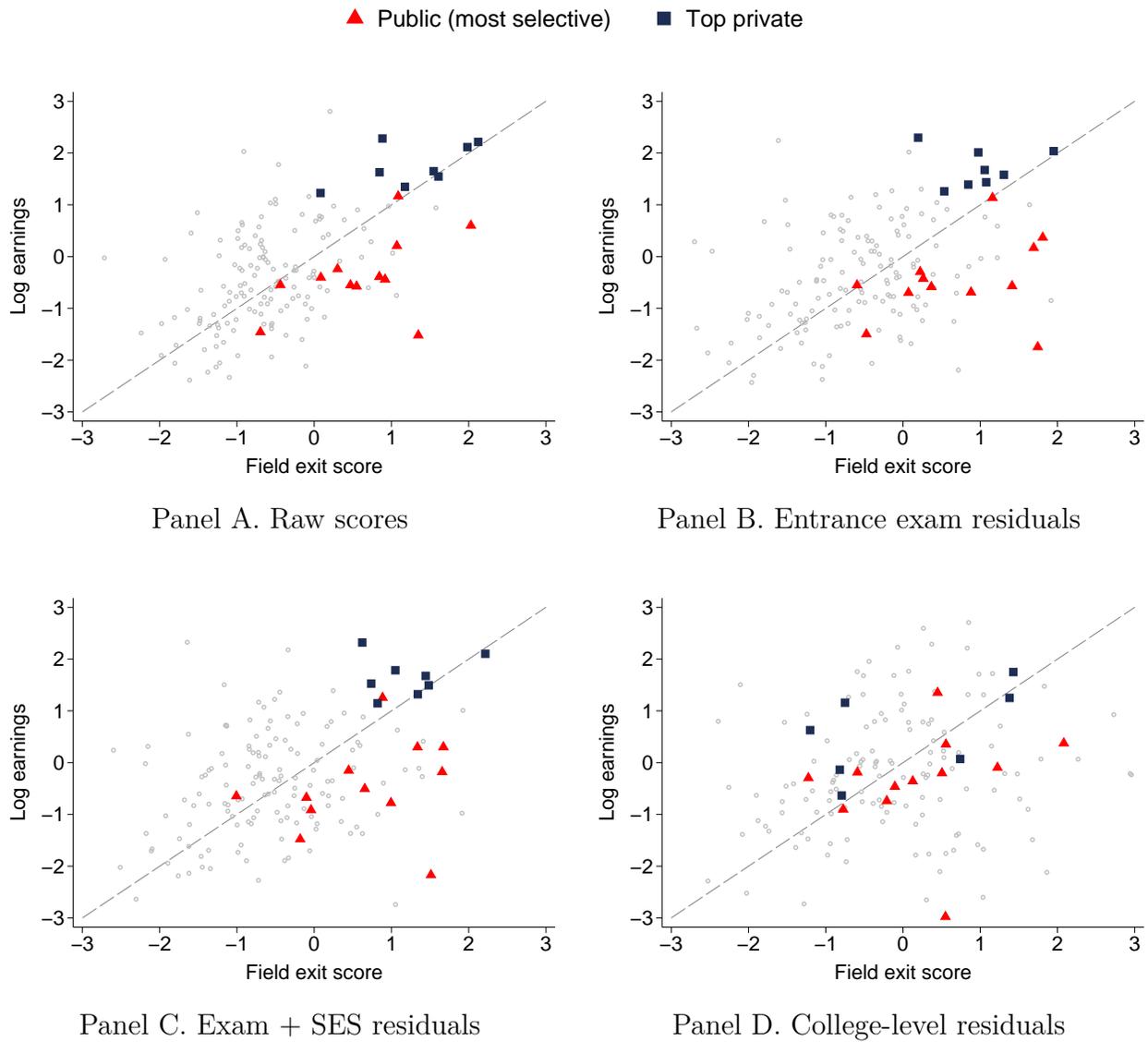


FIGURE 3. Earning vs. field-specific learning

Notes: Light red triangles represent the most selective public colleges as defined in Table 2. Dark blue squares represent top private colleges, and grey circles depict all other colleges.

