



NATIONAL
CENTER *for* ANALYSIS of LONGITUDINAL DATA *in* EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

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



*How High Schools
Explain Students'
Initial Colleges and
Majors*

Rajeev Darolia
Cory Koedel

How High Schools Explain Students' Initial Colleges and Majors

Rajeev Darolia
University of Missouri

Cory Koedel
University of Missouri/CALDER

Contents

Acknowledgements.....	ii
Abstracts.....	iii
1. Introduction.....	1
2. Context and Data.....	3
3. Defining Students' Academic Indices and University-Major Quality.....	6
4. Variation in University-by-Major Cell Quality and Student Sorting.....	10
5. The Role of High Schools in Student Sorting.....	12
6. Results.....	14
7. Extension.....	15
8. Conclusions.....	19
References.....	22
Figures & Tables.....	26
Appendix A.....	35

Acknowledgements

We thank the Missouri Department of Higher Education for providing access to data and gratefully acknowledge research support from the National Center for Analysis of Longitudinal Data in Education Research (CALDER) funded through grant #R305C120008 to American Institutes for Research from the Institute of Education Sciences, U.S. Department of Education. We thank session participants at Purdue University and the APPAM and CALDER conferences for valuable feedback, in particular Mark Long and Carrie Conaway. The usual disclaimers apply.

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

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

How High Schools Explain Students' Initial Colleges and Majors

Rajeev Darolia, Cory Koedel

CALDER Working Paper No. 165

January 2017

Abstract

We use statewide administrative data from Missouri to examine the role of high schools in explaining students' initial college and major placements at 4-year public universities. To facilitate our investigation of postsecondary sorting, we develop a "preparation and persistence index" (PPI) for each university-by-major cell in the Missouri system to measure the quality of that cell. The PPI depends on the pre-college academic qualifications of degree completers. Our analysis of high schools shows that the high school attended predicts the quality of the initial university, as measured by PPI, conditional on a student's own academic preparation. Consistent with previous research, we further show that students from lower-SES high schools systematically enroll at lower-quality universities relative to their similarly-qualified peers from higher-SES high schools. However, high schools offer little explanatory power over major placements within universities. Correspondingly, there are not meaningful differences in the quality of these placements by high-school SES.

1 Introduction

College and major placements play an important role in shaping students' academic and post-college outcomes (Arcidiacono, 2004; Astin, 1993; Carnevale et al., 2016; Cohodes and Goodman, 2014; Stinebrickner and Stinebrickner, 2014). These placements also collectively influence the human capital of the workforce, which is important in light of concerns that students in the United States are no longer keeping pace with their global competitors in developing the key skills that promote long-term economic prosperity (Committee on Prospering in the Global Economy of the 21st Century, 2007). For these reasons, and because the socioeconomic backgrounds of students are unequally distributed across universities and majors, recent research has focused increasingly on the factors that explain how and why students enroll in different colleges and pursue different majors (Arcidiacono, Aucejo, and Hotz, 2016; Bowen, Chingos and MacPherson, 2009; Hoxby and Turner, 2014; Hurwitz et al., forthcoming; Porter and Umbach, 2006; Wiswall and Zafar, 2015).

We contribute to the literature on college and major sorting by examining the role of high schools in explaining students' initial university and major placements. An innovation of our study is to develop an empirical measure of academic quality for each university-by-major cell in the state university system we study (Missouri), which we refer to as a "preparation and persistence index" (PPI). The PPI for each cell is a function of the pre-college academic qualifications of students who complete a degree in that cell. The PPI therefore varies across university-by-major cells because of differences in admissions decisions, students' initial university and major choices, persistence within cells, and cross-cell transfers. Our PPI measures allow for a detailed investigation of postsecondary sorting both across and within universities and offer several conceptual benefits that derive from their flexible, empirical foundation. One benefit is that they

facilitate rankings of major quality that overlap across universities with different levels of overall selectivity – for example, in the case of two universities that differ in selectivity on average, the PPI measures permit cases where some majors at the less-selective university are of higher quality than some majors at the more-selective university. PPI also allows us to move away from traditional, subjective divisions of college majors that have been used in the past – most notably, between majors broadly grouped as science, technology, engineering and mathematics (STEM) and other majors. Relatedly, it allows for a better accounting of heterogeneity within groups of STEM and non-STEM majors.¹

We use our PPI measures to document across- and within-university variance shares of the quality of university-by-major cells in the Missouri system. Universities explain a substantial fraction of the total variance of PPI – about 64 percent – but within-university variance is also important (36 percent). We also explore related variability in the academic alignment between students and their entering university-by-major cells. This analysis complements previous research focusing on “undermatching” in the alignment between students’ academic preparation and university of attendance (Arcidiacono and Lovenheim, 2016; Dillon and Smith, forthcoming; Hoxby and Turner, 2014; Rodriguez, 2015; Smith, Pender, and Howell, 2013), which we extend to consider alignment between students and majors within universities. We show that within- and across-university sorting both contribute substantially to the total system-wide variance in academic alignment for individual students. This finding has implications for the way that we

¹ As an example, rather than lumping all STEM fields together, our measures differentiate relatively more-selective STEM disciplines (e.g., engineering fields) from less-selective ones (e.g., biological sciences). There is growing awareness of the diversity of majors within STEM and non-STEM categories. For example, Webber (2016) estimates wage premia for many majors and although STEM (and business) majors are associated with higher earnings on average, the premia for some fields are lower than others; e.g., Webber reports that the premium for biology majors is in line with the premium for arts/humanities majors (p. 305).

evaluate and design policies related to the alignment of students and their postsecondary placements.

A number of studies examine how high schools influence academic performance in college, focusing on outcomes such as college grades, persistence, and graduation (Black, Lincove, Cullinane, and Veron, 2015; Fletcher and Tienda, 2010; Long, Iatarola, and Conger, 2009; Pike and Saupe, 2002; Niu and Tienda, 2013). We complement this body of research by examining the predictive power of high schools over the quality of students' initial university-by-major placements conditional on their own pre-entry academic preparation. We observe large numbers of students who enter and exit the university system via various college and major pathways from hundreds of high schools in the state. Our results indicate that high schools are strong predictors of the quality of the entering university-by-major cell. This finding is driven primarily by the explanatory power of high schools over university placements. Consistent with prior research, we find that students from lower-SES high schools systematically enroll at lower-quality universities conditional on their own academic preparation – i.e., universities where their peers have worse pre-entry academic preparation. However, despite the presence of substantial variation in the quality of entering-major cells within universities, high schools explain a negligible fraction of the variance in these placements.

2 Context and Data

We use administrative microdata provided by the Missouri Department of Higher Education (DHE) for the empirical analysis. We focus our attention on six cohorts of full-time, state-resident, non-transfer students who entered the public 4-year university system in Missouri from a public high school between 1996 and 2001 as college freshman. In total, our analytic sample

includes 58,377 students. Basic descriptive statistics of students in our analytic sample are provided in Appendix Table A.1.²

We identify collegiate major pathways based on the Classification of Instructional Programs (CIP) taxonomy developed by the US Department of Education.³ We define majors as specific to each university. This means that we treat students who enter the same major (i.e., same CIP code) at different universities as entering via separate pathways. We also note that in Missouri, like in other states, university enrollment is not entirely separable from major enrollment because universities have different major offerings. In total, over the course of our data panel we identify 476 unique university-by-major cells in the Missouri 4-year public university system.

The initial major that we use to define the entering cell is best interpreted as an “intended” major because there are no requirements or formal system rules that govern the initial selection (e.g., a student can declare herself to be a business major upon entry, prior to being officially accepted into the business program). Though not formally binding, the initial major is important because it shapes students’ initial plans of study, peers, and advisors.⁴ We match enrollment data to completion records to identify a final university and major for each graduate. Each student is tracked for eight years to determine graduation outcomes; all individuals who do not obtain a degree within eight years from a university in the Missouri system are coded as non-completers.

