A Seven-College Experiment Using Algorithms to Track Students: Impacts and Implications for Equity and Fairness*

Peter Bergman†, Elizabeth Kopko, Julio Rodríguez‡

February 2023

ABSTRACT

Tracking is widespread in education. In U.S. post-secondary education alone, at least 71% of colleges use a test to track students into various courses. However, there are concerns that placement tests lack validity and unnecessarily reduce education opportunities for students from under-represented groups. While research has shown that tracking can improve 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 that uses algorithms to predict college readiness and track students into courses. Compared to the most widely-used placement tests in the country, the algorithms are more predictive of future performance. We conduct an experiment across seven colleges to evaluate the effects of algorithmic placement. Placement rates into college-level courses increase substantially without reducing pass rates. Algorithmic placement generally, though not always, narrows differences in college placement rates and remedial course taking across demographic groups. We use the experimental design and variation in placement rates to assess the disparate impact of each system. Test scores exhibit substantially more discrimination than algorithms; a significant share of test-score disparities between Hispanic or Black students and white students is explained by discrimination. We also show that the selective labels problem nearly doubles the prediction error for college English performance but has almost no impact on the prediction error for college math performance. A detailed cost analysis shows that algorithmic placement is socially efficient: it increases college credits earned while saving costs for students and the government.

---

* 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 Will Dobbie, Avi Feller, Nathan Hendren, Peter Hull, Judy Scott-Clayton, Sarah Turner, and various seminar participants for their comments.

† University of Texas at Austin and NBER.

‡ Teachers College, Columbia University.

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 test scores to track students (National Center for Public Policy and Higher Education and Southern Regional Education Board, 2010; Fields and Parsad, 2012).[1] These rates are higher in two-year colleges, which enroll nearly half of all post-secondary students but fewer than 40% of whom receive any credential (Bound, Lovenheim, and Turner, 2010; Fields and Parsad, 2012; Chen, 2016; Denning et al., 2022).[2] While research has demonstrated large potential benefits of tracking (Duflo et al., 2011; Banerjee et al., 2016; Banerji and Chavan, 2016; Card and Giuliano, 2016; Banerjee et al., 2017), *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.

The potential for inaccurate placement is a concern because there is evidence that several widely used tests lack validity and unnecessarily reduce educational opportunities for students from under-represented groups (Rothstein, 2004; Scott-Clayton et al., 2014). Given that most placement tests aim to predict students' readiness for college-level courses, using an algorithm with multiple inputs, such as high school GPA, to formulate predictions could improve validity (Scott-Clayton et al., 2014; Mullainathan and Spiess, 2017).[3] However, algorithmic screening often raises its own concerns about fairness.[4]

In this paper, we develop and evaluate placement algorithms to track students into college-level courses and implement them via an experiment across seven colleges and 12,544 college students. We recruited community colleges across New York and gathered historical data on their students to estimate models predicting students' likelihood of

[1] Outside the U.S., test scores are frequently used as the sole criterion for admission to colleges (Hastings et al., 2013; Kirkeboen et al., 2016; MacLeod et al., 2017; Riehl, 2019; Aguirre et al., 2020).

[2] Community college enrollment has also been significant driver of increases in student borrowing (Chakrabarti, Lovenheim and Morris, 2016).

[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 try to optimize placement systems.

[4] For example, this opinion piece in *The Washington Post* by the director of the American Civil Liberties Union's Racial Justice Program cites the impact of AI in different contexts, including college admissions.

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 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' test-score placement system—from the most widely-used company in the country—or the placement algorithms.

The impacts of algorithmic placement systems depend on several factors. Improving the validity 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 and Zanoni, 2014; Borghans, et al., 2016).[5] The algorithms also help 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 assigned by the algorithm and test scores will be the same. But if colleges choose to maintain pass rates, the algorithm may place more students into college-level courses, which could increase students' credit accumulation and save students money if the algorithm's predictions are sufficiently accurate.

Estimating the algorithm poses substantial challenges, however. Relying on historical data and placement records for estimation requires overcoming the "selective labels problem" (Lakkaraju et al., 2017; Kleinberg et al., 2018). Selection into college-level courses is based on observables, but the historical data can only show pass rates conditional on placement into the college-level courses. The algorithm will recommend placement into the college-level for some students who never would have been placed as such before. At best, such recommendations rely on extrapolations that could reduce their validity. Colleges also required rapid placement results and integration into their existing systems, which limited the complexity of algorithms we could implement.

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

Experimental evaluation of algorithmic placement is important for testing how well it performs in practice.

We show how colleges implemented the placement algorithm, how it affected students' placement outcomes, what impacts this had on credit accumulation and costs, and its implications for fairness and equity. We find that colleges choose cut points to hold pass rates constant. This results in large changes in placement rates: relative to the test-score placement system, 23% of math placements change and 55% of English placements change. Compared to what would have occurred using the test-score placements, the algorithm places 15% into a higher-level math class and 49% into a higher-level English course. Hence the algorithm places 8% of students into a lower-level math course and a 6% lower-level English course.

Placement via the algorithm leads to immediate increases in enrollment into college-level courses. Algorithmic placement yields 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. Algorithmic placement also leads to reductions in remedial course taking and increases in college credits earned—*without* reducing pass rates. Placement via the algorithm reduces remedial credits attempted by 1.1 credits and increased college credits earned by 0.53 credits. For students tracked in both math and English the number of college credits earned increases by 1.3 because of algorithmic placement.[6]

When the two placement mechanisms disagree on college placement, the algorithm is substantially more accurate. We observe counterfactual placements for all students. When the algorithm recommend placement into college-level math, but the test-score system does not, pass rates are ten percentage points *higher* than when the test-score system recommends placement into college-level math, but the algorithm does not. For the analogous disagreement in English, pass rates are 12 percentage points higher.

We find evidence algorithmic placement narrows certain demographic gaps in placement rates. The algorithm increases placement into college-level courses for all subgroups we looked at. After controlling for multiple-hypothesis-testing, increases are

[6] Each college has a small number of 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.

significantly larger for female students in math relative to male students and Black students in English relative to white students. Lower-income students have a larger reduction of remedial credits relative to higher-income students. However, the increases in placement into college-level math, though positive and significant, are not as large for Hispanic students compared to white students.

In terms of equity, colleges were concerned algorithmic placement would differentially affect pass rates across subgroups relative to the status quo. The algorithm does not consider any protected characteristics. However, algorithmic placement could reduce or improve pass rates for some groups and not others relative to test-score placement. Looking across multiple subgroups, we find that pass rates in college-level courses remain extremely similar to pass rates in the test-score placement system. Thus, placement rates and credit accumulation improve across subgroups without reducing pass rates.