The main conclusion from Table 5 is that public colleges receive higher rankings from the learning measures than from the earning measures. Conversely, private colleges are relatively higher ranked using earnings. This finding holds using both the individual- and college-level residual procedures for nearly all colleges types.

The different measures can thus lead to starkly different conclusions about colleges' relative productivity. In Panel B, for example, high cost private colleges are ranked higher on average than the most selective public colleges using earnings, but their average rank is 25 percentile points lower using the learning measure. As discussed above, comparisons of colleges with

TABLE 5. Average institution rank by college type

College type	Panel B. Entrance exam residuals		Panel D. College-level residuals	
	Field exit score	Log earnings	Field exit score	Log earnings
Public (most selective)	0.88	0.58	0.64	0.54
Public (medium selective)	0.54	0.44	0.43	0.46
Public (least selective)	0.26	0.20	0.39	0.35
Top private	0.89	0.95	0.42	0.65
Other private (high cost)	0.63	0.70	0.62	0.60
Other private (low cost)	0.36	0.49	0.45	0.55

Notes: This table displays percentile ranks of colleges using the measures listed in the column header. We sort all colleges according to each measure, and then calculate average ranks within the college types depicted in Table 2. Averages are weighted by each college’s number of graduates.

different mean entrance scores are more complicated under the college-level residual method of Panel D. Nonetheless, a similar conclusion applies to the relative rankings of the most selective public colleges and top private colleges, which have similar mean entrance scores (see Table 2). Top private colleges receive higher ranks under the earning measure, while selective public colleges appear more favorably when one uses the learning measure.

5.2. Correlations with other college characteristics. The fact that the learning and earning measures are not perfectly correlated suggests that they likely have different relationships with other student and college characteristics. In this section we explore how learning and earning productivity are related to two other factors that influence students’ college choice. We first consider socioeconomic status as defined by whether a student’s mother attended college. We then consider a proxy for student demand: each graduate’s annual tuition in the prior year.

For both the SES and tuition variables, we follow the same procedures described in Section 4.2 to compute college averages. This yields measures of college mean SES and college mean tuition corresponding to the raw means, entrance exam residuals, entrance exam + SES residuals, and college-level residuals methods. Note that we do not present the SES measures from equation (3) as this method includes SES controls also defined by mother’s education. As above, we normalize each measure to mean zero and standard deviation one across the sample of 157 colleges.

Figure 4 displays the correlations of SES with the field-specific learning measures and the earning measures. In all cases, the earning measures are more strongly correlated with

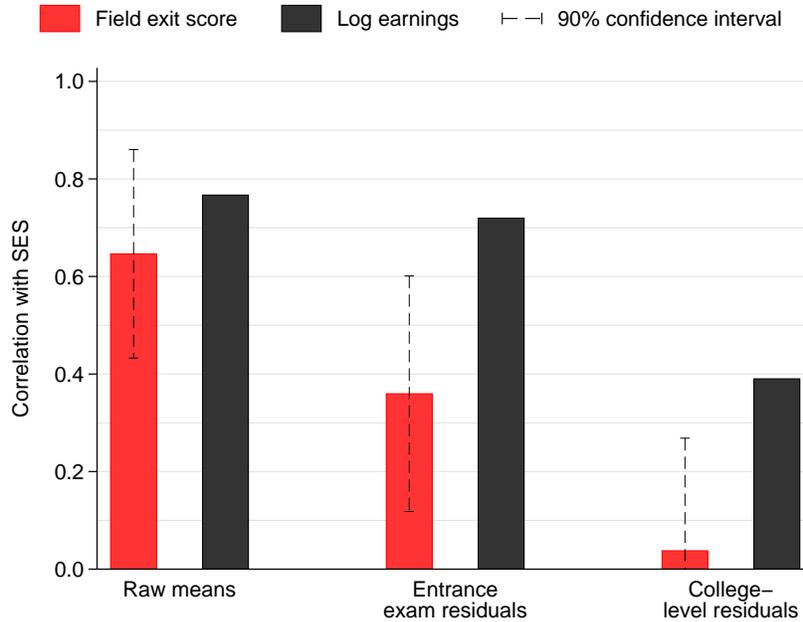


FIGURE 4. Correlations with SES

Notes: Light red bars depict the correlations of our SES measures with our field-specific learning measures (the first row in Table 6). Black bars show the correlation of our SES measures with our earning measures (the fourth row in Table 6). Dashed lines are 90 percent confidence intervals using robust standard errors.