² Our dataset is similar to the dataset used by Arcidiacono and Koedel (2014). Notable differences between the datasets are that we include students from all racial and ethnic groups in our data, whereas they restrict their analysis to African American and white students, and we restrict our attention to students who matriculate into the system from public high schools.

³ We aggregate majors at the 4-digit CIP code level. For sparsely populated university-by-major cells (those with less than 10 who start or less than 5 that finish in the cell), we aggregate them with other majors within the 2-digit CIP code level, but this type of aggregation affects a small number of students (approximately four percent of completers obtain a degree with a CIP code that must be aggregated).

⁴ Furthermore, as documented below, the initial major is highly predictive of the final major. In cases where students list multiple majors, we identify the primary major based on the first listed major.

We observe students' high schools of attendance and for many high schools we observe large numbers of students entering the 4-year university system.⁵ Thus, our data are well-suited to examine the transition from high schools to university-by-major cells, given that we typically have large unit-level samples at both levels. The DHE data additionally include detailed information on the pre-college academic preparation of individual students – most notably, students' class percentile ranks and ACT scores. We use these data to (a) construct the empirically-derived PPIs for each university-by-major cell as described in the next section, and (b) investigate the role of high schools in determining student sorting conditional on students' own pre-entry academic preparation.

There are 13 public 4-year universities in the state system, mapped in Figure 1.⁶ The University of Missouri-Columbia is the flagship university and only university with the highest research activity distinction. The other highly selective universities are Truman State University and the STEM-focused Missouri University of Science and Technology.⁷ There are also two historically black universities in the system, Harris-Stowe State University and Lincoln University (the latter is a land grant university).

We provide additional information about Missouri universities in Table 1. The universities are ordered by the average of an individual academic preparation index for entering students in the first column (we describe the preparation index in the next section). There are several notable features of the system. Beginning with how enrollment is distributed across universities, the third

⁵ We drop records from approximately 3 percent of in-state students who do not have an assigned high school of attendance in the DHE data or who come from high schools that send a small number (<10) of students to an in-state, public university during the period. We observe students who attended 455 different public high schools.

⁶ We use the word “system” to describe all 13 Missouri universities. In terms of governance, there are several subsystems of universities (e.g., the 4-campus “University of Missouri” system) but we do not distinguish between these subsystems in our work.

⁷ Based on the 2015 Carnegie Classifications of Higher Education. See <http://carnegieclassifications.iu.edu>. We use the term “highly selective” to characterize institutions with an undergraduate profile considered “more selective” in the Carnegie lexicon (the highest level of selectivity).

column shows that over forty percent of students in the analytic sample enter into just two universities: the University of Missouri-Columbia and Missouri State University. No other university has more than a 10-percent enrollment share. The three universities with the highest average pre-entry preparation indices also exhibit the least variation in the index among entrants.

The fourth column of Table 1 shows the eight-year graduation rate for each campus (determined by tracking students in our sample for up to eight years after entry to see if a bachelor's degree was obtained). Graduation rates map fairly closely to the pre-entry preparation index in column 1. The most notable differences occur at the urban campuses, University of Missouri-Kansas City and University of Missouri-St. Louis, which have lower graduation rates than would be predicted by students' pre-entry preparation alone. The low graduation rates at the urban campuses are consistent with similar results reported using Missouri data in Arcidiacono and Koedel (2014), and more broadly for urban campuses in Bowen, Chingos and McPherson (2009), who show that graduation rates are negatively related to commuter share.

Finally, the last two columns of Table 1 display the average and standard deviation of the academic preparation index among graduates. As expected, the average index is higher among graduates than non-graduates, which can be seen by comparing the inclusive index values in column 1 with the graduate-only values in column 5. The average index difference between entrants and graduates is negatively related to the average index of entrants.

3 Defining Students' Academic Indices and University-Major Quality

3.1 Students' Academic Indices

We begin by constructing academic indices for individual students. The first step is to regress graduation outcomes on students' academic qualifications prior to college entry:

$$Y_{ijmt} = \beta_0 + ACTM_i\beta_1 + ACTR_i\beta_2 + CR_i\beta_3 + \gamma_t + \theta_{jm} + \varepsilon_{ijmt} \quad (1)$$

In equation (1), Y_{ijmt} is an indicator for whether student i in year-cohort t , who entered the system in the university-by-major cell defined by university j and major m , completed a degree within eight years of entry. The variables $ACTM_i$ and $ACTR_i$ are the student's math and reading ACT scores, and CR_i is the student's class percentile rank in high school. γ_t are cohort fixed effects, θ_{jm} are fixed effects for each university-by-major cell, and ε_{ijmt} is an error term, which we specify as having a Type I extreme value distribution implying that the probability of graduation follows a logit. This model is similar to the one developed by Arcidiacono and Koedel (2014).⁸

We use the output from equation (1), and in particular our estimates of $\beta_1 - \beta_3$, to construct an index of pre-entry academic qualifications, AI , for each student as follows:

$$AI_i = ACTM_i \hat{\beta}_1 + ACTR_i \hat{\beta}_2 + CR_i \hat{\beta}_3 \quad (2)$$

The index is a weighted average of students' pre-entry academic qualifications, where the weights are empirically derived from the graduation model in equation (1) so that the pre-entry qualifications that best predict success in college (as measured by graduation) are given the most weight. A critical aspect of the index is that by the inclusion of γ_t and θ_{jm} in equation (1), we ensure that the identifying variation for the weighting parameters ($\beta_1 - \beta_3$) comes from within university-by-major cells and cohorts.⁹

Table 2 shows results from the estimation of equation (1) – in particular, the coefficient values used to construct the academic index in equation (2) – to provide a sense of the relative importance of students' pre-entry characteristics in shaping the index. Focusing on the estimates

⁸ Unlike Arcidiacono and Koedel, we do not allow for students' pre-entry academic qualifications to differentially predict student success across university-by-major cells. This adjustment is mainly to preserve analytic tractability, as our much narrower definition of a major (CIP code), multiplied across up to 13 system universities, greatly expands the parameter space relative to their study.

⁹ We exclude explicit measures of high school quality (high school fixed effects) from the graduation model and index. This allows for a more straightforward examination of the explanatory power of high schools over student placements below.

from our preferred specification in column 1, a student’s class percentile rank is the strongest predictor of graduation conditional on the entering cell. A one standard deviation change in the class rank corresponds to a change in the index of about 0.65. One standard deviation changes in ACT math or reading scores correspond to index changes of 0.16 and 0.05, respectively. The point estimate on the ACT reading score is negative in column 1, but this is because we also condition on high school class rank – ACT reading scores positively predict graduation independently.

The index in column 2 excludes the class rank, which means that no locally-normed information is used to construct the index. This can be useful for interpretation. For example, a key finding below is that students from lower-SES high schools enroll in lower-quality university-by-major cells conditional on their own index values. One explanation is that a high class rank at a low-SES high school is a weaker indicator of academic preparation. We explore this possibility below using the sparser academic index shown in column 2.

3.2 Preparation and Persistence Indices for University-by-Major Cells

The PPI for each university-by-major cell is based on the academic index values of individuals who complete a degree in that cell, regardless of the entering cell. Therefore, variation in cell quality arises from differences in initial selection (which can be driven by students’ own choices and the behavior of admissions officials), student persistence within cells, and cross-cell student transfers. We start by taking the average academic index among degree completers in cell jm :

$$Q_{jm} = \frac{1}{N_{jm}} \sum_{i=1}^{N_{jm}} AI_i \tag{3}$$

where N_{jm} is the number of individuals who complete a degree in the cell defined by university j and major m . We then define δ_{jm} , an empirical Bayes estimate of quality for cell jm , as follows:

$$\delta_{jm} = \alpha_{jm} * Q_{jm} + (1 - \alpha_{jm}) * Q_j \quad (4)$$

In equation (4), Q_j for university j is defined analogously to Q_{jm} as shown in equation (3), but at the university level, and is treated as deterministic. The parameter α_{jm} , with $0 < \alpha_{jm} < 1$, shrinks the overall quality estimate for cell jm toward the university mean (i.e., the prior). The degree of shrinkage depends on the precision with which Q_{jm} is measured, with more-precisely measured values corresponding to higher values of α_{jm} . The formula we use for α_{jm} is:

$$\alpha_{jm} = \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{\lambda}_{jm}} \quad (5)$$

In equation (5), $\hat{\sigma}^2$ is an estimate of the true variance of Q across university-by-major cells, net of sampling variance, and $\hat{\lambda}_{jm}$ is an estimate of the estimation-error variance of Q_{jm} .

