NBER WORKING PAPER SERIES

USING PREDICTIVE ANALYTICS TO TRACK STUDENTS: EVIDENCE FROM A SEVEN-COLLEGE EXPERIMENT

Peter Bergman
Elizabeth Kopko
Julio E. Rodriguez

Working Paper 28948
http://www.nber.org/papers/w28948

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2021

The research reported here was undertaken through the Center for the Analysis of Postsecondary Readiness and supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C140007 to Teachers College, Columbia University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Numerous people—including several from the Community College Research Center at Teachers College, Columbia University—have helped make this work happen. We particularly thank Elisabeth Barnett for her leadership, as well as Clive Belfield, Magdalena Bennett, Dan Cullinan, Vikash Reddy, and Susha Roy for their contributions. We also thank Judy Scott-Clayton for her comments and advice. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Using Predictive Analytics to Track Students: Evidence from a Seven-College Experiment
Peter Bergman, Elizabeth Kopko, and Julio E. Rodriguez
NBER Working Paper No. 28948
June 2021
JEL No. I0,I20,I24

ABSTRACT

Tracking is widespread in U.S. education. In post-secondary education alone, at least 71% of colleges use a test to track students. However, there are concerns that the most frequently used college placement exams lack validity and reliability, and unnecessarily place students from under-represented groups into remedial courses. While recent research has shown that tracking can have positive effects on student learning, inaccurate placement has consequences: students face misaligned curricula and must pay tuition for remedial courses that do not bear credits toward graduation. We develop an alternative system to place students that uses predictive analytics to combine multiple measures into a placement instrument. Compared to colleges' existing placement tests, the algorithm is more predictive of future performance. We then conduct an experiment across seven colleges to evaluate the algorithm's effects on students. Placement rates into college-level courses increased substantially without reducing pass rates. Adjusting for multiple testing, algorithmic placement generally, though not always, narrowed gaps in college placement rates and remedial course taking across demographic groups. A detailed cost analysis shows that the algorithmic placement system is socially efficient: it saves costs for students while increasing college credits earned, which more than offsets increased costs for colleges. Costs could be reduced with improved data digitization, as opposed to entering data by hand.

Peter Bergman
Department of Education Policy
and Social Analysis
Columbia University
525 W. 120th Street
New York, NY 10027
and NBER
bergman@tc.columbia.edu

Elizabeth Kopko
Columbia University
525 W 120th St Box 174
New York, NY 10027
e.kopko@columbia.edu

Julio E. Rodriguez
Columbia University
525 W 120th St.
New York, NY 10027
julio.rodriguez@columbia.edu

1. Introduction

Tracking students by prior test scores is widespread in U.S. education. In higher education alone, at least 71% of post-secondary institutions use a test to track students (Fields & Parsad, 2012; National Center for Public Policy and Higher Education & Southern Regional Education Board, 2010). These rates are higher in two-year colleges, which enroll nearly half of post-secondary students but graduate 39% of their students (Fields & Parsad, 2012; Chen, 2016).[1] While recent research has demonstrated large potential benefits of tracking (Card & Giuliano, 2016; Duflo, Dupas, & Kremer, 2011), *inaccurate* placement has consequences: students face misaligned curriculum and, in higher education, must pay tuition for remedial courses that do not bear credits toward graduation. Inaccurate placement is a concern because there is evidence that placement exams lack validity and reliability, and they unnecessarily track students from under-represented groups into remedial courses (Rothstein, 2004; Scott-Clayton et al., 2014).[2] Given most placement tests aim to predict students' readiness for college-level courses, additional measures, such as high school GPA, when combined with predictive analytics, could mitigate concerns about validity and fairness (Rothstein, 2004; Scott-Clayton et al., 2014; Mullainathan & Spiess, 2017).[3]

In this paper, we develop a placement algorithm that combines multiple measures using predictive analytics and implement it via an experiment with 12,544 students. To do so, we recruited seven community colleges across New York and gathered historical

---

[1] This is the graduation rate within six years.
[2] See Heubert & Hauser (1999) for more information about key characteristics for sound testing instruments.
[3] Alternatively, a measure could be constructed to predict treatment effects of specific course placements as opposed to pass rates or readiness. In practice, targeting treatment effects does not seem to be how colleges explicitly try to optimize placement systems.

data on their students to estimate models predicting students' likelihood of passing college-level math and English courses. These predictions incorporated measures such as placement-exam scores, high school GPA, high school rank, diploma status, and time since high school graduation. We created college-specific placement algorithms for math and English that placed students into a remedial course if a student's predicted probability of passing a college-level course was below a cut point chosen by each college. We then randomly assigned students to either colleges' business-as-usual placement system or the placement algorithm.

For several reasons, it is unclear what impacts this algorithm will have on student outcomes. Improving the validity and the reliability of the placement instrument could help place students into courses better aligned to their incoming skills. Measures such as high school GPA reflect a wider array of cognitive and non-cognitive skills than test scores alone (Kautz, et al., 2014; Kautz & Zanoni, 2014; Borghans, Golsteyn, & Heckman, 2016).[4] However, estimating the algorithm requires overcoming the selective labels problem (Kleinberg et al., 2018). Selection into college-level courses is based on observables, but the algorithm's predictions still rely on extrapolations that could reduce their validity. Experimental evaluation of algorithmic placement is important for testing how well it performs in practice.

The algorithm also helps colleges choose cut points for placements into the college-level courses, which affect the number of students placed into these courses and their expected pass rates conditional on placement. This choice means algorithmic placement does not necessarily imply that placement rates will change either on net or for a given individual. At particular thresholds (e.g., the extremes), it is possible that the placements

[4] GPA also has a high degree of reliability, but there are concerns that grading standards too school-specific for it to be useful (Bacon & Bean, 2006).

assigned by the algorithm and the test score will be the same. But if colleges choose to maintain pass rates, the algorithm may place more students into college-level courses (Scott-Clayton et al., 2014), which could increase students' credit accumulation and save students money if the algorithm's predictions are sufficiently accurate.

We show how colleges implemented the placement algorithm, how it affected students' placement outcomes, and what impacts this had on credit accumulation and costs. We find that colleges generally chose cut points to hold *pass* rates constant. This resulted in large changes in placement rates: relative to colleges' business as usual, 20% of math placements changed and 40% of English placements changed. Compared to what would have occurred using the business-as-usual placement tests, the algorithm placed 12% into a higher-level math class and 34% into a higher-level English course. The algorithm placed 8% of students into a lower-level math course and a 6% lower-level English course.

Placement via the algorithm led to immediate increases in enrollment into college-level courses. Algorithmic placement yielded first-term enrollment increases in college-level math by 2.6 percentage points and in college-level English by 13.6 percentage points relative to the control group. Roughly two-thirds of the treatment group complied with their algorithmic placement recommendation.

Algorithmic placement also led to reductions in remedial course taking and increases in college credits earned—*without* reducing pass rates. Placement via the algorithm reduced remedial course taking by 1.1 credits and—using treatment as an instrument for compliance with their placement recommendation—by 2.3 credits for algorithmic placement compliers. Assignment to algorithmic placement increased college credits by 0.53 credits and by 1.1 credits for compliers. These changes are larger for students who are being tracked in *both* math and English: these students earned 1.2 more college credits and compliers earned 2.1 college credits because of algorithmic placement.[5]

The magnitude of these effects persists for our earliest cohort of students whom we observe for 2.5 years. Across all terms, students in this cohort earned 1.6 college credits more and attempted 1.2 fewer remedial credits than the control group. For students tracked in both math and English, college credits increased by 2.0 credits and remedial credits decreased by 1.2 credits. These effects are roughly 50% larger for compliers.

We also find evidence algorithmic placement narrows certain demographic gaps in placement rates. Though the algorithm increased placement into college level courses for all subgroups we looked at, increases were significantly larger for female students in math relative to male students, Black students in English relative to white students, and lower-income students took fewer remedial credits relative to higher-income students—after controlling for multiple-hypothesis-testing. However, the increases in placement into college-level math, though positive and significant, were not as large for Hispanic students compared to white students.

The algorithmic placement system also results in cost savings for students. We conducted a detailed cost analysis for colleges and students, separating fixed and variable costs, and costs to students versus costs to colleges. We find that students saved $150, on average relative to the business as usual, which is due to reductions in remedial course taking. These savings were $310 dollars for compliers. This implies an average saving to students equal to $871,200 per cohort, on average.

For colleges to implement such a placement algorithm, decision makers must weigh the potential benefits to students against the costs to the colleges. We estimate that the

[5] Each college has automatic exemptions from taking a placement exam for a given subject, and, because our placement mechanism was integrated within the testing platforms, not all students could be placed by the algorithm for both math and English.

cost per student in the initial year of the study—above and beyond the business as usual—is $70 to $360 dollars, depending on the college. Much of these costs are driven by the need to hand enter data from high school transcripts; processes or technologies to ease this data collection would greatly reduce costs. Furthermore, the first year of implementation involves large fixed costs. We estimate operating costs of the placement algorithm are $40 dollars per student. Again, this cost would be even lower if colleges collected historical data from students more efficiently. College administrators asked whether further savings, without a loss in quality, could arise by not paying to use placement exams. We examine the extent to which the algorithm would place students differently if test scores were not used for prediction. We find that placement rates would change substantially more for math courses than for English courses; for English courses, only 5 to 8 percent of placements would change.

Our research contributes to a broader literature that focuses on the effects of tracking. Historically, tracking is controversial. Oakes (1985) argued that the evidence on tracking is inconsistent, and, in practice, higher-track classes tend to have higher-quality classroom experiences than lower-track classes. More recently, Duflo, Dupas, & Kremer (2011) randomized students in Kenya to schools that either tracked students by test scores or assigned students randomly to classrooms. They found that test scores in schools with tracking improved relative to the control group, both for students placed in the higher-scoring and the lower-scoring tracks. Card and Giuliano (2016) studied a district policy in which students are placed into classrooms based on their test scores. This program caused large increases in the test scores of Black and Hispanic students.