We use variation in placement rates combined with the experimental design to assess the disparate impact of each placement system and consequences of selective labels. Conditional on a student's success in the college course, the placement tests discriminate against several under-represented groups, and more so than the algorithm scores. We further show that the selective labels problem nearly doubles the prediction error for college English performance but has almost no effect on the prediction error for college math performance.

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 test-score placement system, which is due to reductions in remedial course taking. This implies an average saving to students equal to \$145,200 per cohort, per college in our sample.

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 cost per student in the initial year of the study—above and beyond the test-score placement system—is \$140. Much of these costs are driven by the need to hand enter data from high school transcripts. Simple process enhancements, such as requesting GPAs on applications, or data-transfer relationships with high schools, could simply and greatly reduce the cost of this data collection. The first year of implementation also involves large, fixed costs. We estimate operating costs of the placement algorithm are $40 dollars per student. The implementation costs are more than offset by the savings to the government from reduced credit taking, however, yielding an infinite Marginal Value of Public Funds (Hendren and Sprung-Keyser, 2020).

Our paper contributes to research arguing that data-driven algorithms can improve human decision making and reduce biases (Mullainathan and Spiess, 2017; Li et al., 2020; Arnold et al., 2021; Arnold et al., 2022). 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.

Our paper also relates to a broader literature on tracking students. Historically, tracking is controversial. Oakes (1985) argues 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 and Kremer (2011) randomize students in Kenya to schools that either tracked students by test scores or assigned students randomly to classrooms. They find 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) study 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.

Multiple studies look at the effects of being placed into a higher track versus a lower track. Bui et al., (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.[7] In higher education, there is evidence of negative effects on course completion and graduation for marginally admitted students in high-ability classes (de

[7] 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; Abdulkadiroğlu et al., 2014; Dobbie and Fryer, 2014).

Roux and Riehl, 2022). On the other hand, 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 and Long, 2008; Bettinger and Long, 2009; Boatman and Long, 2010; Martorell and McFarlin, 2011; Allen and Dadgar, 2012; Hodara, 2012; Scott-Clayton and Rodriguez, 2015).

The rest of our 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 main findings. Section 6 investigates measures of equity, fairness, and the implications of the selective labels problem for predicting course outcomes. Section 6 provides a detailed cost analysis, and Section 7 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 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 and 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 and 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 and 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.[8] 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.1 of the Appendix provides each colleges name and an overview of their 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

[8] 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.

algorithm that did 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 cases, 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.

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 tests whether the assumptions necessary for the practical application of this estimation are sufficiently satisfied 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, who agreed to define success 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, which had the advantage of producing placement decisions immediately following placement exam completion. The software placed significant constraints on what we could implement because it could largely only incorporate basic, Boolean operators. Nonetheless, the non-linear models yielded similar placement decisions as LPMs around the relevant cut points that colleges chose to determine placements.

For each college, we estimated regressions relating placement test scores, high school GPA and other predictors to "success" in initial college-level classes for a given subject:

$$[1] \qquad \mathbf{1}[\text{C or Better}]_{ic} = \beta_0 + \beta_1(\text{HS GPA}_{ic}) + \beta_2(\text{ACCUPLACER}_{ic}) + \mathbf{X}'_{ic}\boldsymbol{\beta_3} + \boldsymbol{\varepsilon}_{ic}.$$

We added additional covariates 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 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 a value and added indicators for missing. Identical procedures were followed for both English and math. We estimated the models on prior years of historical data excluding the most recent year, and then examined fit criteria using data from that most recent year.[9]

Conditional on placement into the college-level course, exam scores explain very little variation in English course outcomes but more variation in math outcomes; including additional measures adds explanatory power.[10] Appendix Figures A.1 and A.2 list the full set of variables used by each college to calculate students' math and English algorithm scores, respectively. Tables A.2 and A.3 show typical examples of our regression results for math and English. Across colleges, explanatory power is much higher for math course grades than for English course grades. Placement scores typically explain less than 1% of the variation in passing grades for English. Test scores are better predictors for passing math grades, explaining roughly 10% of the variation. This pattern is likely due to the consistency in content between coursework across colleges in math and math tests. ELA college courses may exhibit more heterogeneity across colleges. Prior research has also shown that math exams tend to create stronger incentives for test prep (Riehl, 2019; Riehl and Welch, 2022). Adding high school grades typically explains an additional 10% of the variation in both subjects. We find that indicators for which high school a student attended, which could reflect different

[9] 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 (varying variables in $\mathbf{X}_i$). 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; Hastie et al., 2009). Under certain conditions, choosing model specifications according to the AIC is asymptotically equivalent to leave-one-out-cross-validation (Stone, 1977). 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).

[10] The unconditional explained variation differs from what we show in the appendix table. For context, pass rates average are around 55% in math and 61% in English.

grading standards, add 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:

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

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.[11] 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 in a college-level course 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

[11] In practice, we showed results from many different 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.

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.[12] These rules created the weights on various student characteristics that, when combined with the colleges' thresholds for placement, produced a placement recommendation for a student.

[12] As mentioned above, this process placed constraints on the algorithm's complexity. Interaction terms and non-linear models, for instance, were difficult to implement within the ACCUPLACER system.

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.[13] Random assignment was at the student level and stratified by college. We integrated the assignment procedure into each college's placement platform as described above, such that, upon taking their placement exams, students were randomly assigned to be placed using either the test-score placement system or the algorithmic system. 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 test-score placement system for both subjects or the algorithmic placement system 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 resulted in a well-powered study, given the 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 and Giuliano, 2016 and Duflo et al., 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

[13] 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.

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. Therefore, we observe the algorithmic *and* the test-score placement results for every student in our sample, irrespective of whether they were assigned to the treatment or control group. 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. 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.1 shows that the study sample characteristics match the overall characteristics of students each college serves.

Outcomes

We study the effects of assignment to the placement algorithm on several primary outcomes, by subject. First, we examine how placements change as a result of the algorithm: what share of treated students had their placement change relative to the status quo, and of these, what share had their placement change from a remedial-course assignment to a college-level assignment, and what share had their placement change 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 algorithmic placement versus the test-score placement system. We estimate the following model:

$$\text{[2]} \qquad \mathrm{Y}_{ic} = \alpha_0 + \alpha_1 \mathrm{Treatment}_{ic} + \varphi_c + \mathbf{X}'_{ic}\boldsymbol{\alpha_2} + \psi_{ic},$$

where Y_{ic} are academic outcomes for student i in college c, such as placement into a college-level course and passing a college-level course; $\mathrm{Treatment}_{ic}$ indicates whether the individual was randomly assigned to be placed using the algorithmic placement system or the test-score placement system; φ_c are college (strata) fixed-effects; $\mathbf{X}_{ic}$ is a vector of baseline covariates (gender, race, age, financial aid status), including students' math and English algorithm scores, which are baseline measures of academic preparedness, and ψ_{ic} is the error term. The coefficient of interest is α_1, which is the effect of assignment to the placement algorithm on outcomes discussed above. We estimate Huber-White-Heteroskedasticity robust standard errors following the experimental design (Huber, 1967; White, 1980; 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 (which cannot be affected by algorithmic assignment), and therefore are assigned to placement by the algorithm for one *or* two courses. We also estimate these regressions for those who took placement exams in both subjects and therefore can be assigned to placement by the algorithm for two subjects.