We follow the recent literature on teacher quality to estimate the parameters used in equation (5) (Koedel, Mihaly and Rockoff, 2015). Briefly, we first estimate the following regression using degree completers in our analytic sample:

$$AI_{ijm} = \mathbf{D}_{jm} \boldsymbol{\pi} + e_{ijm} \quad (6)$$

where AI_{ijm} is the academic index for individual i who completes a degree in cell jm , and \mathbf{D}_{jm} is a vector of indicators for each cell. The raw variance of the academic index for completers across cells is estimated by the variance of $\hat{\pi}_{jm}$, which we adjust to obtain an estimate of the true variance – $\hat{\sigma}^2$ in equation (5) – by netting out the total estimation-error variance following the procedure

outlined in Koedel (2009).¹⁰ We estimate λ_{jm} for each cell as the square of the standard error of $\hat{\pi}_{jm}$ from equation (6).

An appealing aspect of our quality measures is their objectivity. As noted in the introduction, PPI is not influenced by subjective assessments of college majors, either within or across universities, as it depends entirely on the pre-entry academic qualifications of graduates. In Appendix Table A.2 we list the ten highest- and lowest-quality university-by-major cells in the Missouri system based on PPI for illustrative purposes.¹¹

4 Variation in University-by-Major Cell Quality and Student Sorting

A basic variance decomposition of cell-level PPI indicates that 64 percent of the variance occurs across universities and 36 percent occurs within. While this split affirms the literature’s focus on the importance of institutional sorting (Arcidiacono and Lovenheim, 2016; Dillon and Smith, forthcoming; Hoxby and Turner, 2014; Rodriguez, 2015; Smith, Pender, and Howell, 2013), it also highlights the presence of substantial variability in the quality of majors within institutions.

In addition to the decomposition, we also use measures of academic alignment between individual students and their initial university-by-major cells to contextualize system sorting. To do so, we first define academic alignment for student i who enters cell jm as $M_{i,jm} = AI_i - \delta_{jm}$. We compare observed alignment based on actual student sorting to alignment under two types of

¹⁰ Koedel’s procedure is similar to related procedures found in other studies such as Aaronson, Barrow and Sander (2007), but is better suited to handle situations where there is larger sample-size variance across units (in this case a unit is a university-by-major cell).

¹¹ There are some system cells in which students enter but none graduate – the most prominent example includes students who initially enroll as an undecided major. We cannot construct PPI measures using our base methodology for these cells because the quality measures depend on completers. As an alternative, we construct analogous measures of entry-cell PPI that are a weighted average of final-cell PPI among completers, who by construction must have switched to a different cell. This is an imperfect but functional solution to permit the inclusion of these individuals in our sample. Below we examine the robustness of our findings to dropping students who enter these cells and we obtain similar results.

counterfactual sorting conditions: (1) random assignment of students to system cells; and (2) perfect sorting of students to system cells (where we assign the highest-*AI* students to the cells with the highest values of δ_{jm}). For each set of counterfactual conditions, we consider two scenarios: (a) a “global” scenario in which the counterfactual sorting occurs across and within universities; and (b) a “local” scenario where the counterfactual sorting is conditional on the initial university. For example, with global random assignment, we randomly assign students to majors and universities; whereas with local random assignment, we randomly assign students to majors holding the entering university fixed. The variance of the alignment measure, $M_{i,jm}$, will be minimized in the global perfect-sorting case because students’ own academic indices will align most closely with the hypothetical entering university and major.¹² The variance will be at its practical maximum with global random assignment. These comparisons provide context for observed sorting.

Table 3 reports the results. The top row shows the variance of $M_{i,jm}$ based on students’ actual university-by-major placements. Subsequent rows report the variance under the four counterfactuals. The observed variance of $M_{i,jm}$, 0.43, falls comfortably in between the two global counterfactual bounds of 0.22 (perfect sorting) and 0.60 (random sorting).

The counterfactual scenarios provide useful insight into the potential for cross-university and within-university sorting to affect alignment. For example, the within-university, perfect sorting condition minimizes within-university misalignment (last row of Table 3). The variance of $M_{i,jm}$ in this scenario is 0.28, which is close to the global perfect-sorting condition (0.22); certainly much closer than to the observed sorting condition (0.43). The implication is that

¹² This minimization is subject to the pre-existing structure of the system, and in particular the size of system cells, which we hold fixed for this descriptive analysis.

resorting students to majors with closer academic alignment, without any switching across universities, would increase alignment nearly as much as resorting students across the entire system. This does not diminish the importance of college placements in studying postsecondary sorting; rather, it motivates the importance of also studying sorting within universities.

5 The Role of High Schools in Student Sorting

Having defined each student’s own preparation index and the PPI of the entering university-by-major cell, we examine the explanatory power of high schools over student placements into colleges and majors conditional on each student’s own academic preparation. We start with the following linear regression model:

$$\delta_{jm,is} = \gamma_0 + AI_i \gamma_1 + \mathbf{HS}_{is} \boldsymbol{\gamma}_2 + u_{jm,is} \quad (7)$$

In equation (7), the PPI of university-by-major cell jm into which student i from high school s enters, $\delta_{jm,is}$, is a function of the student’s own academic index, AI_i , and the high school attended, where \mathbf{HS}_{is} is a vector of indicator variables in which the student’s own high school indicator is set to one and all others are set to zero. We do not allow a student’s own academic index to contribute to $\delta_{jm,is}$ to prevent spurious correlations. Thus, if a student starts and completes a degree in cell jm , her own academic index is jack-knifed out of the calculation of $\delta_{jm,is}$. The parameter γ_1 is identified using within high-school variation in AI_i to estimate the empirical relationship between a student’s own academic preparation and the PPI of the initial cell. Conditional on this relationship, the vector of high school fixed effects, $\boldsymbol{\gamma}_2$, captures systematic differences in the PPI of placements across high schools. $u_{jm,is}$ is the residual in the regression. We estimate standard errors using a 2-way clustering structure to account for dependence in the data within university-by-major cells and high schools following Petersen (2009; also see Cameron and Miller, 2015).