SES than that learning measures, though the difference between the two is not statistically different from zero using raw means.

Table 6 presents these correlations for all of our learning and earning measures. The top panel displays the correlation of the measures with college mean SES, while the bottom panel displays the difference between each learning measure and the earning measure. In nearly all cases, the learning measures are less correlated with SES than the earning measures, and this difference is statistically significant using the two residual methods (columns (B) and (D)). The only exceptions arise with two of the English learning measures, which, as noted above, may be more influenced by socioeconomic background than the field and reading scores.

Table 7 is analogous to Table 6, but it presents the correlations of learning and earning measures with tuition rather than with SES. The same pattern holds; the learning measures are in all cases substantially less correlated with graduates' average tuition than the earning measures.

The results in Tables 6 and 7 are consistent with a college's earning performance being a stronger driver of its demand than its learning performance. Though none of our measures may fully isolate college value added, these findings suggest that learning measures may be less related to other factors that affect student outcomes, which may not be observable in all contexts. This is particularly relevant if learning outcomes are ultimately under greater

TABLE 6. Correlations with SES

		(A)	(B)	(D)
		Raw	Entrance	College-
		means	exam	level
			residuals	residuals
<i>Correlations</i>	Field exit score	0.65***	0.36***	0.04
	Reading exit score	0.59***	0.16	0.08
	English exit score	0.83***	0.75***	0.20***
	Log earnings	0.77***	0.72***	0.39***
<i>Differences from earnings</i>	Field exit score	-0.12	-0.36**	-0.35**
	Reading exit score	-0.18	-0.56***	-0.31**
	English exit score	0.07	0.03	-0.19*

Notes: The top panel displays coefficients from linear regressions of SES (defined by mother’s education) measures on each of our learning and earning measures. All regressions have 157 observations with weights equal to each college’s number of graduates. The bottom panel shows the difference between each of the learning coefficients and the earnings coefficient. Stars indicate statistical significance with robust standard errors.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

control on the part of colleges than earning results. In particular, earning measures, unlike those based on learning, have a natural dynamic component in the years after students enter the labor market. Throughout our analysis we have used earnings measured in the year of each student’s graduation, but there are both conceptual and data-related reasons why earnings might be measured later in a worker’s career.

To explore the potential implications of the timing of earning measurement, we use a different sample than in the above analysis that allows us to measure earnings later in workers’ careers. Specifically, we include 2003–2012 graduates with earnings observed in 2008–2012. With this sample we can observe earnings between zero and eight years of potential experience, defined as earnings year minus graduation year.²² Note that this analysis relies on cross-cohort earning comparisons, meaning that the sample differs across experience levels.

The earning measures analyzed above normalize measures to have a constant standard deviation. Before computing such measures, we display the raw data in Figure 5. This figure shows average log earnings at the 128 colleges that we observe at all experience levels, where we demean earnings by graduation cohort and year. We group the 128 colleges into three terciles of different shadings based on their average earnings at experience zero and hold these terciles constant for all experiences levels.

²² We can actually observe a ninth year of potential experience using 2012 earnings for 2003 graduates, but these ninth-year measures are noisy because they come from only a single cohort and year.

TABLE 7. Correlations with tuition

	(A)	(B)	(C)	(D)	
	Raw	Entrance	Exam	College-	
	means	exam	+ SES	level	
		residuals	residuals	residuals	
<i>Correlations</i>	Field exit score	0.32*	0.16	0.24	0.03
	Reading exit score	0.24	-0.05	0.10	0.05
	English exit score	0.59***	0.63***	0.54***	0.08
	Log earnings	0.67***	0.67***	0.60***	0.48***
<i>Differences from earnings</i>	Field exit score	-0.36*	-0.52***	-0.36*	-0.44***
	Reading exit score	-0.44**	-0.72***	-0.50***	-0.42***
	English exit score	-0.08	-0.04	-0.06	-0.39***