A number of studies look at the effects of being placed into a higher track versus a lower track. Bui, Craig, and Imberman (2014) and Card and Giuliano (2014) find that gifted students' placement into advanced coursework does not change test scores. Cohodes

(2020) and Chan (2020), however, find increases in enrollment in advanced high-school coursework and college.[6] In higher education, the evidence that placement into remedial courses improves academic outcomes for marginal students is more mixed, and several regression-discontinuity analyses find no effects (Calcagno & Long 2008; Bettinger & Long 2009; Boatman & Long 2010; Martorell and McFarlin 2011; Allen & Dadgar 2012; Hodara 2012; Scott-Clayton & Rodriguez, 2015).

Recently, economists have argued that data-driven algorithms can improve human decision making and reduce biases (Mullainathan & Spiess, 2017; Li, Raymond and Bergman, 2020; Arnold, Dobbie and Hull, 2021). Kleinberg et al. (2018) show that a machine-learning algorithm has the potential to reduce bias in bail decisions relative to judges' decisions alone. At the same time, others are concerned that these algorithms could embed biases into decision making and exacerbate inequalities (Eubanks, 2018). We contribute to this literature by comparing the impacts of a simple, data-driven algorithm to another quantitative measure, test scores. We then evaluate the algorithm by conducting a large-scale experiment.

The rest of this paper proceeds as follows. Section 2 provides further background information about tracking in postsecondary institutions and study implementation. Section 3 describes the experimental design, data and empirical strategies. Section 4 presents our findings. Section 5 provides a detailed cost analysis, and Section 6 concludes.

2. Background, Site Recruitment, Algorithm Implementation

Tracking students into remedial education is a major component of the higher education system, both in terms of enrollment and cost. In the 2011-12 academic year, 41% of first-

[6] Several other studies look at the effects of placing into high-test score schools and the results are much more mixed (Jackson 2010; Pop-Eleches and Urquiola 2013; Abdulkadiroglu, Angrist, and Pathak 2014; Dobbie and Fryer 2014).

and second-year students at four-year institutions had taken a remedial course, while at two-year institutions, even more—68% of students—had taken a remedial course (Chen, 2016). The cost of remedial education has been estimated to be as much as $2.9 billion annually (Strong American Schools, 2008).

The primary purpose of remedial education is to provide differentiated instruction to under-prepared students so they have the skills to succeed in college-level coursework (Bettinger & Long, 2009). However, there is evidence that community-colleges' tracking systems frequently "under place" students—tracking them into remedial courses when they could have succeeded in college-level courses—and "over place" students—tracking them into college-level courses when they were unlikely to be successful (Belfield & Crosta, 2012; Scott-Clayton, 2012).

Most institutions administer a multiple-choice test in mathematics, reading, and writing to determine whether incoming students should be placed into remedial or college-level courses. The ACCUPLACER, a computer-adaptive test offered by the College Board, is the most widely used college placement system in the U.S. (Barnett & Reddy, 2017). Colleges choose a cut score for each test and place students scoring above this score into college-level courses and students below the cut score into various remedial courses.[7] Given the placement rules and immediate test results provided by the ACCUPLACER platform, students often learn their placement immediately after completing their exam.

Site Selection and Descriptions

All the participating colleges are part of the State University of New York (SUNY) system, ranging from large to small, and students' backgrounds vary from college to college. Table A.3 of the Appendix provides each colleges name and an overview of their

[7] Certain colleges may offer exemptions from testing; for instance, this can occur for students who speak English as a second language or who have high SAT scores.

characteristics using public data. The smallest of the colleges serves roughly 5,500 students while the largest serves over 22,000 students. As is common in community college settings, a large share of the student body is part-time and many are adult learners, with between 21% and 30% of students over the age of 25. For most of the colleges, the majority of students receive financial aid. The colleges have similar transfer-out rates of between 18% and 22% and three-year graduation rates are between 15% and 29%. The colleges also tend to serve local student populations. Lastly, all of the colleges have an open-door admissions policy. This means that the colleges do not have admission requirements beyond having graduated from high school or earned a GED.

Creating the Placement Algorithm

Colleges preferred that we develop college-specific algorithms. We created separate algorithms for each college in math and English using data on each college's previous cohorts of students.

Five colleges in the study had been using ACCUPLACER for several years. One college had been using ACCUPLACER assessments for English but had transitioned from a home-grown math assessment to the ACCUPLACER math assessments too recently to generate historical data, so we tested an algorithm for English placement only at that college. One college in our sample had been using the COMPASS exam, which was discontinued by ACT shortly after this study began. The college replaced the COMPASS exam with the ACCUPLACER exam. At this college, we tested an algorithm that does not use any placement test scores against a placement system that incorporates only ACCUPLACER test results.

We worked with administrators at each college to obtain the data needed to estimate each algorithm. In some instances, these measures were stored in college databases. In other instances, colleges maintained records of high school transcripts as digital images. For the latter, we had the data entered into databases by hand.

In order to estimate the relationships between predictors in the dataset and performance in initial college-level courses, we restricted the historical data to students who took placement tests and who enrolled in a college-level course without first taking a remedial course. This set of students constituted our estimation sample for developing the algorithm. Importantly, students were selected into college-level courses based on observable characteristics, but this sampling scheme does raise concerns about whether the relationships we estimate between variables will apply to all students. The experiment will test the extent to which the assumptions implicit in this estimation matter in this context.

We aimed to predict "success" in the college course for each student. We met with college personnel to decide how to define success, which we agreed to define as a grade of C or better in the initial college-level course associated with the placement decision. We then regressed an indicator for success in the relevant course on various sets of predictors using Probit and linear probability models (LPM). We used the results of the LPMs because we could not code non-linear models into colleges' existing placement software. Nonetheless, the non-linear models yielded similar placement decisions as LPMs, especially around the relevant cut points that colleges chose to determine placements.

For each college, we estimated regressions relating placement test scores and high school GPA to "success" in initial college-level classes for a given subject. We added additional information from high school transcripts when such information was available. This information included the number of years that have passed since high school completion and whether the diploma was a standard high school diploma or a GED (diploma status). We also tested the benefit of including of additional variables such as

SAT scores, ACT scores, high school rank, indicators for high school attended, and scores on the New York Regents exams, when they were available (often these were missing), as well as interaction terms and higher-order terms for variables. When variables were missing, we imputed values and added indicators for missing. Identical procedures were followed for both English and math.

[1] $1(\text{C or Better})_i = \alpha + (\text{HS GPA}_i)\beta_1 + \varepsilon_i$

[2] $1(\text{C or Better})_i = \alpha + (\text{ACCUPLACER}_i)\beta_1 + \varepsilon_i$

[3] $1(\text{C or Better})_i = \alpha + (\text{HS GPA}_i)\beta_1 + (\text{ACCUPLACER}_i)\beta_2 + \varepsilon_i$

[4] $1(\text{C or Better})_i = \alpha + (\text{HS GPA}_i)\beta_1 + (\text{ACCUPLACER}_i)\beta_2 + X_i\beta_3 + \varepsilon_i$

The focus of this analysis was the overall predictive power of the model. As such, we calculated the Akaike Information Criterion (AIC) statistics for each model. The AIC is a penalized-model-fit criterion that combines a model's log-likelihood with the number of parameters included in a model (Akaike, 1998; Burnham and Anderson, 2002; Mazerolle, 2004).[8] In practice, we did not have many variables to select from and higher-order and interaction terms had little effect on prediction criteria (and additional complexity was difficult to implement). We estimated the models on prior years of historical data excluding the most recent year, and then examined the fit criteria using data from that most recent year.

Placement exam scores explained very little variation in English course outcomes, slightly more variation in math outcomes, and including additional measures adds explanatory power. Figures A.1 and A.2 in the Appendix list the full set of variables used by each college to calculate students' math and English algorithm scores, respectively. Tables A1 and A2 show typical examples of our regression results for math and English.

[8] Under certain conditions, choosing model specifications according to the AIC is asymptotically equivalent to leave-one-out-cross-validation (Stone, 1977).

Across colleges, explanatory power was much higher for math course grades than for English course grades. Placement scores typically explained less than 1% of the variation in passing grades for English. Test scores were better predictors for passing math grades, explaining roughly 10% of the variation. Adding high school grades typically explained an additional 10% of the variation in both subjects. Interestingly, we found that indicators for which high school a student attended, which could reflect different grading standards, added little predictive value. Overall, combining multiple measures with predictive analytics is no panacea for predicting future grades, but it does improve the validity of the placement instrument relative to test scores alone.

Setting cut probabilities: After we selected the final models, we used the coefficients from the regression to simulate placement rates for each college using their historical data. Consider the following simplified example where a placement test score (R) and high school GPA (G) are used to predict success in college-level math (Y), defined as earning a grade of C or better. The regression coefficients combined with data on R and G can then predict the probability of earning a C or better in college-level math for incoming students ($\widehat{\mathrm{Y}}$). A set of decision rules must then be determined based on these predicted probabilities. A hypothetical decision rule would be:

$$\mathrm{Placement}_{\mathrm{i}} = \begin{cases} \text{College Level if } \widehat{\mathrm{Y}}_{\mathrm{i}} \geq 0.6 \\ \text{Remedial if } \widehat{\mathrm{Y}}_{\mathrm{i}} < 0.6 \end{cases}$$

For each college, we generated spreadsheets projecting the share of students that would place into college-level coursework at any given cut-point as well as the share of those students we would anticipate earning a C or better. These spreadsheets were provided to colleges so that faculty in the pertinent departments could set cut-points for students entering their programs.

Figure 1 shows an abbreviated, hypothetical example of one such spreadsheet provided to colleges.[9] The top panel shows math placement statistics and the bottom panel shows statistics for English. The highlighted row shows the status quo at the college and the percent of tested students placed into college level is shown in the second column. For instance, for math, the status quo placement rate is 30%. The third column shows the pass or success rate, which is a grade "C" or better in the first college-level course in the relevant subject. In this example, the status-quo pass rate for math is 50% conditional on placement into the college-level math course.