We are also interested in studying the extent to which high schools explain differences in the PPI of student placements across majors *within universities*. The model in equation (7) can be extended for this purpose as follows:

$$(\delta_{jm} - \delta_j)_{is} = \theta_0 + AI_i\theta_1 + \mathbf{HS}_{is}\boldsymbol{\theta}_2 + e_{jm,is} \quad (8)$$

The only change in equation (8) is that the dependent variable is measured relative to overall university PPI, where universities are subscripted by j . Our measures of university PPI are constructed analogously to our measures of university-by-major PPI per the description in Section 3.¹³

Next we examine whether characteristics of high schools systematically explain the PPI of students' placements. Following on previous research showing that students from disadvantaged backgrounds tend to enroll in universities where their own academic preparation exceeds that of their peers, we are particularly interested in the degree to which measures of socioeconomic disadvantage at the high school level predict placement quality. To investigate this question we estimate the following analogs to equations (7) and (8):

$$\delta_{jm,is} = \rho_0 + AI_i\rho_1 + \mathbf{Z}_{is}\boldsymbol{\rho}_2 + \xi_{jm,is} \quad (9)$$

$$(\delta_{jm} - \delta_j)_{is} = \psi_0 + AI_i\psi_1 + \mathbf{Z}_{is}\boldsymbol{\psi}_2 + \zeta_{jm,is} \quad (10)$$

These equations substitute measures of socioeconomic disadvantage for high schools and their surrounding local areas in the Z-vector for the high school indicators that are in equations (7) and (8). The high-school disadvantage measures we include are the share of the student body that is free or reduced price lunch (FRL) eligible and the share of individuals age-25 and older with less than a bachelor's degree in the high school's zip code. Because socioeconomic factors can also

¹³ In fact, because we treat university PPI as deterministic per Section 3, $\delta_j = Q_j$

vary systematically by the minority composition of the high school, we also include the share of the student body that identifies as a minority race or ethnicity in the Z -vector. The school-level data are taken from the Common Core of Data (CCD) and the local-area data are from the year-2000 U.S. Census.

We explore the predictive influence of high-school SES conditional on students' own academic indices as shown in equation (9) and (10), and also conditional on basic characteristics of high schools including urbanicity (schools are divided into five groups: urban, suburban, town, rural and missing) and school size (enrollment). For ease of interpretation, we normalize the dependent variables and high-school characteristics to have a mean of zero and a variance of one.¹⁴ In our preferred specifications as shown in equations (7)-(10), we also normalize the academic index for individuals and enter it into the models linearly. In Appendix Table A.4 we show that our findings are qualitatively unaffected if we use a more flexible modeling approach where we divide students into twenty equal-sized bins based on their own index values and condition on bin assignment instead.

6 Results

Equations (7) and (8) allow us to assess the general importance of high schools in explaining students' initial placements conditional on their own academic indices. Table 4 reports the overall R-squared and partial R-squared attributable to the vector of high school indicators for each model. High schools explain 11.3 percent of the variance in university-by-major PPI overall. However, they explain just 1.4 percent of the within-university variance.

¹⁴ More precisely, the dependent variables are normalized so that a one-unit change represents a one-standard-deviation change in the true distribution of PPI. In practice, the normalized dependent variables have a standard deviation of less than one because they are normalized by the un-shrunken standard deviations. This facilitates the interpretation of a one-unit change in PPI as corresponding to a one standard deviation change in the true (rather than empirical) distribution (see also Chetty, Friedman, Rockoff, 2014; Jacob and Lefgren, 2008).

In Table 5 we show results from variants of equation (9) where we replace the high school indicators with high school characteristics. We include the minority share and each measure of socioeconomic disadvantage in the model separately and then include them all simultaneously, with and without conditioning on the other basic high school characteristics. In the full specification in the final column of Table 5, one standard deviation increases in the minority share, the percentage of FRL-eligible students, and the share of the local area with less than a bachelor's degree correspond to changes in the PPI of the initial university-major cell of -0.02 (not statistically significant), -0.02 (not statistically significant), and -0.14 standard deviations, respectively. A general takeaway from Table 5 is that students from more disadvantaged backgrounds sort to lower quality university-by-major cells conditional on their own academic preparation, which is in line with previous research on undermatch to universities (Turner, 2017).

Next we extend the analysis to look for systematic placements by high school minority share and SES within universities. Table 6 follows the same structure as Table 5, but focuses on within-university placements per equation (10). Consistent with the limited explanatory power of high schools over within-university sorting documented in Table 4, and in contrast to the results for overall placements in Table 5, the results in Table 6 provide little indication of differences between students from high schools with different characteristics. Neither of the high-school SES measures are meaningfully associated with placement quality within universities, individually or jointly, and the same is true for the minority share.

7 Extensions

7.1 An Alternative Academic Index Excluding Class Rank

The measure of academic preparation that receives the most weight by far in the individual academic index – the high school class percentile rank – is a locally-normed measure. One reason

we might find that students from low-SES high schools enter the system in lower-quality cells is that conditional on the index, their preparation is lower. Put differently, it may be that performing at the top of the class at a low-SES high school does not signify the same level of preparation as performing at the top of the class at a high-SES high school. This explanation is consistent with findings from Black et al. (2015), who show that students in Texas with high class ranks but who attended low-performing high schools have persistently lower grades throughout college than their peers who attended better high schools. We gain insight into this issue by using a version of the academic index that does not include class percentile rank (from column 2 of Table 2).

We present results from this exercise in Table 7, where we replicate our full procedure and show specifications akin to those in Tables 5 and 6 using the restricted academic index without high school class rank. A caveat to these results is that we sacrifice substantial informational content by excluding information about students' class ranks.¹⁵ Bearing this in mind, in the model examining system-wide placements in columns 1 and 2, we find generally similar but weaker results to what we show in Table 5 for the income and education SES measures, but the coefficient on the minority share flips in sign and is statistically significant. This pattern of results is also apparent when we enter the high-school SES and minority share measures into the models separately (not shown for brevity). In columns 3 and 4, where we replicate the results from the full specifications in Table 6, there is also a moderate shift toward the appearance of less under-placement for students from low-SES high schools. Specifically, whereas with our primary specification there is not a detectable pattern of within-university sorting by high school SES conditional on students' own academic preparation, when we use the restricted index we find that

¹⁵ While ACT scores have the benefit of not being locally normed, they are much weaker predictors of college success than class ranks (or similar measures such as high-school GPA). In addition to this being shown in our analysis in Table 2, also see Bowen, Chingos, and McPherson (2009), Fletcher and Tienda (2010), and Rothstein (2004).

students from lower-SES high schools conditionally enroll in modestly *higher-quality* majors within universities. In summary, students from low-SES high schools seem less under-placed when we no longer account for class rank.

This shift in results is consistent with the interpretation that our primary estimates in Tables 5 and 6 are driven partly by the fact that highly ranked students from low-SES high schools are not as well prepared as their highly ranked peers from high-SES high schools. Either by their own application and enrollment actions or the actions of university admissions officials, this is reflected in lower-quality placements conditional on these students' own academic indices. This interpretation has significant social meaning: the unequal value of class rank would directly imply that differential opportunities for human capital development during K-12 schooling between students in high- and low-SES high schools explain some of the differences we observe in entry-cell quality.

7.2 Cells without Completers

Next we turn to the issue that approximately one-third of the students in the sample enter into cells in which there are no completers. The predominant example is students who list their initial field of study as “undecided,” who account for about one-fifth of our sample, or approximately 13,000 students. There are also another 5,800 students who begin in a cell without any finishers, with the most common reason being that the initial cell is a broad field such as “general engineering.” Students who enter into a broad field like “general engineering” do not finish with a general degree. Instead, they either finish in a more specific engineering subfield, such as chemical engineering or mechanical engineering, switch to a completely different discipline, or drop out. In the analysis thus far, we have handled such cells by assigning them a PPI measure that is a weighted average of *finishing cell* PPI across all graduating students who

enter. This is a functional solution, but treats these cells differently than other cells (for other cells, only finishers matter regardless of the entering cell as described in Section 3.2).

In Appendix Table A.3 we examine the sensitivity of our findings to dropping all students who enter university-by-major cells with no completers, since we do not have a consistent strategy for constructing measures of cell PPI for these students. For brevity, we replicate our estimates from the full models shown in Tables 5 and 6 only. The results show that our findings are qualitatively unaffected by whether we include these individuals in the analysis.

7.3 The Mapping Between Initial and Final University-by-Major Cells

Students' initial enrollment placements influence their academic experiences and outcomes (e.g., Astin, 1993; Carrell, Fullerton, and West, 2009; Porter and Umbach, 2006; St. John et al., 2004). However, there is also a robust literature that connects post-college outcomes to *final* college and major (Arcidiacono, 2004; Carnevale et al., 2016; Hamermesh and Donald, 2008; Thomas and Zhang, 2005). An obvious question given our focus on initial university-by-major placements is how initial placements translate to final placements.