Notes: The top panel displays coefficients from linear regressions of tuition (defined as in Table 2) measures on each of our learning and earning measures. All regressions have 157 observations with weights equal to each college’s number of graduates. The bottom panel shows the difference between each of the learning coefficients and the earnings coefficient. Stars indicate statistical significance with robust standard errors.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 5 shows that the spread in average earnings between the highest and lowest colleges increases with worker experience, a result first documented by MacLeod et al. (2015). At experience zero, nearly all colleges have average earnings within 30 percent of the mean, while many colleges lie outside this range after eight years. Further, there is substantial mixing of the terciles over time, such that some colleges with low initial earnings ultimately have mean earnings above those of top tercile colleges. These two findings show that both the magnitude and the ordering of differences in earnings across colleges can change substantially depending on when one measures earnings.

Table 8 formalizes this point by showing how the correlation of earnings with initial measures of college productivity evolve with worker experience. For this table we calculate earnings measures analogous to those above using the same students and colleges as in Figure 5. Panel A displays the raw mean measures (from equation (1)), and Panel D depicts residuals from a regression on college mean entrance exam scores (equation (4)).²³

The top panel of Table 8 shows the correlation of earnings measured at different experience levels with earnings at experience zero and with our field-specific earnings measure from above. The bottom panel shows the difference between the experience 2–8 correlations and the experience zero correlations in the first row. The results show that the correlation of earning measures with initial earnings declines substantially over time, and that this

²³ We do not present individual entrance exam residual measures in Table 8 because we do not observe the full vector of individual exam scores for all 2003–2012 graduates. The panel labels in Table 8 again correspond to the method used, i.e., to the columns in Table 4.

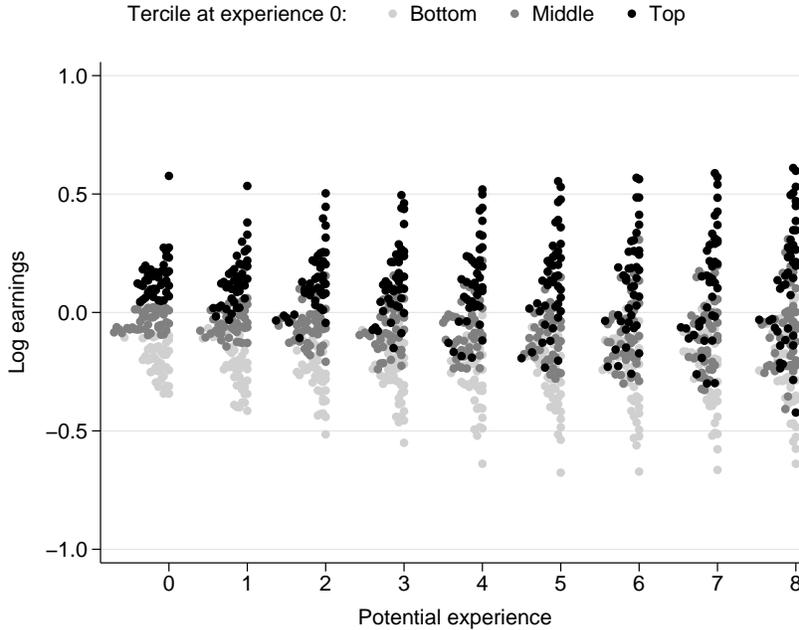


FIGURE 5. Log earnings by potential experience

Notes: The sample includes 2003–2012 graduates with earnings measured at 0–8 years of potential experience, defined as earnings year minus graduation year. Dots depict average log earnings at the 128 colleges in our sample with at least ten earning observations for each experience level. Log earnings are demeaned by graduation year and experience. We group colleges into three terciles based on experience zero earnings and add horizontal spacing to improve visibility.

holds for both the raw and residual methods. By contrast, the earning measures become *more* correlated with the field-specific exit scores over time, though the differences are not significant for the residual measures and fade out in later years.