Below the highlighted row, we show what would happen to placement and pass rates at different cut points for placement. The first column shows these cut points ("Minimum probability of success"). For instance, for math, the first cut point we show is 45%, which implies that for a student to be placed into college-level math under the algorithm, the student must have a predicted probability of receiving a "C" or better in the gate-keeper math course of at least 45%. If this 45% cut point is used, columns two and three show what would happen to the share of students placed into college-level math under the algorithm (column two) and what would happen to the share who would pass this course conditional on placement (column three). In this example, for math, if the 45% cut point is used, the algorithm would place 40% of students into college-level math and we anticipate 60% of those students would pass. The cut point differs from the expected pass rate because the cut point is the lowest probability of passing for a given student; the cut point implies that every student must have that probability of passing or greater. For instance, if the cut point is 40%, then every student has 40% chance or greater of passing the college-level course. Therefore, most students placed into college-level courses according to this rule will have above a 40% chance of passing the course.

[9] In practice, we showed results from many different cut points.

Faculty opted to create placement rules that kept pass-rates in college-level courses the same as historical pass rates. In general, this choice implied increases in the predicted number of students placed into college-level coursework. For instance, in the example, the status quo placement and pass rates for English are 60% and 40%, respectively. A cut point of 45% would induce the same pass rate, 60%, but would place 75% of students into the college-level English course.

Installation of new placement method in college systems: We developed two procedures to implement the algorithms while maintaining the timing of placement decisions. At colleges running our algorithm through the computerized ACCUPLACER-test platform, we programmed custom rules into the ACCUPLACER platform for students selected to be part of the treatment group.[10]

Other colleges ran their placement through a custom server built for the study. Student information was sent to servers to generate the probability of success and the corresponding placement, which was returned to the college.

3. Experimental Design, Data, Empirical Strategy

The sample frame consisted of entering cohorts (fall and spring) enrolling at each college who were required to take the placement exams from 2016 until 2018.[11] Random assignment was at the student level and stratified by college. We integrated the assignment procedure into each college's placement platform described above, such that upon taking their placement exams, students were randomly assigned to be placed using

[10] This process in particular placed constraints on the algorithm's complexity—interaction terms and non-linear models, for instance, are difficult to implement within the ACCUPLACER system.
[11] Colleges preferred to use alternative placement processes for English as a Second Language speakers, and students with high SAT scores or 4.0 GPA were sometimes exempt from placement exams. Note that, as these are non-selective colleges, few students take the SAT. We report exemption rates in Table 3.

either the business-as-usual method or the algorithm. Students and their instructors were blinded to their treatment assignment. If a student took both the English and math placement exams, they were either assigned to the business-as-usual placement for both subjects or the algorithm for both subjects. Some students only took a placement exam in one subject. After taking placement exams, students were notified of their placements either by an administrator or through an online portal, depending on the college.

This experimental design allowed for a well-powered study, given constraints. We interviewed faculty and staff to document any perceived changes they saw in the composition of classrooms and any responses to these changes. As we describe below, faculty did not perceive changes to their classroom compositions and so did not make changes to the curriculum or teaching. Given that prior evidence suggests that tracking can allow instructors to target instruction more effectively (cf. Card & Giuliano, 2016 and Duflo, Dupas, & Kremer, 2011), our results may present a lower bound on effectiveness if instructors were to change their behaviors in response to more significant changes in their classroom compositions.

Data

Data came from three sources: placement test records, administrative data from each college, and qualitative data on implementation and quantitative data on costs was collected from faculty, counselors, and staff using interviews and focus groups. Student-level placement test records include indicators for each students' placement level in math and English, as well as the information that would be needed to determine students' placements regardless of treatment status. Placement test records from each college contained high school grade point averages (when available) and scores on individual placement tests. Additional variables included in placement test records varied by each college's placement algorithm. Examples of additional variables incorporated for certain colleges include the number of years between high school completion and college enrollment, type of diploma (high school diploma vs. GED), SAT scores, and New York State Regents Exam scores.

In addition to placement test records, college administrative data included demographic information, such as gender, race, age, financial aid status, and transcript data that provided course levels, credits attempted and earned, and course grades.

Table 1 shows sample baseline characteristics for students who participated in the study at each college and overall. Our sample consists of 12,544 first-year students across the seven colleges. On average, students in the sample were 43% white and 43% received a federal Pell Grant. There is some variation in demographic characteristics. For instance, Colleges 1, 2, and 3 serve more white students compared to Colleges 5 and 7, which enroll a higher share of Hispanic students. Using Pell Grant receipt as a proxy for income, average family income for study participants also varies across colleges; Pell Grant receipt ranges from 32 percent to 56 percent of students. Comparing these characteristics to Appendix Table A.3 shows that the study sample characteristics match the overall characteristics of students each college serves.

Lastly, a concern is that using a test score as the primary criterion for assignment systematically under places students from one demographic versus another. Figure 2 highlights descriptive results consistent with this concern by showing the gap in placement rates across demographic subgroups. The first two bars show that the white students are placed into college-level math and English courses at rates 14 percentage points and 19 percentage points higher than Black students. These gaps tend be smaller between Hispanic and white students, and between male and female students, but also quite large between Pell recipients and non-Pell recipients—16 and 12 percentage points for math and English, respectively. The experiment will allow us to compare the algorithm's placement rates by subgroup relative to the status quo and students' success rates in these courses as well. If students can be placed into college-level courses at higher rates without sacrificing pass rates, this would indicate students are being under placed.

Outcomes

We study the effects of assignment to the placement algorithm on several primary outcomes, by subject. First, we examine how placements changed as a result of the algorithm: what share of treated students had their placement changed relative to the status quo, and of these, what share had their placement changed from a remedial-course assignment to a college-level assignment, and what share had their placement changed from a college-level course assignment to a remedial assignment. Second, we show treatment effects on enrollment and pass rates for math and English separately. Lastly, we study college and remedial credits attempted and completed. We show these results in the short run—the first term after placement—as well the longer run for subsample of students we observe for more than two years.

Empirical Strategy

We use an intent-to-treat analysis to examine the impacts of using the placement algorithm versus the single-placement test status quo. We estimate the following model:

$$[7]\quad Y_i = \alpha + \beta \mathrm{Treatment}_i + \lambda\varphi_i + \eta X_i + \delta Z_i + \varepsilon_i \ ,$$

where Y_i are academic outcomes for student *i*, such as placement into a college-level course and passing a college-level course; $\mathrm{Treatment}_i$ indicates whether the individual was randomly assigned to be placed using the algorithm or the business as usual; φ_i is a vector of indicators for the institution (strata) a student attends; X_i is a vector of baseline covariates (gender, race, age, financial aid status); Z_i is students' math and English algorithm scores, which are baseline measures of academic preparedness, and ε_i is the error term. The coefficient of interest is β, which is the effect of assignment to the placement algorithm on outcomes at the end of the first semester discussed above. We estimate Huber-White-Heteroskedasticity robust standard errors (Huber, 1967; White, 1980) following the experimental design (Abadie et al., 2020).

As not everyone takes a placement exam in both subjects, we estimate these regressions for those who took any placement exam, and therefore could be assigned to placement by the algorithm for one *or* two courses, and we also estimate these regressions for those who took placement exams in both subjects, and therefore could be assigned to placement by the algorithm for two subjects.

Treatment-on-the-Treated Analyses

Because not everyone follows their recommended placement, we also estimate treatment-on-the-treated effects for those who comply with their placement recommendation. Compliance here is an indicator equal to one for following the algorithm's placement recommendation and equal to zero for the control group if they follow the business as usual. For students who took both math and English exams, compliance is defined as following the algorithm's recommendation in at least one subject. The second stage equation is as follows:

$$[8] \quad Y_i = \alpha_{iv} + \beta_{iv} Placement\_comply_i + \lambda_{iv}\varphi_i + \eta_{iv}X_i + \delta_{iv}Z_i + \eta_i \ .$$

We instrument the endogenous compliance variable with treatment assignment. This analysis will estimate local average treatment effects—effects on those who comply with their placement recommendation. We show control complier means in each IV results table.

Subgroup Analyses

We also study the potential disparate impact the placement algorithm has on the composition of students placed into remedial and college-level courses. We estimate equation [7] above for each subgroup, but also test the significance of the interaction terms, shown below.

[9] $Y_i = \alpha_k + \beta_{1k}\text{Treatment}_i + \beta_{2k}\text{Treatment}_i \times \text{Subgroup}_k + \lambda_k\varphi_i + \eta_k X_i + \delta_k Z_i + \varepsilon_{ki}$.

The outcomes, Y_i, are placement in college-level math, placement in college-level English, and credit accumulation. For each subgroup of interest, the sample is restricted to the reference group and the subgroup. Therefore, the coefficient β_{1k} shows the effect for the reference group (listed below), and the coefficient of particular interest is the significance and magnitude of β_{2k}, which indicates whether the difference between groups of students is widening or narrowing as a result of algorithmic placement. The subgroups of interest are Black students, Hispanic students (compared to white students); female students (compared to male students); and Pell recipients (compared to non-Pell recipients). This process yields many tests, which increases the likelihood of type-I errors. To control for the Family Wise Error Rate, we use the step-down procedure formulated by Holm (1979).

Treatment-Control Baseline Balance

Randomization should ensure that, in expectation, students assigned to the treatment are similar to those assigned to the control-group placement rules. Table 2 provides evidence that random assignment was successfully implemented. Participants' demographic and academic characteristics are balanced across treatment and control groups. Students' ACCUPLACER exam scores also are similar across both groups. Overall, the magnitudes of differences between treatment and control groups are small and only one is significant at the 5 percent level, which is expected with the more than 20 variables tested. Though not shown, this balance also holds for the subgroup of students who took both the English and math placement exams as well.