Treatment Compliance

Because not everyone follows their recommended placement, we also estimate the effect of treatment assignment on compliance with the algorithm's placement recommendation independently of being in the treatment or control group. We observe the algorithm's and the test-score placement system's recommendations for each student in our dataset, so compliance is defined as an indicator equal to one for students following the algorithm's placement recommendation and zero otherwise.[14] For students who took both math and English exams, compliance is defined as following the algorithm's recommendation in at least one subject. These results are more to inform implementation, and we do not present treatment-on-the-treated results.[15]

Subgroup Analyses

We also study several differential effects of the placement algorithm has on the composition of students placed into remedial and college-level courses. We estimate equation [2] above for each subgroup and test the significance of the interaction terms, shown below.

$$[3] \qquad \mathrm{Y}_{ic} = \beta_{0k} + \beta_{1k}\mathrm{Treatment}_{ic} + \beta_{2k}\mathrm{Treatment}_{ic} \times \mathrm{Subgroup}_{k} + \gamma_c + \mathbf{X}'_{ic}\boldsymbol{\beta}_{\boldsymbol{3k}} + \zeta_{ic}.$$

The outcomes, Y_{ic}, are placement in college-level math, placement in college-level English, and credit accumulation. For each k subgroup of interest, we restrict the sample 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 because of algorithmic placement. The subgroups of interest are Black students and 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 a simple step-down procedure formulated by Holm (1979).

Treatment-Control Baseline Balance

Randomization should ensure that, in expectation, students assigned to the treatment group are similar to those assigned to the control group. Table 2 provides evidence that random assignment was successfully implemented. Participants' demographic and academic characteristics are balanced across treatment and control groups. Students'

[14] Therefore, compliance is equal to zero for students in the control group if they follow their test-score placement recommendation, regardless of whether it is the same as the algorithmic recommendation.

[15] Treatment-on-the-treated coefficients can be computed as the estimated ITT coefficient divided by the difference in compliance between the treatment and the control group (Imbens and Angrist, 1994).

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 unsurprising given 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 test-score placement system. 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, 82% were tracked in math and 100% were tracked in English. Among those students who took a math placement exam, 23% experienced a math placement different from what would have been expected under the test-score placement system. Of those with a changed math placement, 67% were placed into a higher-level math course than would have been expected under the test-score placement system, and 33% placed in a lower-level math course. Of those who took the English placement exams, approximately 55% of program-group students experienced a change in the level of their English level placement, of which 90% placed into a higher-level English course and 10% placed into a lower-level course than they would have under the test-score placement system.

Table 4 shows compliance with algorithm's placement recommendations. Overall, the treatment group complies with their algorithmic placement recommendation 80% of the time. Treatment assignment increases compliance with the algorithm's decision relative to the control group by 76 percentage points. The first stage is slightly lower for the spring cohort, when there are fewer first-time enrollees, but is generally consistent. Students may be placed into a particular course, but they may decide to delay enrollment and enroll in courses in different subjects.

Treatment Effects on Placement, Course Taking, and Credits

Algorithmic placement causes increases in 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. As described above, one explanation for the difference between placement and enrollment into a college-level math course is that students placed into college-level math do 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 are 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 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. The first three columns in Table 6 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).

Table 6 shows algorithmic assignment reduces remedial credits attempted by 1.1 credits relative to a mean of 3.5 credits—a 31% reduction. The effect is almost the same for those tracked in both English and math, but, relative to a mean of 4.1 credits, represents a 26% reduction. Table 6 also shows there is an increase in credit accumulation of 0.53 credits for those tracked in at least one subject and 1.3 credits for those tracked in both subjects. At the outset, it was possible students might substitute math and English college credits for other college-level credits and not just for remedial credits. However, Table A.4 shows the net positive effects on credit accumulation is also due to an increase in other (non-math and non-English) college-level credits and not just a substitution between types of credits.[16]

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%.

Another indicator of performance relative to test-score based placement is what happens when the two systems disagree on placement into college-level courses. We observe counterfactual placements for each student. The first row of Table 7 shows pass rates in each subject when the algorithm recommends placement into college level but the test-score placement system does not. The second row of Table 7 shows pass rates in each subject when the test-score system recommends placement into college level, but the algorithm does not. The first two columns show pass rates in math and English during the first term, and the second two columns show pass rates for students who ever enroll in the college-level math and English course. When the systems disagree, pass rates are at least 10 percentage points higher for the algorithmic placement system than for the test-score based placement system.

The overall pattern of results holds over the longer run as well. Table 8 shows results for our earliest cohort of students 2.5 years later. It is possible that, after more than two years, students in the control catch up to students in the treatment group. 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. The evidence we have suggests that the benefits are consistent (or grow slightly) as students progress through community college.

5. Equity, Fairness, and the Implications of Selective Labels

Equity

[16] Pass rates for the non-math and non-English credits are the same (around 65%) as for math and English credits.

One measure of equity is how algorithmic placement affected differences in placement rates across subgroups. Table 9 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 control group mean and the observation count for the outcome and subgroup of interest is shown immediately below the standard error.

The effects of the algorithm 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 college credits increase a statistically significant amount only for Black students, female students, and Pell-grant recipients.

College administrators were interested in how the algorithmic placement system affected differences in placement rates across subgroups. Given this interest, we focus on the extent to which algorithmic placement widened or narrowed differences in key outcomes across subgroups. This question implies we are interested in the interaction terms from equation [3], which tests 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 (conservatively) 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 increase relative to white students and placement rates for female students into college-level math relative to male students increase as well. Though placement rates into college-level math and English courses increase for Hispanic students overall, the increase in math is smaller than it is for white students. Lastly, the decrease in remedial credits is significantly larger for Pell recipients than it is for non-Pell recipients. Thus, though all students seem to benefit from algorithmic placement, there is evidence that many of the benefits accrue to students traditionally under-represented in college courses.

One concern is whether algorithmic placement increases placement rates for certain subgroups but reduces pass rates in college courses for those groups. As mentioned, relative to the test-score placement system, college-level pass rates remain constant overall under the new placement system, but algorithmic placement could have differential effects on college-level pass rates across subgroups.