The main takeaway from Figure 5 and Table 8 is that one can arrive at very different conclusions for a college’s earning productivity depending on when one measures earnings. This highlights the fact that colleges do not have complete control over the earnings of their graduates, which also depend on the post-schooling actions of workers and employers. This leaves open the possibility that learning measures do a better job of isolating a college’s contribution to students’ human capital.

5.3. Learning and earning across majors. Our final set of results concern one way in which colleges might be able to influence these productivity measures: their choice of which majors to offer. To explore how our measures vary across majors, we repeat the four procedures described in Section 4.2, but instead of calculating productivity at the institution level we do so at the institution/major level. In other words, we calculate separate learning and earning productivity measures for each major offered by each college.²⁴ We then

²⁴ We include only institution/major pairs that have at least 20 graduates in our sample.

TABLE 8. Correlations by potential experience

		Panel A. Raw means		Panel D. College-level residuals	
		Log earnings at exp. 0	Field exit score	Log earnings at exp. 0	Field exit score
<i>Correlations</i>	Log earnings at exp. 0	1.00***	0.44***	1.00***	0.15*
	Log earnings at exp. 2	0.93***	0.63***	0.92***	0.24***
	Log earnings at exp. 4	0.88***	0.68***	0.85***	0.24***
	Log earnings at exp. 6	0.83***	0.70***	0.78***	0.21**
	Log earnings at exp. 8	0.76***	0.69***	0.67***	0.14
<i>Differences from earnings at exp. 0</i>	Log earnings at exp. 2	-0.07*	0.20	-0.08**	0.08
	Log earnings at exp. 4	-0.12***	0.25*	-0.15**	0.08
	Log earnings at exp. 6	-0.17***	0.26**	-0.22***	0.05
	Log earnings at exp. 8	-0.24***	0.26**	-0.33***	-0.01

Notes: The top panel displays coefficients from linear regressions of earning measures at different experience levels on experience zero earning measures and the field-specific learning measures. The sample is the same as that for Figure 5. All regressions have 128 observations with weights equal to each college’s number of graduates. The bottom panel shows the difference between each of the experience 2–8 coefficients and the experience zero earnings coefficient. Stars indicate statistical significance with robust standard errors.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

sort the roughly 1,100 college/major pairs according to each measure and calculate each college/major’s percentile rank. This is analogous to the procedure used to calculate institution ranks in Table 5.

Table 9 summarizes the resulting ranks using nine broader major “areas” defined by the Ministry of Education.²⁵ The first column displays the proportion of all graduates in our sample in each major area. More than half of all graduates are in majors related to business and engineering, which are offered by almost all colleges in the country. Majors related to fine arts and natural sciences are less popular, and are offered by only a small number of colleges.

The other columns in Table 9 show the average ranks from the 1,100 college/major pairs using different learning and earning measures. Panel B presents ranks based on the entrance exam residuals method, and Panel D displays ranks based on the college/major level residual method. Using either method, the results show that some majors—such as those in engineering or law—receive much higher ranks under the earning measures than under the learning

²⁵ The Ministry’s categorization actually combines social sciences and law, but we split these major groups because they have vastly different properties with respect to our productivity measures.

TABLE 9. Average institution/major rank by major area

Major area	Prop. of grads	Panel B. Entrance exam residuals		Panel D. College/major level residuals	
		Field exit score	Log earnings	Field exit score	Log earnings
Business/economics	0.35	0.50	0.53	0.53	0.55
Engineering	0.29	0.51	0.60	0.47	0.60
Law	0.14	0.48	0.81	0.42	0.80
Social sciences	0.14	0.55	0.41	0.50	0.36
Health	0.07	0.52	0.66	0.52	0.67
Education	0.06	0.55	0.27	0.59	0.30
Fine arts	0.05	0.50	0.46	0.42	0.32
Agronomy	0.02	0.52	0.35	0.48	0.34
Natural sciences	0.02	0.75	0.62	0.56	0.50

Notes: This table includes all college/major pairs with at least 20 graduates in our sample, where majors are defined by the program name at each college. The Ministry of Education records aggregate these majors into the nine listed “areas.” The first column shows the proportion of graduates from each major area, and the remaining columns display percentile ranks of college/major pairs using the learning and earning measures in the column header. For these we sort college/majors according to each measure, and then calculate average ranks within the major areas. Averages are weighted by each college/major’s number of graduates.

measures. Conversely, majors related to education, fine arts, and social or natural sciences are much lower ranked using the earning measures.