4. Results

Descriptive Changes in Placements

We begin with a descriptive summary of placement changes to show the various ways the algorithm changed students' placements relative to the business as usual. As stated above, it is not obvious how the algorithm will change net placement rates. Table 3 summarizes these changes for students placed by the algorithm. Of the more than 6,000 students assigned to the program-group, 94% were tracked in math and 80% were tracked in English. Among those students who took a math placement exam, 21% experienced a math placement different from what would have been expected under the status quo placement rules. Of those with a changed math placement, 61% were placed into a higher-level math course than would have been expected under a single test placement system, and 39% placed in a lower-level math course. Of those who took the English placement exams, approximately 50% of program-group students experienced a change in the level of their English level placement, of which 86% placed into a higher-level English course and 14% placed into a lower-level course than they would have under the status quo placement strategy.

Table 4 shows compliance with placement recommendations. Overall, the treatment group complies with their algorithmic placement recommendation 62% of the time. Treatment assignment increases compliance with the algorithm's decision relative to the control group by 48 percentage points. The first stage is slightly larger for the spring cohort, when there are fewer first-time enrollees, but is generally consistent.

Treatment Effects on Placement, Course Taking, and Credits

The placement algorithm resulted in increases in all of the outcomes: placement into college-level courses, enrollment in college-level courses, and total college-level credits earned. Table 5 summarizes the first-term results. Students assigned to the placement algorithm are 6.6 percentage points more likely to be placed into a college-level math course, 2.6 percentage points more likely to enroll in a college-level math course, and 1.9 percentage points more likely to pass a college-level math course during the first term. All of these results are statistically significant at the 1 percent level. One explanation for the difference between placement and enrollment into a college-level math course is that some students placing into college-level math did not have to complete a college-level math course prior to enrolling in *other* college-level courses in the first term.

There are positive and substantially larger effects for English placement, enrollment, and completion than for outcomes on math courses. Students who were placed by the algorithm were 32 percentage points more likely to place into a college-level English course, 14 percentage points more likely to enroll in a college-level English course, and 7 percentage points more likely to pass a college-level English course in the first term. All results are significant at the 1 percent level. Again, the difference between placement and enrollment into a college-level English course may occur for the same reason as above for college-level math enrollment.

We also find reductions in total remedial credits taken—irrespective of subject—and increases in total college credits earned. These effects are generally larger for students who are placed via the algorithm in both math and English. Panel A of Table 6 shows intent to treat effects while Panel B shows treatment-on-the-treated effects. The first three columns in each panel show results for students who took any placement exam while columns four through six show results for students who took a placement exam in both subjects (and so are tracked in both courses).

Panel A of Table 6 shows that algorithm assignment reduced remedial credits attempted by 1.1 credits relative to a mean of 3.5 credits—a 31% reduction. Panel B shows that, for those who complied with their placement recommendation, algorithmic placement reduced remedial course taking by 2.3 credits relative to a complier mean of 5.9 credits (a 38% reduction). The reductions are slightly smaller for those who were tracked in both English and math: the ITT is -1.1 remedial credits and TOT is -1.7 credits.

Table 6 also shows there is an increase in credit accumulation of 0.53 credits for those tracked in at least one subject and 1.2 credits for those tracked in both subjects. For those who complied with their placement, these effects are 1.1 credits and 2.1 credits, respectively. These net positive effects reflect that students are not entirely substituting math and English college credits for other college-level credits.

The increase in college-level placement does not result in a reduction in pass rates. We can calculate pass rates by dividing credits earned by credits attempted. For students who are tracked in at least one subject, the control group passes 64% of their college-level credits attempted while the treatment group pass rate is 63%. For those students tracked in both subjects, the control group pass rate is 62% and the treatment group pass rate is 63%.[12]

Table 7 shows these results hold over the longer run as well. It is possible that, by the end of two years, students in the control catch up to students in the treatment group. We can track our initial cohort of students from Fall 2016 for more than two years. The control group for this cohort has higher total credit accumulation than the overall sample, as expected, but the increase in total credits earned and the decrease in remedial education credits earned are each larger than what is observed in the short run. Thus, if

[12] Similarly, for those tracked in at least one subject, the complier (IV) pass rates are: 62% for the control group, and 61% for the treatment group. For those tracked in both subjects, the complier (IV) pass rates are: 62% for the control group, and 63% for the treatment group.

anything, the benefits appear to grow as students progress through community college.

Subgroup Effects

Table 8 shows the subgroup effects on placement outcomes and credit accumulation for each subject and subgroup. Each cell is a separate regression restricted to the subgroup specified in the column header. The observation count for the outcome and subgroup of interest is shown immediately below the standard error.

Treatment effects on placement into college-level math and English are large and positive for all subgroups, except male students in math. Remedial credits earned also decrease for all subgroups (including male students), and credits increase a statistically significant amount only for Black and female students and Pell-grant recipients.

College administrators were particularly interested in how the new placement system affected gaps in outcomes across subgroups. Given these administrators are making the decision to maintain the system, we focus our analysis of heterogeneous effects on the extent to which algorithmic placement widened or narrowed gaps in key outcomes across subgroups. This question implies we are interested in the interaction terms from equation [9], which assess whether there are differential effects for Black and Hispanic students (separately) relative to white students, female students relative to male students, and Pell recipients relative to non-Pell recipients. Including outcomes in placement for math and English and credit accumulation in remedial and college-level courses, there are 16 interaction terms of interest. We use the step-down method from Holm (1979) to control for the Family-wise Error Rate at the five percent level.

Four interaction terms remain significant after this adjustment. Placement rates for Black students into college-level English increased relative to white students and placement rates for female students into college-level math relative to male students increased as well. Though placement rates into college-level math and English courses increased for Hispanic students overall, relative to white students, the increase in math is smaller than it is for white students. Lastly, the decrease in remedial credits is significantly smaller for Pell recipients than it is for non-Pell recipients. Thus, though all students seem to benefit from algorithmic placement, there is evidence that most (though not all) of the benefits accrue to students traditionally under-represented in college courses.

5. Cost Analysis

In this section, we present the cost-effectiveness analysis for the algorithmic placement system and business-as-usual placement systems for six colleges using the ingredients method (Levin et al., 2017); we could not collect complete cost data at one college.[13] The cost estimates reflect the annual expected cost during the first five years of implementing and operating the new placement system at college of similar size and organization as the six sample colleges.

The new placement method resulted in cost savings for students: students earned more college credits and took fewer remedial credits with a net effect of lower tuition payments. Relative to the business as usual, implementation and operation costs were larger for colleges, $140 per student; operating costs, however, are $40 per student over the status quo. Overall, algorithmic placement is more cost-effective from a social perspective than the existing placement systems. That is, while the implementation and operating costs are larger for colleges, the cost reduction for students more than offsets the increased cost to colleges, so total costs are lower for the algorithmic placement system. Moreover, costs could be reduced substantially if data to estimate the algorithm did not have to be hand

[13] What we could collect does not suggest this seventh college had costs significantly different from the others; personnel changes prevented us from collecting all the necessary data.

entered and if data collection were centralized into a single system. We detail the calculations underpinning these findings below.

Defining Costs and Cost Data

To better understand the details of our cost-effectiveness analysis, we start by defining several terms. First, *fixed costs* are those costs that do not vary with college enrollment. *Direct costs* are the costs of implementing and operating the placement system. *Implementation costs* include one-time costs incurred to develop and test the placement method (e.g., evaluator) and the operating costs to keep it fully functional. *Operating costs* refer to running a placement system after the initial method has been developed and tested (i.e., personnel, facilities, administering placement test, etc.). *Indirect costs* are associated with the price and quantity of credits attempted by the students. The *total costs* are the sum of the *indirect* and *direct costs*. *Student costs* include only the cost of the credits attempted and not the *direct costs*, as they do not pay for the additional costs of implementing the alternative method. In contrast, *college costs* include *direct costs* of implementing the alternative system and any costs from course offerings (e.g., changes in the number of remedial courses offered). Finally, *cost-efficiency*, in our context, compares the costs of the algorithmic placement system to the business-as-usual placement system (Levin et al., 2017).

We collected data on ingredients from two primary sources. One source for this information was from direct interviews with faculty and staff who implemented the new testing protocols. The second source for input prices and overhead costs was from secondary sources, such as the Integrated Postsecondary Education Data System (IPEDS), described below.

Sources of Costs in the Placement Systems

Understanding the different cost components of the placement systems helps to distinguish fixed costs from operating costs. The initial investment to implement the algorithm has three components. First, data on students' characteristics (including high school transcripts), placements based on test results, and subsequent college outcomes must be collected. In some colleges, these data are already available, but other colleges required more extensive data collection. Second, data must be analyzed to estimate the new placement algorithm. Third, resources must be allocated to create and implement the new system within the college, which includes training personnel. After the initial investment, implementation requires collecting data from entering students and personnel to assign students to either remedial or college-level courses. For the algorithm, one driver of costs was data entry. Data entry costs were lower if the college had all high school information pre-loaded into their databases; in contrast, data entry costs were higher if each student's information had to be entered into the computing system individually.

For both placement systems there are costs for administering placement tests. Also, for both systems, future resources may be required as students progress into college-level courses after completing remedial coursework. If more students progress into college-level courses, colleges may have to shift resources toward college courses and away from remedial courses in conjunction with any changes in revenue per student.

College faculty, counselors, and administrators did not indicate significant resource changes with respect to instruction. Potentially, the new placement system may change assignments such that more students are now in college-level classes, which would require more college-level faculty and more sections of college-level courses. However, colleges indicated that faculty could be reassigned from teaching remedial classes to teaching college-level classes, and few changes in class size were anticipated even given the changes in placement rates.

Along with the direct implementation and operational costs, there were also indirect costs associated to the different total number of credits attempted by students under the algorithmic placement system. To compute the indirect costs, we used IPEDS information on the six colleges considered in this analysis. The overall cost per college-level credit and remedial credit was approximately $520 (Barnett et al., 2020).