To check this, we independently estimate equation [2] for each subgroup on pass rates and present the results in separate columns of Table 10. While there seems to be a slight reduction in pass rates for white students in English and male students in math and English, the few (3 out of 21) statistically significant coefficients become non-significant (at the five percent level) after adjusting for multiple testing using the Holm's step-down method. Overall, pass rates in college-level courses are similar across subgroups.

Assessing a Measure of Fairness

While the algorithm does not include any group membership indicator (i.e., there is no disparate treatment), a placement score may discriminate by causing disparate impact. In the context of course placement, this criterion means that a score—a test score or an algorithm score—discriminates if people from two different demographic groups have different expected scores *conditional on their likelihood of passing the college-level course*. Following the notation of Arnold, Dobbie, and Hull (2021), a score results in disparate impact if:

[4]
$$\begin{aligned}\Delta = (E[S_i|Y_i^* = 0, R_i = a] - E[S_i|Y_i^* = 0, R_i = b])(1-\mu) \\ +(E[S_i|Y_i^* = 1, R_i = a] - E[S_i|Y_i^* = 1, R_i = b])\mu \neq 0.\end{aligned}$$

S_i is an individual's score, such as their test score or algorithm score. Y_i^* indicates whether a student would take and pass the class, if directly placed. R_i indicates a particular demographic group, and μ is the average pass rate across students in either group. The first term on the righthand side of the equation represents the difference in expected scores between groups a and b conditional on students who *would not* pass the course if placed. The second term on the righthand side represents the difference in expected scores between the same groups conditional on students who *would* pass the course. These two differences are weighted by the overall share of students who would not pass the course and the overall share of students who would pass the course, respectively. A Δ different from zero indicates disparate impact. For example, let S_i be the ACCUPLACER score used to place students into college-level math, let demographic group a indicate White students, and let demographic group b indicate Black students. If $\Delta > 0$, this implies Black students who are equally qualified for the college-level course score *lower* on the ACCUPLACER test than equally qualified white students.

The challenge in estimating Δ is that we only observe Y_i^*—whether a student passes the college-level course if placed—for a selected set of students. This set is selected because placement into the college-level course is conditional on sufficiently high placement scores. The experiment helps solve this selection problem because the algorithm and test-score placement systems place different sets of students into the college-level course, and the experimental design selects a random sample of students from each of these groups.

For instance, imagine the algorithm selects one set of students and the test score system selects the complement of this set. If we only saw one of these systems implemented, we would still only observe Y_i^* for those who meet the placement criterion for that system. However, since we implement both systems simultaneously and randomly assign students to each system, we observe a random sample of students placed by either system. In this example, we would see a random sample of all students placed into the college-level course because the two systems select complementary sets of students. While the latter solves the selection problem, for most colleges it is not true: there remains a large share of students whom neither system would place into the college-level course. However, one college in the study comes close to this scenario in English and nearly does so in math.

We use this college to examine disparate impact in both the ACCUPLACER tests and algorithm scores, as well as to understand the selective-labels problem for predicting college-course performance. At this college, 99% of all students eligible for placement in English (N=905) would have been placed into college-level English by at least one of the placement systems, and 93% of students eligible for placement in math (N=1,679) would have been placed into college-level math by at least one of the placement systems.[17] We use students' observed placements across the two systems to estimate Δ as follows:

$$[5] \qquad \widehat{\Delta} = \left(\frac{\sum_{n_a=1}^{N_a} S_{ia}(1-Y_{ia}^*)}{\sum_{n_a=1}^{N_a}(1-Y_{ia}^*)} - \frac{\sum_{n_b=1}^{N_b} S_{ia}(1-Y_{ib}^*)}{\sum_{n_b=1}^{N_b}(1-Y_{ib}^*)} \right) \left(1 - \frac{1}{N}\sum_{n=1}^{N} Y_i^* \right) + \left(\frac{\sum_{n_a=1}^{N_a} S_{ia} Y_{ia}^*}{\sum_{n_a=1}^{N_a} Y_{ia}^*} - \frac{\sum_{n_b=1}^{N_b} S_{ia} Y_{ib}^*}{\sum_{n_b=1}^{N_b} Y_{ib}^*} \right) \left(\frac{1}{N}\sum_{n=1}^{N} Y_i^* \right).$$

S_{ia}, S_{ib}, Y_{ia}^*, and Y_{ib}^* are a placement score (the algorithm or ACCUPLACER score) and an indicator for passing the course, respectively, conditional on student i belonging to demographic group a or b. We standardize S_i to be mean zero and standard deviation one based on the mean and standard deviation of the score calculated across all colleges. Lastly, $\frac{1}{N}\sum_{n=1}^{N} Y_i^*$ is an estimate of the pass rate if all students were placed into the college-level course.

While our design helps address the selective labels problem, we still do not observe Y_i^* for all students.[18] We impute these missing values by numerically searching for the values of Y_{ia}^* and Y_{ib}^* that generate the largest and smallest estimates of $\widehat{\Delta}$.[19] We then bootstrap these estimates 1,000 times to obtain standard errors for the upper and lower bounds of $\widehat{\Delta}$. We estimate Δ between different demographic groups for the ACCUPLACER scores and the algorithm scores in math and English. Table 11 shows the resulting estimates and standard errors.

For the ACCUPLACER math scores, we find evidence of disparate impact across multiple demographic groups. Focusing on the lower bounds, Black students who are equally likely to succeed in college math as white students score 0.37 standard

---

[17] Note that this does not imply that either 99% of students in English or 93% of students in math were placed into the respective college-level course either within or across the two systems.

[18] Y_i^* is missing <1% in English and 7% in math.

[19] We could tighten these bounds further by assuming pass rates are weakly higher for students with higher test scores or higher algorithm scores. Let p_t, p_a be the placement decisions by the test score system and algorithm system, respectively, which can take on a value of either 1 for college level or 0 for remedial. Dropping the conditioning on race, $E[Y^*] = E[Y_i^*|p_t = 0, p_a = 0]E[p_t = 0, p_a = 0] + E[Y_i^*|p_t = 1, p_a = 0]E[p_t = 1, p_a = 0] + E[Y_i^*|p_t = 0, p_a = 1]E[p_t = 0, p_a = 1] + E[Y_i^*|p_t = 1, p_a = 1]E[p_t = 1, p_a = 1]$. We do not observe the first term—pass rates for students who had low test *and* algorithm scores—however, it is reasonable to assume it is not a higher value than any of the other terms.

deviations lower on the test (77% of the raw test score difference).[20] Hispanic students score 0.27 standard deviations lower compared to equally qualified white students (84% of the raw test score difference). Similarly, female students score 0.30 standard deviations lower than male students (88% of the test score difference). All of these results are statistically significant at the 1% level.