Figure 6 elaborates on this result using a slightly more granular grouping of majors. The horizontal axis displays the average rank in each major group using the field-specific learning measure from Panel B of Table 9. The vertical axis depicts the average rank using the earning measure from the same procedure. Major groups that lie below the 45 degree line are ranked more highly on learning than on earning; these include many majors in social and natural sciences majors. Major groups above the 45 degree line, including many related to engineering and health, appear more favorable when rankings are based on earnings.

The results in Table 9 and Figure 6 suggest that the use of different productivity measures may create incentives for colleges to favor some majors over others. In particular, if policymakers primarily use earnings to measure performance, this could encourage college administrators to shift resources away from more specialized majors.

6. CONCLUSION

Increasingly, policymakers are looking to provide information on the outcomes that different colleges produce for their graduates. In many ways this reflects a desire to extend school accountability to higher education. Casual observation suggests this desire is particularly

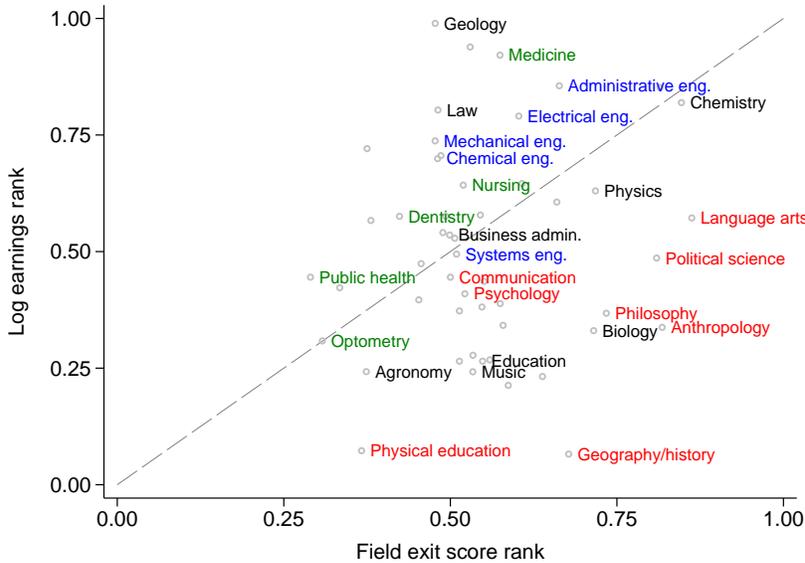


FIGURE 6. Earning vs. field-specific learning ranks by major group

Notes: This figure plots percentile ranks for college/major pairs using the entrance exam residual earning and field-specific learning measures. We calculate these ranks as in Panel B of Table 6, but we display average ranks within a more granular categorization of majors into 51 groups defined by the Ministry of Education. Averages are weighted by each college/major’s number of graduates.

prevalent in countries that have seen some combination of: significant growth in access to college, growth of a substantial (and often relatively unregulated) private sector, and increasing amounts of student debt.²⁶ As with school accountability in K-12 education—despite its much longer history—questions remain as to the informational content and the ultimate effects of initiatives in this area.

Our goal here has been to contribute by calculating, for the country of Colombia, system-wide measures of college productivity in terms of earning and learning. While we do not claim that our measures isolate causal college value added, they allow for analyses beyond those that have been previously feasible. Our findings suggest that measures of college productivity on earning and learning are far from perfectly correlated.

A key implication of this is that the design of accountability systems will affect how these portray different types of colleges, and potentially also how these colleges respond. For instance, we find that in the case of Colombia, top private colleges generally perform better under our earning measure, while selective public colleges appear more favorably under our learning measure.

In addition, in the earnings dimension one can arrive at starkly different conclusions regarding college’ relative productivity depending on when one measures earnings. This is problematic because the more chronologically removed the observation is from graduation,

²⁶ For instance, the U.S., Chile, and Colombia fit some of these criteria.

the more factors extraneous to colleges—such as post-schooling human capital investment decisions taken by employers and employees—will have a chance to affect wages. This leaves open the possibility that learning measures do a better job of isolating a college’s contribution to students’ human capital.