Cost Estimates

Indirect costs: Table A.4 shows the college-level and remedial credits earned and attempted (ITT and TOT). Using our estimates of costs per credit, the indirect costs for the business-as-usual are $5,420 per student ($7,440 for compliers) compared to $5,040 per student ($6,650 for compliers) for the algorithmic placement system. The lower costs of the latter stem from the net decrease of 0.74 (1.5 for compliers) in total credits attempted. Thus, the implementation of the algorithmic placement system results in an indirect cost reduction of $380 per student ($790 for compliers).

Student costs: Students do not pay all the costs associated with each credit attempted. The relevant costs for students are tuition and fees paid for these credits. Using IPEDS data for the six colleges, the cohort-weighted average for tuition and fees is 39% of total expenditures per credit (Barnett et al., 2020). Therefore, of the $520 cost per credit, students pay $200 and public funding covers the remaining $320. Consequently, as shown in Table A.5, students attempted fewer credits in total with the new placement method relative to business-as-usual and therefore saved $150 ($310 for compliers).

Direct costs: Table A.6 shows the direct costs to implement and operate the algorithmic placement system and the business-as-usual placement system for five years (amortized over cohorts). For a typical college cohort in the sample of 5,808 students, the cost of implementing the algorithmic placement system is $958,810. The cost of the placement exam system (business as usual) is $174,240. These estimates imply an incremental cost per student of $140 for algorithmic placement. The remaining two columns show upper and lower bounds for this cost per student, which ranges from $70 to $360. This variation is driven by substantial fixed costs, so colleges with larger enrollments show much smaller per student costs. One implication of these findings is that costs could be reduced substantially with more efficient, centralized data collection. Minimizing hand data entry and centralizing high school student information into a single data system would help automate algorithm estimation and reduce costs.

Total costs: We summarize the total costs—direct and indirect for both students and colleges— for each placement system in Table A.7. The total cost per student is $240 *less* ($650 for compliers) for the algorithmic placement system compared to the business-as-usual placement system. This result is a consequence of the lower indirect costs due to fewer total credits attempted under the algorithmic placement system, which more than offsets the higher direct cost relative to business-as-usual (see Table A.6).

The lower total costs of the algorithmic placement system suggest it is cost-effective from a social perspective relative to business-as-usual system: algorithmic placement is more effective regarding the number of college-level credits earned and its total cost is lower. As shown in Table A.7, the cost-per college credit earned is $100 ($220 for compliers) less for the new placement method relative to business-as-usual.

Finally, cost effectiveness from the colleges' perspective is harder to establish. On the one hand, colleges must incur the higher costs to implement and operate the new placement method (as shown in Table A.6). On the other hand, we do not incorporate potential increases in net revenues from the additional coursework. These revenue changes will depend on the characteristics of each institution (e.g., enrollment numbers, funding strategy, etc.), which makes it more difficult to determine these changes relative to the status quo. However, as the algorithmic placement method's total cost is lower and leads to greater credit accumulation, we believe this system is likely cost-effective from each colleges' perspective relative to the business-as-usual system as well.

Lastly, colleges could also save money by not purchasing the ACCUPLACER exams, and they asked whether students could be placed via the algorithm as accurately without using these test scores. We examined the extent to which the algorithm would place students differently if test scores were not used for prediction. We find that placement rates would change substantially for math courses—by 18%—however, for English courses, only 5% to 8% of placements would change. This finding is in line with the increased predictive value we find for math test scores over English test scores.

6. Conclusion

Our findings indicate that combining predictive analytics with multiple measures significantly impacted how colleges track students into either college-level or remedial courses. First, the placement algorithm allowed colleges to choose cut points that explicitly targeted predicted placement rates and pass rates. Second, the algorithm led to changes in the placement of students. Across the seven study colleges, more students were placed into college-level math and English courses—without reducing pass rates in either course. There were particularly large increases in college-level placements in English courses.

While the algorithm’s predictive validity was greater than placement scores alone, the algorithms we developed could be improved. Most notably, our model was constrained by implementation in several ways. To produce rapid placement decisions, we had to embed our algorithm into existing systems, which restricted our modeling choices. We could not for instance, implement a non-linear model. Future models could also use richer transcript data; the colleges we worked with could not readily provide course-level high school grades that could be predictive of future performance as well. More generally, as colleges develop more consistent ways to record incoming student information, the ability to predict future performance could improve.

One question is how our results would differ if all students within a college were placed according to the algorithm. Our interviews with college administrators, department chairs, faculty and counselors at each college documented their impressions to the algorithm's implementation. Generally, there was no perceived change in classroom composition. However, this could change if all students were placed via the algorithm, especially in English courses where placement changes were more significant. Prior research suggests this could result in improved academic outcomes for students (Duflo, Dupas, and Kremer, 2011).

Our initial results have important implications because the high cost of remedial education falls onto students placed into these courses and indirectly onto taxpayers whose money helps subsidize public postsecondary institutions. As a result, there is both a private and social benefit to ensuring that remedial education is correctly targeted. Colleges recognize this, and some have begun to implement these placement algorithms. Long Beach City College (LBCC) created a placement formula that uses student high school achievement in addition to standardized assessment scores. The formula weights each measure based on how predictive it is of student performance in college courses (Long Beach City College, Office of Institutional Effectiveness, 2013). This paper provides evidence that these placement systems not only affect student outcomes through changes in the placement instrument, but also through colleges' improved ability to target pass rates explicitly. Future research could test more intricate predictive models than we could implement in the current study, and perhaps focus on algorithms that predict treatment effects of each course rather than pass rates.

REFERENCES

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2020). Sampling-Based versus Design-Based Uncertainty in Regression Analysis. *Econometrica*, *88*(1), 265-296.

Abdulkadiroğlu, A., Angrist, J., & Pathak, P. (2014). The elite illusion: Achievement effects at Boston and New York exam schools. *Econometrica*, *82*(1), 137-196.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagwa (Eds.) *Selected Papers of Hirotugu Akaike.* Springer Series in Statistics (Perspectives in Statistics). New York, NY: Springer.

Allen, D., & Dadgar, M. (2012). Does dual enrollment increase students' success in college? Evidence from a quasi-experimental analysis of dual enrollment in New York City. New Directions for Higher Education, 2012, 11–19. doi:10.1002/he.20010

Arnold, D., Dobbie, W., & Hull, P. (2021, May). Measuring racial discrimination in algorithms. In *AEA Papers and Proceedings* (Vol. 111, pp. 49-54).

Bacon, D. R., & Bean, B. (2006). GPA in research studies: An invaluable but neglected opportunity. Journal of Marketing Education, 28, 35–42.

Barnett, E. A., Kopko, E., Cullinan, D., & Belfield, C. R. (2020). Who Should Take College-Level Courses? Impact Findings from an Evaluation of a Multiple Measures Assessment Strategy. *Center for the Analysis of Postsecondary Readiness*.

Barnett, E., & Reddy, V. T. (2017). College Placement Strategies: Evolving Considerations and Practices (A CAPR Working Paper).

Belfield, C., & Crosta, P. M. (2012). Predicting success in college: The importance of placement tests and high school transcripts (CCRC Working Paper No. 42). New York, NY: Community College Research Center. Retrieved from: http://ccrc.tc.columbia.edu/publications/predicting-success-placement-tests-transcripts.html

Bettinger, E. P., & Long, B. T. (2009). Addressing the needs of underprepared students in higher education: Does college remediation work? *Journal of Human Resources*, 44(3), 736–771.

Boatman, A., & Long, B. T. (2010). Does remediation work for all students? How the effects of postsecondary remedial and remedial courses vary by level of academic preparation (An NCPR Working Paper). New York, NY: National Center for Postsecondary Research.

Borghans, L., Golsteyn, B. H. H., Heckman, J. J., & Humphries, J. E. (2016). What grades and achievement tests measure. *Proceedings of the National Academy of Science, 113*(47), 13354-13359.

Bui, S. A., Craig, S. G., and Imberman, S. A. (2014). Is gifted education a bright idea? Assessing the impact of gifted and talented programs on students. *American Economic Journal: Economic Policy*, 6(3):30 – 62.

Burnham, K. P. & Anderson, D. R., (2002). *Model selection and multimodel inference: A practical information-theoretic approach*, *second edition.* New York, NY: Springer.

Calcagno, J. C., & Long, B. T. (2008). The impact of postsecondary remediation using a regression discontinuity approach: Addressing endogenous sorting and noncompliance (NBER Working Paper No. 14194). Cambridge, MA: National Bureau of Economic Research. http://www.nber.org/papers/w14194

Card, D. and Giuliano, L. (2014). Does gifted education work? For which students? Technical Report 20453, National Bureau of Economic Research.

Card, D. and Giuliano, L. (2016). Can tracking raise the test scores of high-ability minority students? *The American Economic Review*, 106(10), 2783–2816.

Chan, E. W. K. (2020). *Heterogenous Parental Responses to Education Quality. Mimeo*, Babson College.

Chen, X. (2016). *Remedial Coursetaking at U.S. Public 2- and 4-year Institutions: Scope, Experiences, and Outcomes* (NCES 2016-405). U.S. Department of Education. Washington, DC: National Center for Education Statistic s. Retrieved from: https://nces.ed.gov/pubs2016/2016405.pdf

Cohodes, S. R. (2020). The long-run impacts of specialized programming for high-achieving students. *American Economic Journal: Economic Policy*, *12*(1), 127-66.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review* 101(5), 1739–74.

Dobbie, W., & Fryer Jr, R. G. (2014). The impact of attending a school with high-achieving peers: Evidence from the New York City exam schools. *American Economic Journal: Applied Economics*, *6*(3), 58-75.

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor.* New York, NY. St. Martin's Press.

Fields, R. & Parsad, B. (2012). *Tests and Cut Scores Used for Student Placement in Postsecondary Education: Fall 2011.* Washington, DC: National Assessment Governing Board.

Heubert, J. P., & Hauser, R. M. (Eds.). (1999). *High-stakes: Testing for Tracking, Promotion, and Graduation*. Washington, DC: National Academy Press.

