NBER WORKING PAPER SERIES

IMPROVING THE TARGETING OF TREATMENT: EVIDENCE FROM COLLEGE REMEDIATION

Judith Scott-Clayton
Peter M. Crosta
Clive R. Belfield

Working Paper 18457
http://www.nber.org/papers/w18457

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2012

Funding was provided by the Bill and Melinda Gates Foundation and the National Center for Postsecondary Research at Teachers College, Columbia University. The authors gratefully acknowledge: community college personnel for access to the data, research support from Olga Rodriguez, Michelle Hodara, and Emma Garcia, comments from Davis Jenkins and Tom Bailey, and editorial assistance from Betsy Yoon and Doug Slater. Funding was provided by the Bill and Melinda Gates Foundation and the National Center for Postsecondary Research, Teachers College, Columbia University. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Improving the Targeting of Treatment: Evidence from College Remediation
Judith Scott-Clayton, Peter M. Crosta, and Clive R. Belfield
NBER Working Paper No. 18457
October 2012
JEL No. H75,I23,J24

ABSTRACT

At an annual cost of roughly $7 billion nationally, remedial coursework is one of the single largest interventions intended to improve outcomes for underprepared college students. But like a costly medical treatment with non-trivial side effects, the value of remediation overall depends upon whether those most likely to benefit can be identified in advance. Our analysis uses administrative data and a rich predictive model to examine the accuracy of remedial screening tests, either instead of or in addition to using high school transcript data to determine remedial assignment. We find that roughly one in four test-takers in math and one in three test-takers in English are severely mis-assigned under current test-based policies, with mis-assignments to remediation much more common than mis-assignments to college-level coursework. We find that using high school transcript information—either instead of or in addition to test scores—could significantly reduce the prevalence of assignment errors. Further, we find that the choice of screening device has significant implications for the racial and gender composition of both remedial and college-level courses. Finally, we find that if institutions took account of students' high school performance, they could remediate substantially fewer students without lowering success rates in college-level courses.

Judith Scott-Clayton
Teachers College
Columbia University
525 W.120th Street, Box 174
New York, NY 10027
and NBER
scott-clayton@tc.columbia.edu

Clive R. Belfield
Queens College / CUNY
Economics Department
300 Powdermaker Hall
65-30 Kissena Boulevard
Flushing, NY 11367
clive.belfield@gmail.com

Peter M. Crosta
Teachers College, Columbia University
525 W. 120th St., Box 174
New York, NY 10027
peter.crosta@tc.columbia.edu

I. Introduction

Only about half of degree-seeking college entrants will complete any type of degree or certificate within six years.[1] One of the primary explanations for college non-completion is that many entrants, despite having graduated from high school, nonetheless lack the basic academic skills required for success in college coursework (Greene & Forster, 2003; Bailey et al., 2010). As a result, most two-year colleges and many four-year colleges require incoming students to be screened for possible remediation, which provides basic skills instruction but does not bear college credit, before they may enroll in college-level courses.

Besides financial aid, remedial education is perhaps the most widespread and costly single intervention aimed at improving college completion rates. Half of all undergraduates will take one or more remedial courses while enrolled; among those who take any the average is 2.6 remedial courses.[2] With over three million new students entering college each year, this implies a national cost of nearly $7 billion dollars annually.[3] This figure accounts only for the direct cost of remediation: it does not include the opportunity cost of time for students enrolled in these courses, nor does it account for any impact, positive or negative, that remediation may have on students' future outcomes.

The impacts of remediation are likely heterogeneous across individuals. Thus, like a costly medical intervention with non-negligible side effects, the net value of remediation in

---

[1] Authors' calculations based on BPS:2009 data (National Center for Education Statistics [NCES], 2012). Bachelor's degree attainment rates are 59% for those entering with a four-year degree goal, and bachelor's/associate's degree attainment rates are 30% for those entering with a two-year degree goal.