The math algorithm exhibits smaller or no bias. For Black students compared to white students, the upper and lower bounds of $\widehat{\Delta}$ are 0.07 and 0.05, respectively (both significant at the 1% level). For Hispanic students compared to white students, the upper and lower bounds are 0.01 and 0.00, neither of which is statistically significant. Similarly, for female versus male students, the upper and lower bounds of $\widehat{\Delta}$ are 0.02 and 0.00, neither of which is statistically significant.

For English ACCUPLACER scores, the results are more variable. For Hispanic and Black students compared to white students, $\widehat{\Delta}$ is 0.19 and 0.28 standard deviations (95% and 97% of the raw test score differences), respectively. Female students, however, score 0.13 standard deviations *higher* than male students (118% of the raw difference as female students score higher on the exam than male students).

The algorithm for English exhibits similar directional bias for female versus male students; $\widehat{\Delta}$ is 0.05 standard deviations in favor of female students (significant at the 10% level). Similarly, $\widehat{\Delta}$ for Hispanic students compared to white students is 0.10 standard deviations in favor of the former (significant at the 1% level). For Black students compared to white students, $\widehat{\Delta}$ is 0.05 standard deviations in favor of Black students (not statistically significant).

Though the results above assess a measure of discrimination across various placement scores, the extent to which they exhibit any bias is still within the context of the college's broader placement system and its associated rules and practices. For instance, colleges may change the set of test takes by providing exemptions from taking

[20] For context, Black and Hispanic students score 0.29 and 0.20 standard deviations lower than white students on ACCUPLACER English exams. Female students score 0.11 standard deviations higher than male students. On the ACCUPLACER math exams, Black, Hispanic, and Female students score 0.48, 0.32, 0.34 standard deviations lower, respectively. Differences in algorithm scores are much smaller across groups: Black and Hispanic students score 0.01 and 0.07 standard deviations higher than white students in English and 0.08 and 0.01 standard deviations *lower* in math. Female students score 0.05 standard deviations higher than male students in English and 0.01 standard deviations lower in math.

the placement tests due to prior credits earned in high school, high SAT scores, or high New York Regents test scores. Colleges may offer formal or informal test preparation guidance to students as well. While these policies tend to be similar across the colleges in our sample, their similarity (or not) to colleges outside the context of our study could limit the external validity of our results.

The Implications of Selective Labels for Predicting Course Outcomes

We also use this college to examine how the selective labels problem affects our ability to predict success in the college-level courses. To do so, we regress an indicator for passing the college-level course on baseline covariates. The college's placements across both systems, combined with random assignment, allow us to add an important interaction term: an indicator for whether a student would have been placed into college-level by the test-score system. The latter group of students—those placed into the college-level course—is the group typically used for estimating a model predicting course outcomes (cf. Scott-Clayton, 2012 and Scott-Clayton, Crosta, & Belfield, 2014). Without the experiment, these students would be the only ones for whom we observe the relevant outcome because they were placed into the college level course.

Specifically, we estimate the following model that compares coefficients from the "selected model," which conditions on placement into the college level by test-score system, with those who are not placed into college level:

[6] $$\mathrm{Y}_i = \gamma_0 + \mathbf{X}_i'\boldsymbol{\gamma}_\mathbf{1} + \mathrm{selected}_i\gamma_2 + \mathrm{selected}_i * \mathbf{X}_i'\boldsymbol{\theta} + \omega_i,$$

Y_i is an indicator for passing the college-level course, $\mathbf{X}_i$ is a vector of baseline covariates, such as high school GPA and diploma type, and $selected_i$ is an indicator for whether the test-score placement system would have placed student i into the college-level course. The coefficients of interest are represented by the vector $\boldsymbol{\theta}$, which show whether predictors are weighted differently among those placed into college level by the test score system versus those not placed into college level. One test of whether the selective labels problem "matters" is reflected in the significance of these interaction terms. Significant differences could imply that the selected model, which conditions on placement into the college level, will make different predictions than an unselected model that does not condition on placement. Table 12 shows the resulting coefficients and standard errors from estimating [6] for predicting success in college-level English and math.

Several interaction terms are statistically significant. For college English, ACCUPLACER scores and high school GPA have smaller coefficients in the selected model. We fit a Seemingly Unrelated Regression model to conduct a joint test of whether the coefficients estimated using the unselected sample—not conditioned on placement—are equal to the coefficients estimated using the selected sample. We reject this hypothesis at the 1% level. For college math, GED status is significantly larger in the selected sample. A joint test for the equality of coefficient estimates across the two models for math rejects this hypothesis at the 10% level.

We also use these model estimates to generate predictions of success in college English and math for each student. For English, the predicted probabilities generated from the unselected model are only weakly correlated with those generated from the selected model; the correlation is 0.22. The correlation of the predicted probabilities for math are much higher, however: 0.94.

What implications do these differences have for prediction errors? To answer this question, we randomly sort the data and partition it into five “folds.” We then fit each model on four of the folds and compute the Mean Squared Error (MSE) on the remaining fold (the “hold-out” or “test” sample). We repeat this process by holding out each of the five folds and then computing the MSE each time on the test sample. We average the MSE across the five folds to compute the out-of-sample Mean Squared Error (OOSMSE) for the selected and unselected models in each subject.

In English, the OOSMSE is nearly twice as high for the selected model (the model estimated on the sample that conditions on placement into the college level course via the test-score system) as the unselected model. The OOSMSEs for each model are 0.39 and 0.23, respectively. In math however, the OOSMSEs are nearly identical across the two models: 0.23 for the selected model versus 0.22 for the unselected model. Overall, the evidence suggests that the selective labels problem is more consequential for predicting college English outcomes than predicting college math outcomes.

6. Cost Analysis

In this section, we present the cost-effectiveness analysis for the algorithmic placement system and the test-score placement systems for six colleges using the ingredients method (Levin et al., 2017). We could not collect complete cost data at one college.[21] 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.

Algorithmic placement 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 test-score placement system, 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-efficient 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. Following Hendren and Sprung-Keyser (2020), we conduct a comparative welfare analysis by estimating the Marginal Value of Public Funds (MVPF): students' willingness to pay for algorithmic placement divided by the costs to the government. The latter is negative and the former positive, resulting in an infinite MVPF.

Costs could be reduced substantially if data to estimate the algorithms did not have to be hand 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 time) 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.).

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

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 students do not pay for the additional costs of implementing the algorithmic placement system. 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 test-score 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.5 shows the college-level and remedial credits earned and attempted. Using our estimates of costs per credit, the indirect costs for the test-score placement system are $5,420 per student compared to $5,040 per student for the algorithmic placement system. The lower costs of the latter stem from the net decrease of 0.74 in total credits attempted. Thus, the implementation of the algorithmic placement system results in an indirect cost reduction of $380 per student.