Hodara, M. (2012). Language Minority Students at Community College: How Do Developmental Education and English as a Second Language Affect Their Educational Outcomes? (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (Accession Order No. [3505981]).

Holm, S. (1979). A Simple Sequentially Rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65-70.

Huber, P. J. (1967, June). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, No. 1, pp. 221-233).

Kautz, T., Heckman, J. J., Diris, R., ter Weel, B., and Borghans, L. (2014), *Fostering and Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success*, OECD, Paris.

Kautz, T. D., & Zanoni, W. (2014). Measuring and fostering noncognitive skills in adolescence: Evidence from Chicago Public Schools and the OneGoal Program. Unpublished manuscript, Department of Economics, University of Chicago, Chicago, IL.

Kirabo Jackson, C. (2010). Do students benefit from attending better schools? Evidence from rule-based student assignments in Trinidad and Tobago. *The Economic Journal*, *120*(549), 1399-1429.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *The Quarterly Journal of Economics*, 133(1), 237-293.

Levin, H.M., McEwan, P.J., Belfield, C.R., Bowden, A.B., & Shand. R. (2017). *Economic evaluation of education: Cost-effectiveness analysis and benefit-cost analysis.* Thousand Oaks, CA: SAGE Publications.

Li, D., Raymond, L., & Bergman, P. (2020). *Hiring as Exploration* (No. w27736). National Bureau of Economic Research.

Long Beach City College, Office of Institutional Effectiveness. (2013). Preliminary overview of the effects of the fall 2012 Promise Pathways on key educational milestones. Long Beach, CA: Office of Institutional Effectiveness.

Martorell, P., & McFarlin, I., Jr. (2011). Help or hindrance? The effects of college remediation on academic and labor market outcomes. *The Review of Economics and Statistics*, *93*(2), 436–454.

Mazerolle, M. J. (2004). Appendix 1: Making sense out of Akaike's Information Criterion (AIC): its use and interpretation in model selection and inference from ecological data. Retrieved from http://theses.ulaval.ca/archimede/fichiers/21842/apa.html

Mullainathan, S. and Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87–106.

National Center for Public Policy and Higher Education and the Southern Regional Education Board. (2010). Beyond the Rhetoric: Improving College Readiness Through Coherent State Policy. Retrieved from: http://www.highereducation.org/reports/college_readiness/CollegeReadiness.pdf

Oakes, J. (1985). Collaborative Inquiry: A Congenial Paradigm in a Cantankerous World.

Pop-Eleches, C., & Urquiola, M. (2013). Going to a better school: Effects and behavioral responses. *American Economic Review*, *103*(4), 1289-1324.

Rodríguez, O., Bowden, A.B., Belfield, C.R., & Scott-Clayton, J. (2014) Remedial placement testing in community colleges: What resources are required, and what does it cost? (CCRC Working Paper No. 73). New York, NY: Community College Research Center. Retrieved from: https://ccrc.tc.columbia.edu/media/k2/attachments/remedial-placement-testing-resources.pdf

Rodríguez, O., Bowden, Belfield, C., & Scott-Clayton, J. (2015). Calculating the costs of remedial placement testing (CCRC Analytics). New York, NY: Community College Research Center.

Rothstein, J. M. (2004) College performance predictions and the SAT. *Journal of Econometrics,* 121(1-2), 2917-317.

Scott-Clayton, J. (2012). Do high-stakes placement exams predict college success? (CCRC Working Paper No. 41). New York, NY: Community College Research Center. Retrieved from http://ccrc.tc.columbia.edu/media/k2/attachments/high-stakes-predict-success.pdf

Scott-Clayton, J., & Rodriguez, O. (2015). Development, discouragement, or diversion? New evidence on the effects of college remediation policy. *Education Finance and Policy*, 10(1), 4–45.

Scott-Clayton, J., Crosta, P. M., & Belfield, C. R. (2014). Improving the targeting of treatment: Evidence from college remediation. *Educational Evaluation and Policy Analysis*, 36(3), 371–393.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 44-47.

Strong American Schools. (2008). Diploma to nowhere. Retrieved from: http://paworldclassmath.webs.com/8534051-Diploma-To-Nowhere-Strong-American-Schools-2008.pdf.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817-838.

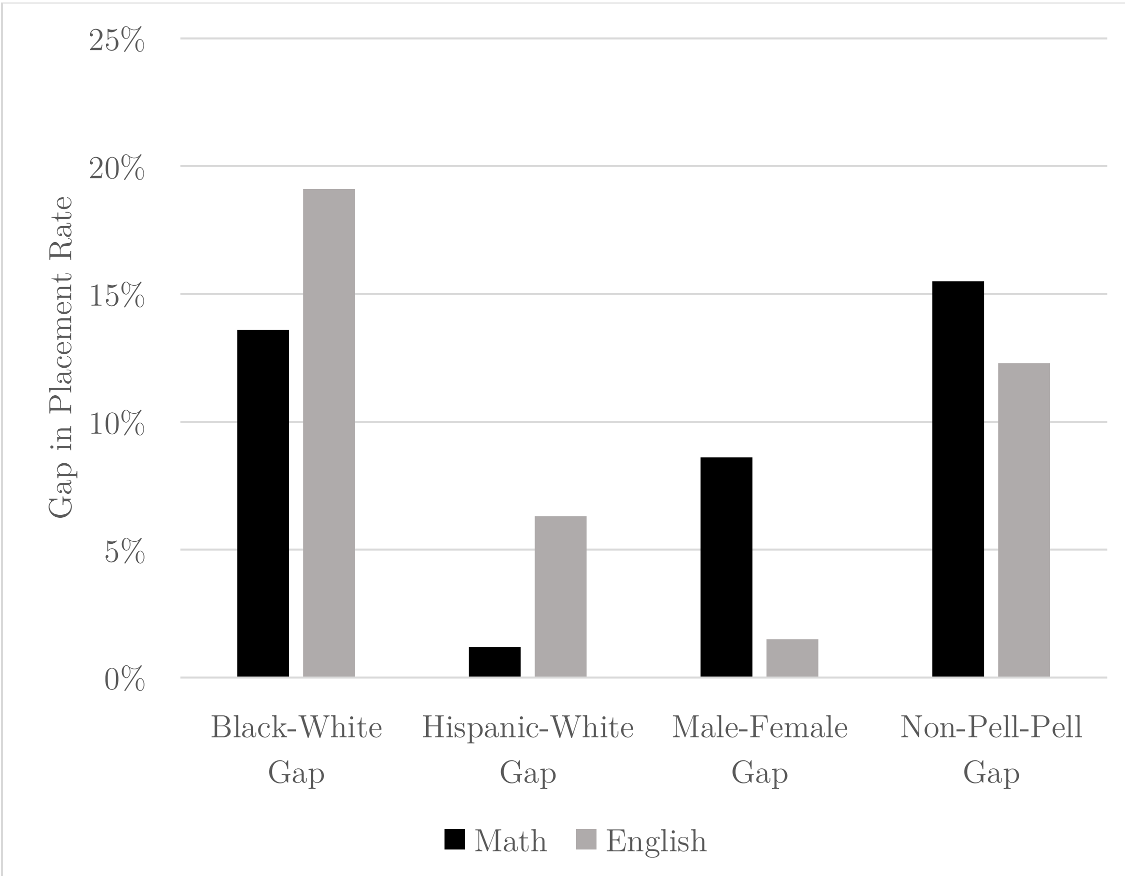
FIGURES

Figure 1. Hypothetical spreadsheet provided to colleges on placement projections

Example Community College		
Math Success: C or above		
Minimum Probability of Success	Percent Placed into College Level	Percent Passing College Level
Cohort 3, Status Quo	30%	50%
45%	40%	60%
55%	20%	70%
65%	10%	75%
Eng. Success: C or above		
Minimum Probability of Success	Percent Placed into College Level	Percent Passing College Level
Cohort 3, Status Quo	40%	60%
45%	75%	60%
55%	60%	65%
65%	20%	70%

Notes: This figure is a hypothetical version of the information presented to college faculty and administrators to help them choose a threshold for being placed into college-level course in math or English. The placement algorithm outputs a probability of success in the college-level math and/or English course for each student. Colleges then choose what probability is the "minimum probability" acceptable for placement into the college-level course. Several possible minimum probabilities are shown in the leftmost column. The middle column and the rightmost then show the predicted percent of students placed into the college-level course and the predicted pass rate for those students, respectively, associated with the minimum probability shown in the same row.

Figure 2. Gaps in Placement Rates Across Demographic Groups



Notes: Sample includes any student who took a placement exam in at least one subject and first enrolled at one of the seven study colleges in the fall of 2016. Gap in placement rate is the difference in placement rates into college-level math (shown in black) and English (shown in gray).

Table 1. Sample Demographics by College

	Overall	College 1	College 2	College 3	College 4	College 5	College 6	College 7
Female	50%	58%	54%	53%	48%	51%	55%	46%
Race								
White	43%	81%	69%	56%	53%	36%	41%	24%
Asian	2%	1%	1%	1%	2%	5%	9%	2%
Black	20%	9%	17%	20%	23%	21%	31%	19%
Hispanic	20%	5%	3%	4%	11%	28%	14%	33%
Native American	1%	1%	1%	1%	1%	0%	1%	1%
Two or more races	3%	1%	3%	4%	6%	3%	3%	3%
Age at entry	20.93	20.82	22.91	22.04	20.23	21.51	23.02	19.92
Pell Grant recip.	43%	52%	47%	49%	41%	32%	56%	42%
Total	12,544	672	1,228	1,818	2,003	1,756	350	4,717

Notes: Sample is any student who took a placement exam in at least one subject and first enrolled at one of the seven study colleges during the study period.