To answer this question we begin with basic summary statistics. A substantial proportion of the students in our sample finish in the university-by-major cell in which they start. Among those that declared a major when they entered the system and graduated, almost 40 percent finished in the same cell that they entered. Furthermore, nearly 60 percent finished in the same major group (with the same 2-digit CIP code) as the entering major.

To address this question more generally, we estimate the relationship between the PPI of the initial and final cell using a simple, student-level regression of the following form:

$$\delta_{jm,is}^F = \varphi_0 + \delta_{jm,is}^I \varphi_1 + AI_i \varphi_2 + \zeta_{jm,is} \quad (11)$$

In equation (11), $\delta_{jm,is}^F$ is the normalized PPI of the final cell and $\delta_{jm,is}^I$ is the normalized PPI of the initial cell.¹⁶ The estimation of equation (11) is restricted to degree completers.

Figure 2 plots the unconditional relationship between $\delta_{jm,is}^F$ and $\delta_{jm,is}^I$ among completers. The markers represent the average ending PPI for each bin of beginning PPI, with bin sizes of 0.1 standard deviations. The size of each marker reflects the number of students in the bin. It is visually apparent that the PPI of the initial major is highly predictive of the PPI of the final major, and that this strong relationship holds throughout the distribution of beginning-cell PPI. This is supported formally by results from equation (11) where we estimate φ_1 to be 0.78 with a standard error of 0.02.

The strong link we identify between PPI of the starting and ending cells should not be interpreted causally and it is important not to infer that simply changing initial placements will necessarily change final placements. That said, the link between the PPI of the initial and final cell is quite strong, which implies that policies that change students' initial placements *and the factors that underlie these placements* can meaningfully change the distribution of university-by-major quality at exit.

8 Conclusion

We develop new empirical measures of the quality of university-by-major cells that depend on the pre-college academic qualifications of degree completers. Our measures – which we term “preparation and persistence indices” (PPIs) – afford us great flexibility in examining student sorting within the 4-year public university system in Missouri. We find that the quality of

¹⁶ As in the preceding analysis, the normalizations are performed to facilitate interpretations in terms of the real (rather than empirical) distributions of PPI. Because the PPI measures are shrunken, estimates of φ_1 will not be affected by attenuation bias (Jacob and Lefgren, 2008).

university-by-major cells varies substantially both within and across universities, and relatedly, that sorting both within and across universities meaningfully contributes to the total system-wide variance in academic alignment for students.

Our examination of the explanatory power of high schools over students' initial university-by-major placements, conditional on students' own academic preparation, yields the insights that high schools explain (a) a substantial share of the variance in the PPI of university placements, and (b) little of the variance in the PPI of major placements within universities. Corroborating previous research, the socioeconomic status of high schools and their surrounding areas is a clear predictor of the PPI of students' initial university placements, with lower-SES students systematically enrolling at lower-PPI universities conditional on their own academic preparation (Dillon and Smith, forthcoming; Hoxby and Avery, 2013; Hoxby and Turner, 2014; Smith, Pender, and Howell, 2013). A similar pattern is not present for within-university sorting to majors, which is consistent with the limited explanatory power of high schools over within-university placements that we document generally.

Our findings have several important implications for research and policy. First, they point toward the value of interventions that inform students of the educational options for which they are academically qualified, which can better align students from low-SES high schools with universities (Hoxby and Turner, 2014). The differences in sorting to institutions that we document between these students and their higher-SES peers are disconcerting in light of compelling evidence that more-selective institutions, as measured by the academic qualifications of entering students, improve educational outcomes (Cohodes and Goodman, 2014; Hoekstra, 2009; Melguizo, 2010). Even if some of the disparate sorting behavior between seemingly similarly-qualified students from high- and low-SES high schools is driven by true gaps in student

preparation owing to unequal opportunities during high school (per Table 7), the greater efficacy of more selective institutions will still likely benefit lower-SES students (Arcidiacono and Koedel, 2014; Dillon and Smith, forthcoming).

Second, we document substantial within-university variation in PPI between majors within universities. While majors can affect learning and influence students' academic environments, including interactions with faculty and the development of peer groups (e.g., Astin, 1993; Carrell, Fullerton, and West, 2009; Porter and Umbach, 2006; St. John et al., 2004; Umbach and Wawrzynski, 2005), little is known about the practical importance of quality differences across majors in terms of affecting student outcomes, or about the malleability of student allocations to departments within universities should reallocations be desirable. Our findings at least raise the possibility that, like with the aforementioned recent literature on college selectivity, educational production could be improved by more purposeful allocations of students to majors within colleges. Said another way, students across the ability distribution may benefit from placements in high quality majors; future research probing the significance of within-university variability in major quality and student sorting can shed light on this issue.

Finally, we show that the quality of the initial university-by-major cell is a strong predictor of the quality of the final university-by-major cell among degree completers. This is driven in large part by cell persistence, but it is also the case that cell changes tend to be PPI-aligned. An implication is that a pressure point for policy interventions that aim to affect the skill distribution of the workforce through human capital development in college occurs prior to college entry.

References

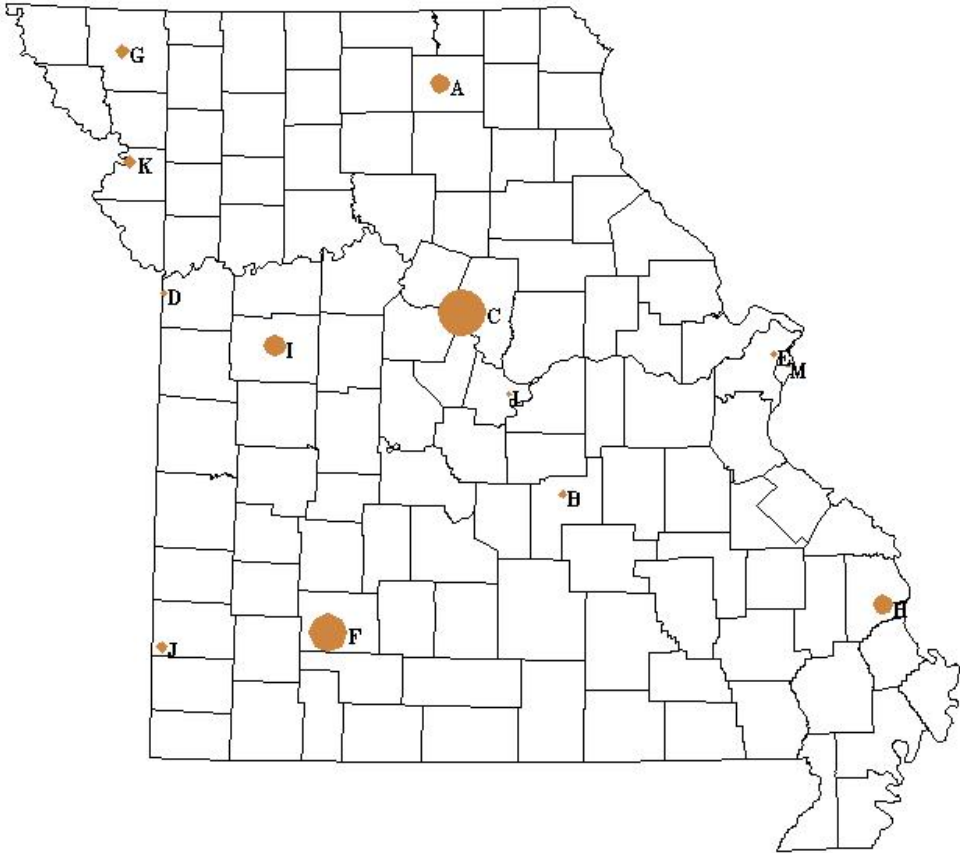
- Aaronson, Daniel, Lisa Barrow and William Sander. 2007. Teachers and Student Achievement in the Chicago Public High Schools. *Journal of Labor Economics* 25(1), 95-135.
- Arcidiacono, Peter. 2004. Ability Sorting and the Returns to College Major. *Journal of Econometrics* 121(1-2), 343-375.
- Arcidiacono, Peter, Esteban Aucejo, and V. Joseph Hotz. University Differences in the Graduation of Minorities in STEM Fields: Evidence from California. *American Economic Review*, 106(3), 525-562.
- Arcidiacono, Peter and Cory Koedel. 2014. Race and College Success: Evidence from Missouri. *American Economic Journal: Applied Economics*, 6(3), 20-57.
- Arcidiacono, Peter and Michael Lovenheim. 2016. Affirmative Action and the Quality-Fit Tradeoff. *Journal of Economic Literature*, 54(1), 3-51
- Astin, Alexander. 1993. *What Matters in College: Four Critical Years Revisited*. San Francisco: Jossey-Bass.
- Black, Sandra, Jane Lincove, Jennifer Cullinane, and Rachel Veron. 2015. Can You Leave High School Behind? *Economics of Education Review* 46, 52-63.
- Bowen, William G., Mathew M. Chingos and Michael S. McPherson. 2009. *Crossing the Finish Line*. Princeton, New Jersey: Princeton University Press.
- Cameron, A. Colin and Douglas L. Miller. 2015. A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, 50(2), 317-372.
- Cameron, Stephen V. and James J. Heckman. 2001. The Dynamics of Educational Attainment for Black Hispanic, And White Males. *Journal of Political Economy* 109(3), 455-499.
- Carnevale, Anthony P., Megan L. Fasules, Andrea Porter and Jennifer Landis-Santos. 2016. African Americans: College Majors and Earnings. Policy Report. Center on Education and the Workforce: Georgetown University.
- Carrell, Scott, Richard Fullerton, and James West. 2009. Does Your Cohort Matter? Measuring Peer Effects in College Achievement. *Journal of Labor Economics*, 27(3), 439-464.
- Chetty, Raj, John N. Friedman and Jonah E. Rockoff. 2014. Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review* 104(9), 2633-79.