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 government funding covers the remaining $320. Consequently, as shown in Table A.6, students attempted fewer credits in total with the algorithmic placement system relative to the test-score placement system and therefore saved $150.

Direct costs: Table A.7 shows the direct costs to implement and operate the algorithmic placement system and the test-score 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 test-score placement system 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 the algorithm's 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.8. The total cost per student is $240 *less* for the algorithmic placement system compared to the test-score 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 (see Table A.7).

The lower total costs of the algorithmic placement system suggest it is cost-effective from a social perspective relative to the test-score placement 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.8, the cost-per college credit earned is $100 less for the algorithmic placement.

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.7). 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 test-score placement system as well.

Marginal Value of Public Funds: A conservative estimate of a student's willingness to pay for algorithmic placement is their resulting savings: $150. This assumes there is no effect on utility from avoiding remedial courses or attempting college-level courses.

Most likely there is an additional opportunity cost or disutility to unnecessary remedial course taking.[22] In the denominator, we balance the fixed and operational costs to public colleges by estimating the costs to implementing and operating the algorithmic system over five years, as described above: $140 per student. This cost is offset by the savings to the government from the reduced credit taking.[23] Total costs are negative (see table A.8) and so the MVPF is infinite.

Lastly, colleges could also save money by not purchasing the ACCUPLACER exams, and administrators 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.

7. Conclusion

Our findings indicate that algorithmic placement, which incorporates multiple measures to predict college readiness, significantly impacts how colleges track students into either college-level or remedial courses. First, algorithmic placement allows colleges to choose cut points that explicitly target predicted placement rates and pass rates. Second, the algorithm leads to changes in the placement of students. Across the seven study colleges, more students were placed into college-level math and English courses without reductions in pass rates in either course. There were particularly large increases in college-level placements in English courses. By several measures, the algorithm is more accurate, more equitable, and less discriminatory than the widely used test-score placement system.

[22] One might argue some students might dislike taking college courses as well, for whatever reason.

[23] As described above, the reduction in remedial credits reduces costs by $570 and the increase in college credits increases costs by $190 for a net cost of negative $380. Roughly 39% of this savings goes to students ($150) and the remainder is savings to the government ($230). The $140 cost to public colleges is therefore offset by the $230 savings to the government.

While the algorithm's predictive validity is 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, which 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 should 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 might result in improved academic outcomes for students (Duflo et al., 2011).

Our 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.

Aguirre, J., Matta, J., & Montoya, A. M. (2020). Joining the men's club: The returns to pursuing high-earnings male-dominated fields for women. *Unpublished Manuscript*.

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*, 11–19.

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

Arnold, D., Dobbie, W., & Hull, P. (2022). Measuring racial discrimination in bail decisions. *American Economic Review*, 112(9), 2992-3038.

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

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., ... & Walton, M. (2017). From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives*, *31*(4), 73-102.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., ... & Walton, M. (2016). *Mainstreaming an effective intervention: Evidence from randomized evaluations of "Teaching at the Right Level" in India* (No. w22746). National Bureau of Economic Research.

Banerji, R., & Chavan, M. (2016). Improving literacy and math instruction at scale in India's primary schools: The case of Pratham's Read India program. *Journal of Educational Change*, *17*(4), 453-475.

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

Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2021). Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, *50*(1), 3-44.

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.

Bound, J., Lovenheim, M. F., & Turner, S. (2010). Why have college completion rates declined? An analysis of changing student preparation and collegiate resources. *American Economic Journal: Applied Economics*, *2*(3), 129-57.

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.

Chakrabarti, R., Lovenheim, M., & Morris, K. (2016). *The Changing Role of Community-College and For-Profit-College Borrowers in the Student Loan Market* (No. 20160908). Federal Reserve Bank of New York.

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.

de Roux, N., & Riehl, E. (2022). Do college students benefit from placement into higher-achieving classes?. *Journal of Public Economics*, *210*, 104669.

Denning, J., Eide, E., Mumford, K., Patterson, R., & Warnick, M. (2022). Why Have College Completion Rates Increased? *American Economic Journal: Applied Economics, 14*(3), 1-29.

Duflo, E., Dupas, P., & Kremer, M. (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.

Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. *Advances in neural information processing systems*, *29*.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: springer.

Hastings, J. S., Neilson, C. A., & Zimmerman, S. D. (2013). *Are some degrees worth more than others? Evidence from college admission cutoffs in Chile* (No. w19241). National Bureau of Economic Research.

Hendren, N., & Sprung-Keyser, B. (2020). A United Welfare Analysis of Government Policies. *The Quarterly Journal of Economics,* 135(3): 1209-1318.

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).

Imbens, G. W., & Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, *62*(2), 467–475. https://doi.org/10.2307/2951620

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.

Kirkeboen, L. J., Leuven, E., & Mogstad, M. (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics*, *131*(3), 1057-1111.

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.

Lakkaraju, H., J. Kleinberg, J. Leskovec, J. Ludwig, and S. Mullainathan (2017). The Selective Labels Problem: Evaluating Algorithmic Predictions in the Presence of Unobservables. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 275–284.

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.

MacLeod, W. B., Riehl, E., Saavedra, J. E., & Urquiola, M. (2017). The big sort: College reputation and labor market outcomes. *American Economic Journal: Applied Economics*, *9*(3), 223-61.

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.

Riehl, E. (2019). Fairness in College Admission Exams: From Test Score Gaps to Earnings Equality.

Riehl, E., & Welch, M. (2022). Accountability, test prep incentives, and the design of math and English exams.

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

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., 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.

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.

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.

TABLES

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 Amer.	1%	1%	1%	1%	1%	0%	1%	1%
Two 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: This table shows the demographic characteristics of study participants in each college. Sample is any student who took a placement exam in at least one subject and 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	0.86	0.85	-0.01	0.27	12,544
Female	0.50	0.50	0.01	0.48	11,901
Race					
White	0.44	0.42	-0.02	0.04	12,544
Asian	0.03	0.02	0.00	0.80	12,544
Black	0.19	0.20	0.02	0.03	12,544
Hispanic	0.19	0.20	0.01	0.17	12,544
Native American	0.01	0.01	0.00	0.44	12,544
Two or more races	0.04	0.03	0.00	0.24	12,544
Age at entry	20.95	20.90	-0.05	0.63	12,544
Pell Grant recip.	0.42	0.43	0.01	0.30	12,544
TAP Grant recip.	0.31	0.31	0.00	0.90	12,544
GED recip.	0.07	0.07	0.00	0.92	12,544
HS GPA (100 scale)	77.96	78.12	0.16	0.36	7,869
HS GPA missing	0.37	0.37	0.00	0.95	12,544
ACCUPLACER exam score					
Arithmetic	33.35	34.20	0.85	0.15	10,191
Algebra	48.06	47.99	-0.07	0.89	10,191
College-level math	8.25	8.08	-0.17	0.76	3,656
Reading	58.03	58.07	0.03	0.96	12,544
Sentence skills	35.45	34.05	-1.40	0.07	10,726
Written exam	3.87	3.92	0.06	0.37	10,979
Observations	6,141	6,403			12,544