Table 2. Baseline Characteristics by Treatment Assignment

	Control Mean	Treatment Mean	Difference (T - C)	P-value	Obs.
Enrollment	86%	85%	-1%	0.26	12,544
Female	50%	50%	1%	0.39	11,901
Race					
White	44%	42%	-1%	0.13	12,544
Asian	3%	2%	0%	0.82	12,544
Black	19%	20%	2%	0.03	12,544
Hispanic	20%	20%	1%	0.43	12,544
Pacific Islander	0%	0%	0%	0.93	12,544
Native American	1%	1%	0%	0.46	12,544
Two or more races	4%	3%	0%	0.25	12,544
Race Missing	10%	10%	0%	0.57	12,544
Age at entry	20.94	20.91	-0.02	0.82	12,544
Pell Grant recip.	42%	43%	1%	0.22	12,544
TAP Grant recip.	31%	31%	0%	0.78	12,544
GED recip.	7%	7%	0%	0.98	12,544
HS GPA (100 scale)	77.96	78.12	0.16	0.34	7,869
HS GPA missing	37%	37%	0%	0.77	12,544
ACCUPLACER Exam score					
Arithmetic	33.6	34.0	0.4	0.43	10,191
Algebra	48.1	47.9	-0.2	0.75	10,191
College-level math	8.3	8.0	-0.3	0.61	3,656
Reading	58.1	58.0	-0.1	0.81	12,544
Sentence skills	34.9	34.6	-0.3	0.49	10,726
Written exam	3.9	3.9	0.0	0.69	10,979
Total	6,141	6,403			12,544

Notes: Sample is any student who took a placement exam in at least one subject at one of the study colleges during the study period. Estimates include strata fixed effects (indicators for each college). Observation counts vary for exam scores because students do not necessarily take all exams and gender and HS GPA are not available for all students.

Table 3. Changes in Placement for Program-Group Students

	(1)	(2)	(3)	(4)	(5)
	Took Placement Exam	Same Placement Under Business as Usual	Placement Changed from Business as Usual	Higher Placement than Business as Usual	Lower Placement than Business as Usual
			Math Placement		
% of sample	94.49%	74.67%	19.82%	12.17%	7.65%
N	6,050	4,781	1,269	779	490
			English Placement		
% of sample	80.43%	40.23%	40.20%	34.52%	5.68%
N	5,150	2,576	2,574	2,210	364

Notes: Sample is restricted to treatment group students; students who took a placement exam in at least one subject at one of the seven study colleges during the study period and assigned to the treatment group.

Table 4. Instrumental Variable 1st-Stage:
Complied with Algorithm's Recommendation

	(1)	(2)	(3)	(4)
	Overall Sample	Fall 2016	Spring 2017	Fall 2017
Treatment	0.483***	0.469***	0.529***	0.477***
	(0.0069)	(0.012)	(0.016)	(0.001)
Control Mean	0.140	0.174	0.104	0.125
Observations	12,544	4,688	1,914	5,942

Notes: Robust standard errors in parenthesis. Sample is any student who took a placement exam in at least one subject at one of the study colleges during the study period. Columns (2) to (4) restrict the sample to students tested in the corresponding term. All models include fixed effects for college (strata), controls for demographic indicators (race, gender and age, Pell recipient status), and calculated math and English algorithm values. Compliance is defined as following the algorithm's recommendation and following the business-as-usual in the control group is considered non-compliance.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5. Effect on Math and English College Coursework

	(1)	(2)	(3)	(4)	(5)	(6)
	Placed Math 1st-Term	Enrolled Math 1st-Term	Passed Math 1st-Term	Placed English 1st-Term	Enrolled English 1st-Term	Passed English 1st-Term
Treatment	0.0663***	0.0256***	0.0193***	0.322***	0.136***	0.0700***
	(0.00796)	(0.00812)	(0.00704)	(0.00830)	(0.00855)	(0.00864)
Control Mean	0.376	0.280	0.155	0.491	0.471	0.292
Observations	9,530	9,530	9,530	10,048	10,048	10,048

Notes: Robust standard errors shown in parenthesis. Sample is any student who took a placement exam in at least one subject at one of the study colleges during the study period. Estimates include strata fixed effects (indicators for each college). Columns (1)-(3) restricts to students who took the math exam. Columns (4)-(6) restricts to students who took the English exam. All models include fixed effects for college (strata), controls for demographic indicators (race, gender and age, Pell recipient status), and calculated math and English algorithm values.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 6. Effect on College-Course Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A. Intent to Treat Estimates					
	Remedial Credits Attempted	College Credits Attempted	College Credits Earned	Remedial Credits Attempted	College Credits Attempted	College Credits Earned
Treatment	-1.095***	1.247***	0.530*	-1.061***	1.893***	1.276***
	(0.0714)	(0.311)	(0.302)	(0.104)	(0.418)	(0.401)
Control Mean	3.537	26.19	16.80	4.120	24.61	15.36
	Panel B. Treatment on the Treated Estimates					
Placement Compliance	-2.265***	2.581***	1.097*	-1.740***	3.106***	2.093***
	(0.152)	(0.636)	(0.622)	(0.174)	(0.681)	(0.654)
Control Complier Mean	5.975	29.38	18.28	6.028	29.68	18.47
Sample	All	All	All	Placed in Math and English	Placed in Math and English	Placed in Math and English
Observations	12,544	12,544	12,544	7,034	7,034	7,034

Notes: Robust standard errors shown in parenthesis. Estimates include strata fixed effects (indicators for each college). Columns (1)-(3) is the full sample and columns (4)-(6) restricts the sample to students who were placed in both math and English. All models include fixed effects for college (strata), controls for demographic indicators (race, gender and age, Pell recipient status), and calculated math and English algorithm values. IV models of Panel B use treatment assignment to instrument for compliance with algorithm's recommendation. Compliance is defined as following the algorithm's recommendation and following business-as-usual in the control group is considered non-compliance. Credits attempted and earned are total credits attempted and earned by students.
*** p<0.01, ** p<0.05, * p<0.1

Table 7. Longer-Run Effects: Fall 2016 Cohort

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A. Intent to Treat Estimates					
	Remedial Credits Attempted	College Credits Attempted	College Credits Earned	Remedial Credits Attempted	College Credits Attempted	College Credits Earned
Treatment	-1.181***	2.503***	1.618***	-1.224***	2.692***	2.041***
	(0.129)	(0.605)	(0.598)	(0.164)	(0.706)	(0.688)
Control Mean	3.913	32.82	21.68	4.584	30.32	19.23
	Panel B. Treatment on the Treated Estimates					
Placement Compliance	-2.516***	5.334***	3.448***	-1.895***	4.168***	3.161***
	(0.282)	(1.265)	(1.259)	(0.259)	(1.083)	(1.059)
Control Complier Mean	6.688	35.22	22.17	6.347	35.67	22.41
Sample	All	All	All	Placed in Math and English	Placed in Math and English	Placed in Math and English
Observations	4,688	4,688	4,688	3,277	3,277	3,277

Notes: Robust standard errors shown in parenthesis. Estimates include strata fixed effects (indicators for each college). Columns (1)-(3) is the full sample and columns (4)-(6) restricts the sample to students who were placed in both math and English. All models include fixed effects for college (strata), controls for demographic indicators (race, gender and age, Pell recipient status), and calculated math and English algorithm values. IV models of Panel B use treatment assignment to instrument for compliance with algorithm's recommendation. Compliance is defined as following the algorithm's recommendation and following business-as-usual in the control group is considered non-compliance. Credits attempted and earned are total credits attempted and earned by students.
*** p<0.01, ** p<0.05, * p<0.1

Table 8. ITT by Subgroup on College-Course Outcomes

	White	Hispanic	Black	Male	Female	Pell	Non-Pell
Placed into College Math	0.106***	0.034**	0.057***	0.011	0.139***	0.068***	0.065***
	(0.013)	(0.015)	(0.019)	(0.012)	(0.012)	(0.012)	(0.011)
Observations	3,810	2,116	1,802	4,420	4,486	4,117	5,413
Placed into College English	0.296***	0.308***	0.386***	0.317***	0.333***	0.321***	0.323***
	(0.013)	(0.019)	(0.019)	(0.012)	(0.012)	(0.013)	(0.011)
Observations	4,085	2,081	2,046	4,894	4,543	4,313	5,735
College Credits Earned	0.000	0.712	1.131*	-0.092	1.147**	1.171**	0.037
	(0.498)	(0.688)	(0.622)	(0.446)	(0.453)	(0.459)	(0.392)
Remedial Credits Attempted	-0.692***	-0.792***	-0.734***	-0.555***	-0.834***	-0.887***	-0.500***
	(0.066)	(0.114)	(0.113)	(0.064)	(0.071)	(0.078)	(0.053)
Observations	5,389	2,485	2,471	5,959	5,942	5,386	7,158

Notes: Robust standard errors shown in parenthesis. Estimates include strata fixed effects (indicators for each college). Enrollment and pass-rate outcomes are for "ever enrolled" and "ever passed" the course indicated. Each column restricts the sample to the subgroup in the column header. Each cell is from a separate regression. All models include fixed effects for college (strata), controls for demographic indicators (race, gender and age, Pell recipient status), and calculated math and English algorithm values. Compliance is defined as following the algorithm's recommendation and following business-as-usual in the control group is considered non-compliance. Credits attempted and earned are total credits attempted and earned by students.
*** p<0.01, ** p<0.05, * p<0.1

APPENDIX

Figure A.1. Math Algorithm Components by College

	HS GPA	Years since HS Graduation	GED Status	Regents Math Score	SAT Math Score	Arithmetic Test Score	Algebra Test Score	College-Level Test Math
College 1	X	X	X			X	X	X
College 2	X	X	X	X	X	X	X	X
College 3	X	X	X			X	X	
College 4								
College 5	X	X				X	X	X
College 6								
College 7	X	X	X				X	

Notes: This table indicates what variables colleges used in their respective math algorithm. Test score variables are from ACCUPLACER placement exams. HS abbreviates high school.