- Cohodes, Sarah and Joshua Goodman. 2014. Merit Aid, College Quality and College Completion: Massachusetts' Adams Scholarship as an In-Kind Subsidy. *American Economic Journal: Applied Economics* 6(4), 251-85.
- Committee on Prospering in the Global Economy of the 21st Century. 2007. *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future*. Washington DC: The National Academies Press.
- Deming, David. 2015. The Growing Importance of Social Skills in the Labor Market. NBER Working Paper No. 21473.
- Dillon, Eleanor and Jeff Smith. Forthcoming. Determinants of the Match between Student Ability and College Quality. *Journal of Labor Economics*.
- Fletcher, Jason and Marta Tienda. 2010. Race and Ethnic Differences in College Achievement: Does High School Attended Matter? *The ANNALS of the American Academy of Political and Social Science* 627: 144-166.
- Goldhaber, Dan and Michael Hansen. 2013. Is it Just a Bad Class? Assessing the Long-term Stability of Estimated Teacher Performance. *Economica* 80, 589-612.
- Hamermesh, Daniel S. and Stephen G. Donald, The Effect of College Curriculum on Earnings: An Affinity Identifier for Non-ignorable Non-response Bias. *Journal of Econometrics* 144(2), 479-491.
- Hoekstra, Mark. 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, Caroline and Christopher Avery, 2013. The Missing "One-Offs": The Hidden Supply of High-Achieving, Low-Income Students. *Brookings Papers on Economic Activity* 46(1), 1-65.
- Hoxby, Caroline and Sarah Turner. 2014. Expanding College Opportunities for High-Achieving, Low Income Students. SIEPR Discussion Paper No. 12-014.
- Hurwitz, Michael, Preeya P. Mbekeani, Margaret M. Nipson, and Lindsay C. Page. Forthcoming. Surprising Ripple Effects: How Changing the SAT Score-Sending Policy for Low-Income Students Impacts College Access and Success. *Educational Evaluation and Policy Analysis*.
- Jacob, Brian and Lars Lefgren. 2008. Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education. *Journal of Labor Economics* 26(1), 101-136.
- Koedel, Cory. 2009. An Empirical Analysis of Teacher Spillover Effects in Secondary School. *Economics of Education Review*, 28(6), 682-692.

- Koedel, Cory, Kata Mihaly and Jonah E. Rockoff. 2015. Value-Added Modeling: Review. *Economics of Education Review*, 47, 180-195.
- Long, Mark C., Patrice Iatarola, and Dylan Conger. 2009. Explaining Gaps in Readiness for College-Level Math: The Role of High School Courses. *Education Finance and Policy*, 4(1), 1-33.
- Melguizo, Tatiana. 2010. Are Students of Color More Likely to Graduate From College if They Attend More Selective Institutions? Evidence From a Cohort of Recipients and Nonrecipients of the Gates Millennium Scholarship Program. *Educational Evaluation and Policy Analysis*, 32(2), 230-248.
- Niu, Sunny and Marta Tienda. 2013. High School Economic Composition and College Persistence. *Research in Higher Education*, 54, 30–62.
- Petersen, Mitchell. 2009. Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. *Review of Financial Studies* 22(1): 435-480
- Pike, Gary and Joseph Saupe. 2002. Does High School Matter? An Analysis of Three Methods of Predicting First-Year Grades. *Research in Higher Education*, 43(2), 187-207.
- Porter, Stephen and Paul Umbach. 2006. College Major Choice: An Analysis of Person–Environment Fit. *Research in Higher Education*, 47(4), 429-449.
- Rodriguez, Awilda. 2015. Tradeoffs and Limitations: Understanding the Estimation of College Undermatch. *Research in Higher Education*, 56, 566-594.
- Rothstein, Jesse M. 2004. College performance predictions and the SAT. *Journal of Econometrics* 121(1-2): 297-317.
- Smith, Jonathan, Matea Pender, and Jessica Howell. 2013. The Full Extent of Student-College Academic Undermatch. *Economics of Education Review*, 32.
- St. John, Edward, Shouping Hu, Ada Simmons, Deborah Carter, and Jeff Weber. 2004. What Difference Does a Major Make? The Influence of College Major Field on Persistence by African American and White students. *Research in Higher Education*, 45(3), 209–232.
- Stinebrickner, Ralph and Todd R. Stinebrickner. 2014. A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout. *Review of Economic Studies* 81(1), 426-472.
- Thomas, Scott and Liang Zhang. 2005. Post-Baccalaureate Wage Growth within Four Years of Graduation: The Effects of College Quality and College Major. *Research in Higher Education*, 46(4), 437-459.

- Turner, Sarah. 2017. Education Markets: Forward-Looking Policy Options. Hutchins Center Working Paper #27. Washington, DC: Brookings.
- Umbach, Paul and Matthew Wawrzynski. 2005. Faculty Do Matter: The Role of College Faculty in Student Learning and Engagement. *Research in Higher Education*, 46(2), 153-184.
- Webber, Douglas A. Are College Costs Worth It? How Ability, Major, and Debt Affect the Returns to Schooling. *Economics of Education Review* 53, 296-310.
- Wiswall, Matthew and Basit Zafar. 2015. Determinants of College Major Choice: Identification using an Information Experiment. *Review of Economic Studies*, 82(2), 791-824.
- Zafar, Basit. 2013. College Major Choice and the Gender Gap. *Journal of Human Resources*, 48(3), 545-595.