Notes: Means and Treatment − Control differences are rounded to the nearest hundredth of a point. Sample is any student who took a placement exam in at least one subject and enrolled at one of the seven study colleges during the study period. The difference column shows estimates from a regression of the baseline characteristic shown on the lefthand side on a treatment indicator and 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 Test	Same Placement as Test Score System	Placement Changed from Test Score System	Higher Placement than Test Score System	Lower Placement than Test Score System
			Math Placement		
% of sample	81.6%	58.5%	23.1%	15.4%	7.7%
N	5,226	3,747	1,479	988	491
			English Placement		
% of sample	100%	45.2%	54.8%	49.1%	5.7%
N	6,403	2,891	3,512	3,145	367

Notes: Sample is restricted to treatment group students: students who took a placement exam in at least one subject and enrolled at one of the seven study colleges during the study period and were assigned to the treatment group. Colleges exempted a share of students from placement tests based on certain criteria, shown in column (1).

Table 4. Compliance with Algorithm's Recommendation

	(1)	(2)	(3)	(4)
	Overall Sample	Fall 2016	Spring 2017	Fall 2017
Treatment	0.757***	0.813***	0.699***	0.729***
	(0.005)	(0.007)	(0.014)	(0.008)
Control Mean	0.046	0.027	0.065	0.054
Observations	12,544	4,688	1,914	5,942

Notes: Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.
Sample is any student who took a placement exam in at least one subject and enrolled at one of the seven study colleges during the study period. Columns (2)-(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), GED indicator, and calculated math and English algorithm scores. Compliance is defined as following the algorithm's placement recommendation. For students who took both math and English exams, compliance is defined as following the algorithm's recommendation in at least one subject. Following the test-score placement system's recommendation in the control group is considered non-compliance regardless of whether it is the same as the algorithmic recommendation.

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	0.066***	0.026***	0.019***	0.322***	0.136***	0.070***
	(0.008)	(0.008)	(0.007)	(0.008)	(0.009)	(0.009)
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 parentheses. *** p<0.01, ** p<0.05, * p<0.10.
Sample is any student who took a placement exam in at least one subject and enrolled at one of the seven study colleges during the study period. Columns (1)-(3) restrict to students who took the math exam. Columns (4)-(6) restrict 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), GED indicator, and calculated math and English algorithm scores.

Table 6. Effect on College-Course Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	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***
	(0.071)	(0.311)	(0.302)	(0.104)	(0.418)	(0.401)
Control Mean	3.537	26.19	16.80	4.120	24.61	15.36
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 parentheses. *** p<0.01, ** p<0.05, * p<0.10.
Columns (1)-(3) use the full sample and columns (4)-(6) restrict 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), GED indicator, and calculated math and English algorithm scores. Credits attempted and credits earned are *total* credits attempted or earned by students, respectively.

Table 7. Error Rates of the Algorithm v. Test-Score Placement Systems

		Math Pass Rate (1st-term)	English Pass Rate (1st-term)	Math Pass Rate (Ever)	English Pass Rate (Ever)
Treatment	Placed in College-Level by Algorithm, Counterfactual was Remedial	0.607	0.624	0.499	0.530
Control	Placed in College-level by Test Scores, Counterfactual was Remedial	0.442	0.496	0.401	0.408

Notes: The first row restricts the sample to students assigned to the treatment group who were placed by the algorithm into a college-level course, but their counterfactual test-score placement would have been the remedial course. The second row restricts the sample to students assigned to the control group who were placed by the test-score system (ACCUPLACER) into a college-level course, but their algorithmic placement would have been the remedial course. Each column shows pass rates by subject for the first term and ever. Pass rates are college-level credits earned divided by college-level credits attempted in the corresponding period and subject.

Table 8. Longer-Run Effects: Fall 2016 Cohort

	(1)	(2)	(3)	(4)	(5)	(6)
	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
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 parentheses. *** p<0.01, ** p<0.05, * p<0.10.
Columns (1)-(3) uses the full sample and columns (4)-(6) restrict 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), GED indicator, and calculated math and English algorithm scores. Credits attempted and earned are *total* credits attempted and earned by students, respectively.

Table 9. Subgroup Analysis 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 parentheses. *** p<0.01, ** p<0.05, * p<0.10.
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), GED indicator, and calculated math and English algorithm values. Credits attempted and earned are *total* credits attempted and earned by students, respectively.

Table 10. Subgroup Analysis on Pass Rates

	White	Hispanic	Black	Male	Female	Pell	Non-Pell
Total Credits Pass Rate (Ever)	-0.016	0.011	-0.003	-0.011	0.008	-0.007	0.004
	(0.010)	(0.014)	(0.014)	(0.009)	(0.009)	(0.009)	(0.009)
Control Mean	0.601	0.472	0.411	0.499	0.555	0.484	0.567
Observations	4,932	2,343	2,223	5,424	5,277	5,330	5,371
Math Credits Pass Rate (Ever)	-0.010	-0.009	-0.041	-0.028*	0.008	-0.005	-0.011
	(0.016)	(0.023)	(0.027)	(0.015)	(0.016)	(0.016)	(0.015)
Control Mean	0.580	0.487	0.446	0.492	0.576	0.491	0.565
Observations	2,946	1,426	1,041	3,205	2,926	2,825	3,306
English Credits Pass Rate (Ever)	-0.039***	-0.001	-0.005	-0.030**	-0.004	-0.016	-0.015
	(0.013)	(0.019)	(0.020)	(0.013)	(0.013)	(0.013)	(0.012)
Control Mean	0.667	0.532	0.466	0.560	0.618	0.542	0.631
Observations	4,162	2,021	1,804	4,553	4,435	4,450	4,538

Notes: Robust standard errors shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.
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), GED indicator, and calculated math and English algorithm values. Pass rates are calculated as total college credits earned divided total college credits attempted.