Finally, our results illustrate that the use of different productivity measures may create incentives for colleges to favor some majors over others. For example, our findings suggest that they might encourage institutions to shift resources away from more specialized majors and towards areas such as business and engineering.

REFERENCES

- Angrist, J., E. Bettinger, and M. Kremer (2006). Long-term consequences of secondary school vouchers: Evidence from administrative records in colombia. *American Economic Review*.
- Chetty, R., J. N. Friedman, and J. Rockoff (2014). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review* 104(9), 2593–2632.
- Cooke, M. (1910). *Academic and industrial efficiency*.
- Cunha, J. and T. Miller (2014). Measuring value-added in higher education: Possibilities and limitations in the use of administrative data. *Economics of Education Review*.
- Dale, S. B. and A. B. Krueger (2002, November). 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.
- Dale, S. B. and A. B. Krueger (2014, November). Estimating the effects of college characteristics over the career using administrative earnings data. *The Journal of Human Resources* 49(2), 323–358.
- Facione, N. C. and P. A. Facione (1997). *Critical thinking assessment in nursing education programs: An aggregate data analysis*. California Academic Press.
- Hagedorn, L. S., E. T. Pascarella, M. Edison, J. M. Braxton, A. Nora, and P. T. Terenzini (1999). Institutional context and the development of critical thinking: A research note. *The Review of Higher Education* 22(3), 265–285.
- Hoekstra, M. (2009). The effect of attending the flagship state university on earnings: A discontinuity-based approach. *Review of Economics and Statistics* 91(4), 717–724.
- Hoxby, C. (2002). School choice and school productivity (or could school choice be a tide that lifts all boats?). NBER Working paper 8873, National Bureau of Economic Research.
- Hoxby, C. and G. Bulman (2015). Computing the value added of american postsecondary institutions. Technical report, Stanford University.
- Klein, S., D. Freedman, R. Shavelson, and R. Bolus (2008). Assessing school effectiveness. *Evaluation Review* 32(6), 511–525.
- Klein, S., J. Steedle, and H. Kugelmass (2010). The lumina longitudinal study: Summary of procedures and findings comparing the longitudinal and cross-sectional models. Unpublished manuscript.
- MacLeod, W. B., E. Riehl, J. E. Saavedra, and M. Urquiola (2015). The big sort: College reputation and labor market outcomes. Mimeo, National Bureau of Economic Research Working Paper No. 21230.

- Mentkowski, M., G. Rogers, D. Deemer, T. Ben-Ur, J. Reissetter, W. Rickards, and M. Talbott (1991). Understanding abilities, learning and development through college outcomes studies: What can we expect from higher education assessment? Paper presented at the annual meeting of the american educational research association (chicago, illinois).
- Pascarella, E. T. and P. Terenzini (2005). *How college affects students: A third decade of research*.
- Prendergast, D. L. (1998). Influences of college environments and the development of critical thinking skills in college students. Doctoral dissertation, University of Illinois at Chicago.
- Rothwell, J. and S. Kulkarni (2015). Beyond college rankings: A value-added approach to assessing two- and four-year schools. Technical report, Washington D.C.: Brookings Institution.
- Saavedra, A. and J. E. Saavedra (2011). Do colleges cultivate critical thinking, problem solving, writing and interpersonal skills? *Economics of Education Review*.
- Saavedra, J. (2009). The learning and early labor market effects of college quality: A regression discontinuity analysis. Mimeo, Harvard University.
- Silva, E., T. White, and T. Toch (2015). The carnegie unit: A century-old standard in a changing education landscape. Technical report, Stanford: Carnegie Foundation.
- Stange, K. (2012). Ability sorting and the importance of college quality to student achievement: Evidence from community colleges. *Education Finance and Policy* 7(1), 74–105.
- Sullivan, T., C. Mackie, W. F. Massy, and E. Sinha (2012). Improving measurement of productivity in higher education. Technical report, The National Academies Press: Washington, D.C.