Figure 1: Geographic Distribution of 4-year Public Universities in Missouri

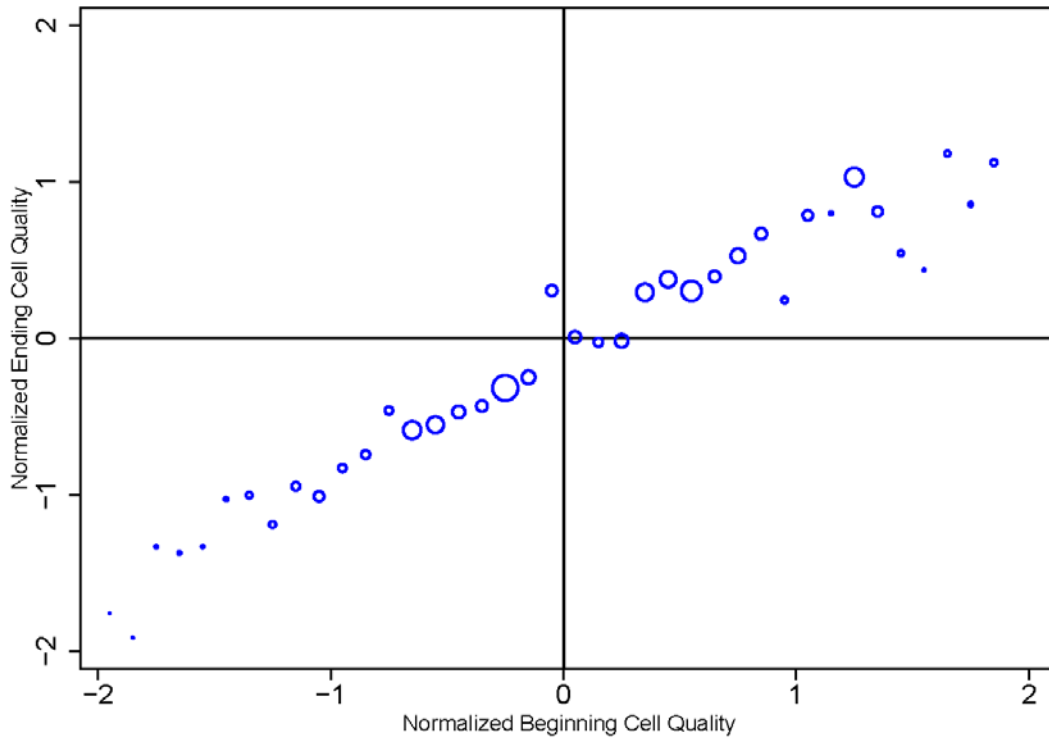


Legend

- A: Truman State University
- C: UM-Columbia
- E: UM-St. Louis
- G: Northwest Missouri State University
- I: University of Central Missouri
- K: Western Missouri State University
- M: Harris Stowe State University
- B: UM-Rolla
- D: UM-Kansas City
- F: Missouri State University
- H: Southeast Missouri State University
- J: Missouri Southern State University
- L: Lincoln University

Note: Circle sizes correspond to enrollment shares from the analytic sample.

Figure 2: Relationship between the PPI of the Final and Initial University-Major Cell



Notes: Graph depicts the relationship between normalized ending university-major cell PPI (on the y-axis) and normalized beginning cell PPI (on the x-axis). Markers are the average ending PPI for the values of beginning PPI, with beginning PPI grouped into bins of 0.1 standard deviations. The size of each marker reflects the number of students in the bin. This chart only includes students who finish.

Table 1. University Descriptive Statistics for Analytic Sample.

University	Average Academic Index Of Entrants	Standard Dev. Academic Index of Entrants	Entry Share	Graduation Rate	Average Academic Index Of Graduates	Standard Dev. Academic Index of Graduates
Overall	2.77	0.73	1.00	0.62	2.97	0.63
Univ of Missouri -Rolla	3.32	0.52	0.04	0.72	3.42	0.46
Truman State Univ	3.12	0.40	0.08	0.78	3.17	0.38
Univ of Missouri-Columbia	3.08	0.59	0.22	0.75	3.16	0.55
Univ of Missouri -Kansas City	3.02	0.65	0.04	0.55	3.09	0.62
Univ of Missouri -St. Louis	2.83	0.65	0.03	0.50	2.87	0.64
Missouri State Univ	2.71	0.66	0.19	0.59	2.88	0.60
Northwest Missouri State Univ	2.61	0.68	0.07	0.64	2.78	0.64
University of Central Missouri	2.59	0.69	0.10	0.60	2.77	0.64
Southeast Missouri State Univ	2.59	0.75	0.09	0.58	2.79	0.68
Missouri Southern State Univ	2.52	0.77	0.05	0.44	2.82	0.68
Western Missouri State Univ	2.19	0.79	0.07	0.41	2.61	0.69
Lincoln Univ	2.05	0.86	0.02	0.39	2.48	0.79
Harris Stowe State Univ	1.94	1.01	0.00	0.30	2.04	1.06

Notes: The analytic sample includes full-time, resident, non-transfer students who entered the system between 1996 and 2001 as college freshman from public high schools. It omits students whose high school of attendance, class rank, and/or ACT scores are unavailable (combined data loss \approx 6 percent). The enrollment shares presented in this table are broadly reflective of the relative sizes of the public universities in Missouri, but can differ from total enrollment shares because we exclude transfer students from community colleges as well as part-time students, and these students are not evenly distributed across the system.

Table 2. Index Parameters from Primary and Alternative Specifications for the Index.

	(1)	(2)
HS Class Percentile Rank	3.12 (0.06)***	
ACT Math Score	0.03 (0.00)***	0.08 (0.00)***
ACT Reading Score	-0.01 (0.00)***	0.01 (0.00)***

Note: Standard errors included in parentheses.

*** p<0.01, ** p<0.05

Table 3. Variance of Student-Level Alignment to University-by-Major Cells with Observed and Counterfactual Sorting Conditions.

	Variance of $M_{i,jm}$
Observed	0.43
<i>Counterfactual Scenarios</i>	
Global Random Assignment	0.60
Global AI-Sorting	0.22
Random Assignment Conditional on Initial University	0.47
AI-Sorting Conditional on Initial University	0.28

Notes: This table reports on the system-wide variance of observed and counterfactual academic alignment, measured by the difference between students' own academic preparation and the PPI of the entering cell. See text for description of counterfactual scenarios.

Table 4. The Explanatory Power of High Schools over the PPI of Student Placements.

	<u>Cell PPI</u> (1)	<u>Cell PPI, Net of University PPI</u> (2)
Coefficient on <i>AI</i> variable	0.33 (0.03)***	0.13 (0.04)***
Total Model R ²	0.269	0.029
Partial R ² Attributable to High School Fixed Effects	0.113	0.014

Note: Standard errors clustered by university-by-major cell and high school are included in parentheses. Cell PPI and the individual academic index are normalized such that estimates can be interpreted as mapping a one-standard-deviation move in a covariate to one standard deviation of the true distribution of PPI.

*** p<0.01

Table 5. Results from High School Covariate Models, Cell PPI.

	(1)	(2)	(3)	(4)	(5)
Academic Index	0.32 (0.04)***	0.33 (0.04)***	0.35 (0.04)***	0.35 (0.04)***	0.34 (0.04)***
% HS Minority	0.01 (0.02)			-0.01 (0.02)	-0.02 (0.02)
% HS FRL		-0.10 (0.01)***		-0.04 (0.01)***	-0.02 (0.01)
Zip % Less than BA			-0.16 (0.01)***	-0.15 (0.01)***	-0.14 (0.02)***
Basic HS Controls					X
R-squared	0.18	0.19	0.22	0.22	0.22

Notes: Standard errors clustered by university-by-major cell and high school are included in parentheses. The basic high school characteristics included in column (5) are indicators for urbanicity (urban, suburban, town, rural, missing) and schools size. Cell PPI, the academic index, Pct Minority, Pct Free/Reduced Price Lunch, and Zip Pct Less than BA are all normalized such that estimates can be interpreted as mapping a one-standard-deviation move in the covariate to one standard deviation of the true distribution of PPI.