Table 11. Selective Labels: Delta Estimates

	White−Black	White−Hispanic	Male−Female
Panel A: ACCUPLACER Math			
Δ_{upper}	0.511***	0.358***	0.417***
	(0.069)	(0.068)	(0.056)
Δ_{lower}	0.375***	0.270***	0.295***
	(0.076)	(0.072)	(0.058)
Panel B: Algorithm Math			
Δ_{upper}	0.073***	0.013	0.015
	(0.020)	(0.015)	(0.013)
Δ_{lower}	0.052***	0.003	0.002
	(0.019)	(0.015)	(0.013)
Panel C: ACCUPLACER English			
Δ_{upper}	0.276***	0.185***	-0.134***
	(0.068)	(0.065)	(0.049)
Δ_{lower}	0.276***	0.185***	-0.134***
	(0.068)	(0.065)	(0.049)
Panel D: Algorithm English			
Δ_{upper}	-0.054	-0.096***	-0.055*
	(0.039)	(0.036)	(0.030)
Δ_{lower}	-0.054	-0.096***	-0.055*
	(0.039)	(0.036)	(0.030)

Notes: Bootstrap standard errors shown in parentheses (reps. = 1,000). *** p<0.01, ** p<0.05, * p<0.10. Each cell is an estimation of Δ from Equation [5] for a score (specified in the Panel) and demographic pair (specified in the Column). The estimation in each cell is restricted to students in College 5 eligible for placement in Math (Panels A and B) or English (Panels C and D). See Table 1 for descriptive statistics of College 5. The scores used to estimate Δ in each panel were standardized using the full sample. Δ_{upper} and Δ_{lower} are calculated by imputing missing values for passing the college-level course to maximize and minimize the estimated Δ for the upper and lower bounds, respectively.

Table 12. Selective Labels: Interaction terms from Eq. [6]

	(1)	(2)
	Pass College-level English	Pass College-level Math
selected × age	-0.010	0.002
	(0.009)	(0.007)
selected × $\mathbf{1}[25 < \text{age}]$	0.134	-0.172
	(0.148)	(0.128)
selected × HS GPA	-0.021***	-0.008
	(0.008)	(0.005)
selected × ACPL Algebra	-0.020	-0.015
	(0.079)	(0.053)
selected × ACPL Arithmetic	-0.002	-0.052
	(0.052)	(0.078)
selected × ACPL Reading	-0.020	0.102**
	(0.068)	(0.050)
selected × ACPL Written	-0.226***	-0.021
	(0.081)	(0.056)
selected × ACPL Sentence	-0.008**	-0.006**
	(0.003)	(0.002)
selected × GED	0.134	0.264**
	(0.148)	(0.118)
Observations	903	1,560

Notes: Robust standard errors shown in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

This table reports the interaction terms for Eq. [6] in the text for passing the college-level English course and math course in Column (1) and (2), respectively. The estimation in each column is restricted to students in College 5 eligible for placement in English in Column (1) or Math in Column (2). See Table 1 for descriptive statistics of College 5. All models include the reported interaction terms as well as each variable separably, and indicator variables for missing data and their interactions with $selected_i$.

APPENDIX FIGURES

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	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 English algorithm. Test score variables are from ACCUPLACER or other placement exams. HS abbreviates high school.

APPENDIX TABLES

Table A.1. College Characteristics

	Institution						
	Cayuga	**Jefferson**	**Niagara**	**Onondaga**	**Rockland**	**Schenectady**	**Westchester**
GENERAL 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%
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.2. 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-level 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: Robust standard errors shown in parentheses. *** p<0.01, ** p<0.05, * p<0.10.
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.3. 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)
ACPL Reading[3]		0.001*	0.001*	0.001
		(0.001)	(0.001)	(0.001)
ACPL Sentence Skills[3]		0.000	0.000	0.000
		(0.001)	(0.001)	(0.001)
ACPL Written Exam[4]		0.000	-0.002	-0.001
		(0.002)	(0.002)	(0.002)
ACPL Reading Missing[2]		0.315***	0.332***	0.210**
		(0.073)	(0.074)	(0.077)
ACPL Sentence Skills Missing[2]		-0.027	-0.147*	-0.154*
		(0.077)	(0.074)	(0.074)
ACPL Written Exam Missing[2]		0.021	0.008	0.017
		(0.027)	(0.026)	(0.025)
Years Since HS Graduation				0.009***
				(0.001)
HS Graduation Year Missing[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	4,893.2	5,161.4	4,879.8	4,823.8

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

Notes: Robust standard errors shown in parentheses. *** p<0.01, ** p<0.05, * p<0.10.
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.4. Non-English and Non-Math Credits

	(1)	(2)	(3)	(4)
	Non-English/Math Credits Attempted	Non-English/Math Credits Earned	Non-English/Math Credits Attempted	Non-English/Math Credits Earned
Treatment	0.889***	0.402*	1.369***	0.956***
	(0.248)	(0.243)	(0.328)	(0.318)
Control Mean	19.31	12.81	17.78	11.54
Sample	All	All	Placed in Math and English	Placed in Math and English
Observations	12,544	12,544	7,034	7,034

Notes: Robust standard errors shown in parentheses. *** p<0.01, ** p<0.05, * p<0.10.
Columns (1)-(2) is the full sample and columns (3)-(4) 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), GED indicator, and calculated math and English algorithm values. Credits attempted and earned are total credits attempted and earned by students.

Table A.5. Impacts on Credits Attempted and Earned. Full Sample

Per-student outcomes	Control	Treatment	Difference
Remedial credits:			
Attempted	3.537	2.442	-1.095***
Earned	1.761	1.100	-0.661***
College credits in math/English:			
Attempted	6.890	7.248	0.358***
Earned	3.986	4.114	0.128

Notes: *** p<0.01, ** p<0.05, * p<0.10.
All models include fixed effects for each college, controls for demographic indicators including race, gender and age, Pell recipient status, GED indicator, and calculated math and English algorithm values.

Table A.6. Changes in Total Credits Attempted and Costs for Students

	Treatment Effect
Credits attempted relative to status quo	-0.737
Difference in credits paid by students	-$150

SOURCE: Table A.5; authors' calculations. Cost figures rounded to nearest 10.

Table A.7. Costs for Implementation and Operation of the Algorithmic Placement System

		Range Per College	
	Total (six colleges)	Lower Per-student Incremental Cost Bound	Upper Per-student Incremental Cost Bound
Students per semester	5,808	2,750	505
Total Placement Cost:			
Algorithm	$958,810	$268,890	$196,170
Test Score Placement	$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 test-score placement system and the new, algorithmic placement system'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.8. Cost-Effectiveness Analysis: Social Perspective

Per-student Costs	Control	Treatment	Difference
Direct cost: Placement	$30	$170	$140
Indirect cost: Attempted remedial credits	$1,840	$1,270	-$570
Indirect cost: Attempted math and English college credits	$3,580	$3,770	$190
Total Cost	***$5,450***	***$5,210***	***-$240***
Earned math and English college credits	3.986	4.114	0.128
Cost per earned college credit	$1,370	$1,270	-$100

SOURCE: Tables A.5 and A.7 and authors' calculations. Cost figures rounded to nearest 10.