Figure A.2. English Algorithm Components by College

	HS GPA	HS Rank	Years Since HS Graduation	GED Status	Reading Score	Sentence Skills Score	Other Writing Score
College 1	X		X	X	X	X	
College 2	X		X	X	X	X	X
College 3	X		X	X	X		X
College 4	X	X	X	X	X	X	X
College 5	X		X		X	X	X
College 6	X		X	X			
College 7	X		X	X	X		

Notes: This table indicates what variables colleges used in their respective math algorithm. Test score variables are from ACCUPLACER or other placement exams. HS abbreviates high school.

Table A.1. Math Algorithm Models

	Model 1	Model 2	Model 3	Model 4
HS GPA[1]	0.035***		0.028***	0.030***
	(0.002)		(0.003)	(0.002)
Missing GPA[2]	2.822***		2.270***	2.583***
	(0.195)		(0.209)	(0.210)
ACPL Algebra[3]		0.006***	0.004***	0.004***
		(0.001)	(0.001)	(0.001)
ACPL Arithmetic missing[2]		0.056	0.038	0.065
		(0.040)	(0.041)	(0.042)
ACPL Algebra missing[2]		0.634***	0.361**	0.335*
		(0.141)	(0.137)	(0.140)
ACPL college math missing[2]		-0.087	-0.088	-0.084
		(0.055)	(0.051)	(0.051)
Years since HS graduation				0.020***
				(0.004)
HS graduation year missing[2]				-0.056
				(0.068)
GED[2]				-0.192**
				(0.071)
Missing Diploma Type[2]				0.121
				(0.100)
Constant	-2.337***	0.038	-2.048***	-2.303***
	(0.192)	(0.122)	(0.217)	(0.213)
N	1,166	1,166	1,166	1,166
R^2	0.125	0.105	0.176	0.207
AIC	1,538.4	1,568.6	1,475.5	1,439.5

[1] 100-point scale
[2] Binary indicator
[3] Test score range 20-120

Notes: This table shows results from regression of the covariates listed on an indicator for getting a C or better in the college-level math course. Models 1 - 3 include different subsets of covariates, with the full model shown in Model 4.

Table A.2. English Algorithm Models

	Model 1	Model 2	Model 3	Model 4
HS GPA[1]	0.022***		0.022***	0.024***
	(0.001)		(0.001)	(0.001)
Missing GPA[2]	1.774***		1.761***	1.959***
	(0.103)		(0.103)	(0.114)
Reading[3]		0.001*	0.001*	0.001
		(0.001)	(0.001)	(0.001)
Sentence Skills[3]		0.000	0.000	0.000
		(0.001)	(0.001)	(0.001)
Written Essay[4]		0.000	-0.002	-0.001
		(0.002)	(0.002)	(0.002)
Missing Reading[2]		0.315***	0.332***	0.210**
		(0.073)	(0.074)	(0.077)
Missing Sentence Skills[2]		-0.027	-0.147*	-0.154*
		(0.077)	(0.074)	(0.074)
Missing Written Essay[2]		0.021	0.008	0.017
		(0.027)	(0.026)	(0.025)
Years since HS graduation				0.009***
				(0.001)
Missing Year of Graduation[2]				0.041
				(0.087)
GED[2]				-0.190*
				(0.083)
Missing Diploma Type[2]				0.032
				(0.094)
High School Rank Percentile				0.000
				(0.000)
Missing High School Rank[2]				-0.006
				(0.041)
Constant	-1.147***	0.478***	-1.218***	-1.301***
	(0.101)	(0.060)	(0.111)	(0.118)
N	3,786	3,786	3,786	3,786
R^2	0.072	0.006	0.078	0.095
AIC	4893.2	5161.4	4879.8	4823.8

[1] 100-point scale
[2] Binary indicator
[3] Test score range 20-120
[4] Test score range 1-8

Notes: This table shows results from regression of the covariates listed on an indicator for getting a C or better in the college-level English course. Models 1 - 3 include different subsets of covariates, with the full model shown in Model 4.

Table A.3. College Characteristics

	Institution						
	Cayuga	Jefferson	Niagara	Onondaga	Rockland	Schenectady	Westchester
GENERAL COLLEGE INFORMATION							
Student Population	7,001	5,513	7,712	23,984	10,098	8,458	22,093
Full-time Faculty	69	80	151	194	122	79	215
Part-time Faculty	170	177	0	480	409	0	2
Student/Faculty Ratio	20	18	16	23	23	23	16
% Receiving Financial Aid	92%	91%	92%	92%	56%	92%	70%
DEMOGRAPHICS							
Race/ethnicity:							
American Indian/Alaska Native	0%	1%	1%	1%	0%	1%	1%
Asian	1%	2%	1%	3%	5%	7%	4%
Black	5%	7%	11%	12%	18%	14%	21%
Hispanic/Latino	3%	11%	3%	5%	20%	6%	32%
Native Hawaiian or Other	0%	0%	0%	0%	0%	1%	0%
White	85%	73%	80%	49%	39%	67%	33%
Multi-Ethnic	2%	3%	2%	3%	2%	2%	2%
Race/Ethnicity Unknown	3%	3%	1%	27%	15%	2%	5%
Non-Resident Alien	1%	1%	0%	0%	1%	0%	1%
Gender:							
Female	60%	58%	59%	52%	54%	53%	53%
Male	40%	42%	41%	48%	46%	47%	47%
Age:							
Under 18	30%	17%	19%	24%	10%	37%	1%
18-24	44%	52%	60%	55%	63%	40%	69%
25-65	26%	31%	21%	21%	26%	23%	30%
Age Unknown	0%	0%	0%	0%	0%	0%	0%
RETENTION/GRADUATION RATES							
Retention							
Full-Time Students	56%	55%	63%	57%	68%	56%	64%
Part-Time Students	28%	30%	47%	34%	56%	50%	53%
Three-Year Graduation Rate	24%	27%	28%	20%	29%	20%	15%
Transfer Out Rate	18%	19%	18%	22%	19%	22%	18%

Notes: This table shows summary statistics for all students enrolled at the seven study colleges from historical data. Data are from the U.S. Department of Education, National Center for Education Statistics, IPEDS, Fall 2015, Institutional Characteristics.

Table A.4. Impacts on Credits Attempted and Earned. Full Sample

	ITT			TOT		
Per-student outcomes	Control	Treatment	Difference	Control	Complier	Difference
Remedial credits:						
Attempted	3.537	2.442	-1.095***	5.975	3.710	-2.265***
Earned	1.761	1.100	-0.661***	2.958	1.590	-1.368***
College credits in math/English:						
Attempted	6.890	7.248	0.358***	8.333	9.073	0.740***
Earned	3.986	4.114	0.128	4.662	4.927	0.265

All models include fixed effects for college, controls for demographic indicators including race, gender and age, Pell recipient status, and calculated math and English algorithm values.
*** Significant at 1%, ** 5%, and * 10%.

Table A.5. Changes in Total Credits Attempted and Costs for Students

	ITT	TOT
Credits attempted relative to status quo	-0.737	-1.525
Difference in credits paid by students	-$150	-$310

SOURCE: Tables A.4; authors' calculations. Cost figures rounded to nearest 10.

Table A.6. Costs for Implementation and Operation of the Algorithmic Placement System

	Total (six colleges)	Range Per College: Lower Per-student Incremental Cost Bound	Range Per College: Upper Per-student Incremental Cost Bound
Students per semester	5,808	2,750	505
Total Placement Cost:			
Algorithm	$958,810	$268,890	$196,170
Business-as-Usual	$174,240	$82,590	$15,150
New placement incremental cost:			
Per semester	$784,560	$186,300	$181,020
Per student	$140	$70	$360

Notes: 2016 dollars. Present values (discount = 3%). Rounded to $10. Ingredients information on full-time equivalents is from interviews with key personnel at six colleges. Lower and upper bounds represent highest and lowest per-student incremental costs across the six colleges. Cost data not available for one college. Costs amortized over cohorts. Student cohorts rounded to nearest 10. *Total placement cost* includes all costs to implement and administer the placement test; personnel (i.e., IT, program, senior/faculty, administrative support, and evaluator's time), fringe benefits, and overheads/facilities.

IT personnel salary data from https://www.cs.ny.gov/businesssuite/Compensation/Salary-Schedules/index.cfm?nu=PST&effdt=04/01/2015&archive=1&fullScreen.

Program personnel annual salary (step 4, grade 13) from https://www.suny.edu/media/suny/content-assets/documents/hr/UUP_2011-2017_ProfessionalSalarySchedule.pdf.

Senior/faculty midpoint MP-IV https://www.suny.edu/hr/compensation/salary/mc-salary-schedule/ https://www.cs.ny.gov/businesssuite/Compensation/Salary-Schedules/index.cfm?nu=CSA&effdt=04/01/2015&archive=1&fullScreen.

Evaluator's time estimated from timesheets. Fringe benefits uprated from ratio of fringe benefits to total salaries (IPEDS data (2013, 846 public community colleges). Overheads/facilities uprated from ratio of all other expenses to total salaries (IPEDS data (2013, 846 public community colleges). Cost to administer placement test from Rodríguez et al. (2014).

New placement incremental cost is the difference between the business-as-usual and the new method's total placement costs. More than two-thirds of the new placement incremental costs are implementation costs, and approximately 30% are operating costs ($40 per-student), which refer to running of new placement system after initial algorithm has been developed and tested.

Table A.7. Cost-Effectiveness Analysis: Social Perspective

	ITT			TOT		
Per-student Costs	Control	Treatment	Difference	Control	Complier	Difference
Direct cost: Placement	$30	$170	$140	$30	$170	$140
Indirect cost: Attempted remedial credits	$1,840	$1,270	-$570	$3,110	$1,930	-$1,180
Indirect cost: Attempted math and English college credits	$3,580	$3,770	$190	$4,330	$4,720	$390
Total Cost	***$5,450***	***$5,210***	***-$240***	***$7,470***	***$6,820***	***-$650***
Earned college credits	3.986	4.114	0.128	4.662	4.927	0.265
Cost per earned college credit	$1,370	$1,270	-$100	$1,600	$1,380	-$220

SOURCE: Tables A.4 and A.6 and authors' calculations. Cost figures rounded to nearest 10.