*** p<0.01, ** p<0.05, * p<0.10

Table 6. Results from High School Covariate Models, Cell PPI Net of University PPI.

	(1)	(2)	(3)	(4)	(5)
Academic Index	0.12 (0.04)***	0.12 (0.04)***	0.12 (0.04)***	0.12 (0.04)***	0.12 (0.04)***
% HS Minority	-0.01 (0.02)			-0.01 (0.02)	-0.01 (0.02)
% HS FRL		0.01 (0.01)		-0.00 (0.01)	0.00 (0.01)
Zip % Less than BA			0.02 (0.02)	0.02 (0.02)	0.02 (0.02)
Basic HS Controls					X
R-squared	0.02	0.02	0.02	0.02	0.02

Notes: Standard errors clustered by university-by-major cell and high school are included in parentheses. The basic high school characteristics included in column (5) are indicators for urbanicity (urban, suburban, town, rural, missing) and schools size. Cell PPI, the academic index, Pct Minority, Pct Free/Reduced Price Lunch, and Zip Pct Less than BA are all normalized such that estimates can be interpreted as mapping a one-standard-deviation move in the covariate to one standard deviation of the true distribution of PPI.

*** p<0.01, ** p<0.05, * p<0.10

Table 7. Alternative Academic Index without High School Class Rank

	Cell PPI		Cell PPI net of Univ PPI	
	(1)	(2)	(3)	(4)
Academic Index	0.49 (0.05)***	0.49 (0.05)***	0.19 (0.05)***	0.19 (0.05)***
% HS Minority	0.06 (0.02)***	0.04 (0.02)**	0.03 (0.02)	0.02 (0.02)
% HS FRL	-0.03 (0.01)**	-0.00 (0.02)	0.00 (0.01)	0.01 (0.01)
Zip % Less than BA	-0.06 (0.01)***	-0.05 (0.02)***	0.05 (0.01)***	0.05 (0.01)***
Basic HS Controls		X		X
R-squared	0.29	0.30	0.03	0.04

Notes: Standard errors clustered by university-by-major cell and high school are included in parentheses. The basic high school characteristics included in columns 2 and 4 are indicators for urbanicity (urban, suburban, town, rural, missing) and schools size. Cell PPI, the academic index, Pct Minority, Pct Free/Reduced Price Lunch, and Zip Pct Less than BA are all normalized such that estimates can be interpreted as mapping a one-standard-deviation move in the covariate to one standard deviation of the true distribution of PPI.

*** p<0.01, ** p<0.05, * p<0.10

Appendix A Supplementary Tables

Appendix Table A.1: Summary Statistics for Student and High School Characteristics in the Sample

	Mean	SD
<i>Students in the sample</i>		
High School Percentile Class Rank	0.72	0.21
ACT Math Score	22.63	4.76
ACT Reading Score	24.38	5.51
White Male	0.39	0.49
African American Male	0.02	0.15
Asian Male	0.01	0.09
Hispanic Male	0.01	0.07
Other Race Male	0.01	0.11
White Female	0.49	0.50
African American Female	0.04	0.19
Asian Female	0.01	0.09
Hispanic Female	0.01	0.08
Other Race Female	0.01	0.12
<i>High schools in the sample</i>		
City	0.18	0.38
Suburb	0.38	0.48
Town	0.21	0.41
Rural	0.17	0.38
Locale Missing	0.06	0.24
Number of Students (000)	1.12	0.66
Pct Minority (%)	12.11	16.81
Pct Free or Reduced Price Lunch (%)	10.48	15.35
Zip Pct Less than BA (%)	77.09	13.50
Number of Students	58377	
Number of High Schools	455	
Number of University-by-Major Cells	476	

Notes: Student data are from DHE state administrative records. High school data are taken from the Common Core of Data (CCD). Area information (the share of individuals age-25 and older with at least a bachelor's degree in the high school's zip code) comes from the year-2000 United States Census. The high school and local-area averages and standard deviations reported in the table are student weighted.

Appendix Table A.2: Ten Highest and Lowest PPI University-by-Major Cells.

University Level (Selective or Less Selective)	Major	Average AI of Finishers
<i>A. Highest Average AI of Finishers</i>		
Selective University	Mathematics	3.61
Selective University	Biochemistry	3.61
Selective University	Applied Mathematics	3.60
Selective University	Computer Engineering	3.57
Selective University	Physics	3.54
Selective University	Industrial Engineering	3.53
Selective University	Chemical Engineering	3.53
Selective University	Agricultural Engineering	3.25
Selective University	Chemistry	3.20
Selective University	Electrical Engineering	3.19
<i>B. Lowest Average AI of Finishers</i>		
Less selective University	Business Administration	1.86
Less selective University	Social Sciences, General	2.04
Less selective University	Journalism	2.05
Less selective University	Education, General	2.15
Less selective University	Criminal Justice and Corrections	2.17
Less selective University	Parks, Recreation and Leisure Facilities Management	2.25
Less selective University	Criminal Justice and Corrections	2.26
Less selective University	Fine and Studio Arts	2.29
Less selective University	General Sales, Merchandising, & Marketing Operations	2.33
Less selective University	Psychology	2.34

Note: Cells displayed in these tables are restricted to those with at least 40 graduates. University names are masked to preserve anonymity; in total, the cells listed in the table are spread across seven of the thirteen universities in the system. “Selective” universities are those with an undergraduate profile considered “more selective” or “selective” in the 2015 Carnegie Classifications of Higher Education. “Less selective” universities in this table are universities with undergraduate profiles that are not considered as selective as “selective” colleges. See <http://carnegieclassifications.iu.edu>.

Appendix Table A.3: Sensitivity Analysis: Dropping Cells without Finishers.

	Cell PPI	Cell PPI net of Univ PPI
	(1)	(2)
Academic Index	0.36 (0.03)***	0.18 (0.05)***
% Minority	-0.04 (0.03)	-0.04 (0.02)
% FRL	-0.02 (0.02)	0.01 (0.02)
Zip % Less than BA	-0.15 (0.02)***	0.02 (0.02)
Basic HS Controls	X	X
R-squared	0.21	0.03

Notes: Standard errors clustered by university-by-major cell and high school are included in parentheses. The basic high school characteristics include school size and indicators for urbanicity (urban, suburban, town, rural, missing). Cell PPI, the academic index, Pct Minority, Pct Free/Reduced Price Lunch, and Zip Pct Less than BA are all normalized such that estimates can be interpreted as mapping a one-standard-deviation move in the covariate to one standard deviation of the true distribution of PPI.

*** p<0.01, ** p<0.05, * p<0.10

Appendix Table A.4: Sensitivity of Primary Findings (Tables 5 & 6, Column 5) to Replacing the Linear AI Control with a 20-Bin AI Control Set.

	<u>Cell PPI</u>	<u>Cell PPI net of Univ PPI</u>
	(1)	(2)
% Minority	-0.02 (0.02)	-0.01 (0.02)
% FRL	-0.02 (0.01)	0.00 (0.01)
Zip % Less than BA	-0.14 (0.02)***	0.02 (0.02)
Basic HS Controls	X	X
R-squared	0.23	0.02

Notes: Standard errors clustered by university-by-major cell and high school are included in parentheses. The basic high school characteristics include school size and indicators for urbanicity (urban, suburban, town, rural, missing). Cell PPI, Pct Minority, Pct Free/Reduced Price Lunch, and Zip Pct Less than BA are all normalized such that estimates can be interpreted as mapping a one-standard-deviation move in the covariate to one standard deviation of the true distribution of PPI. Students are divided into twenty equal-sized bins based on their AI values and we control for the AI bins (coefficients not displayed) in place of the linear AI control used in the main text. This allows for a flexible, highly non-linear relationship between AI and the quality of the university-by-major placement but has no bearing on our findings qualitatively.

*** p<0.01, ** p<0.05, * p<0.10