[2] Estimate based on BPS:2009 transcript data for 2003-04 entrants (NCES, 2012). Estimates based upon student self-reports are substantially lower, potentially because students do not realize the courses are remedial.

[3] This estimate is based on first-time degree-seeking fall enrollees (NCES Digest of Education Statistics, 2011, Table 207). We estimate a cost of roughly $1,620 per student per remedial course, making the assumption that each course is equivalent to a three-credit course or roughly 1/8th of a full-time year of college, and assuming the costs are comparable to the costs at public two-year colleges which have total expenditures of $12,957 per FTE per year (Delta Cost Project, 2012). With an average of 1.3 remedial courses per entrant, this implies costs of 1.3 course*$1,620 per course*3.1 million students=$6.7 billion annually.

practice depends not just on the average effectiveness of the treatment, but also on whether or not the individuals most likely to benefit can be identified in advance. Of the two-year institutions where remediation is particularly concentrated, the vast majority use brief, standardized tests administered to new students just prior to registration in order to determine who needs remediation (Parsad, Lewis, & Greene, 2003). Often, assignment is determined solely on the basis of whether a score is above or below a certain cutoff. While several studies have leveraged the somewhat arbitrary nature of these cutoffs to identify the causal effect of remediation, very little attention has been paid to the diagnostic value of the tests themselves.

This is surprising given the potentially serious adverse consequences of incorrectly assigning a truly prepared student to remediation. Prepared students who are assigned to remediation may garner little or no educational benefit, but incur additional tuition and time costs and may be discouraged from or delayed in their degree plans. Indeed, several studies using regression-discontinuity (RD) analysis to compare students just above and just below remedial test score cutoffs have generally found null to negative impacts of remediation for these "marginal" students. For example, Martorell and McFarlin (2011) examine administrative records for over 250,000 students in Texas public two- and four-year colleges: those just below the test score threshold had significantly lower rates of persistence and college credit accumulation, with no impact on degree attainment and future labor market earnings. Studies in the state of Florida and an anonymous large northeastern urban community college system using similar data and methods found similarly null to negative effects on academic outcomes (Calcagno & Long, 2008; Scott-Clayton & Rodriguez, 2012).

A typical caveat in RD studies is that they identify average treatment effects that are local to students scoring near the cutoff—that is, the highest scoring remediated students—and thus one interpretation of the RD evidence may be that the existing remedial cutoffs are set too high. The available evidence regarding heterogeneity by ability does in fact suggest that the negative effects of remediation may be largest for higher-ability or lower-academic-risk students (Martorell & McFarlin, 2011; Scott-Clayton & Rodriguez, 2012).[4]

Moreover, assigning truly unprepared students directly to college-level coursework implies a different, but no less important set of potential costs. First, there is strong evidence of peer effects in higher education, meaning that truly unprepared students who are incorrectly assigned to college-level coursework might not only do worse academically than they would have otherwise, they might depress the achievement of their better-prepared peers (Sacerdote, 2001; Zimmerman, 2003; Winston & Zimmerman, 2004; Carrell, Fullerton & West, 2009). Second, there is evidence that at least some students fare better in college if they enter remediation. Taking advantage of arbitrary variation in test cutoffs across four-year campuses in Ohio, Bettinger & Long (2009) use distance to college as an instrument for the stringency of the cutoff policy an applicant was likely to face. They find that students who were more likely to be remediated (by virtue of the cutoff policy at the nearest school) were also more likely to complete a bachelor's degree in four years. Similarly, several RD studies examining very low-scoring students at the margin between higher and lower levels of remediation have found some positive effects of being assigned to the more intensive remedial treatment (Boatman & Long, 2010, Dadgar, 2012; Hodara, 2012).

Improving the accuracy of the assignment process is thus of particular importance given the evidence for heterogeneous impacts across individuals, and given that the dominant pattern

[4] Both of these studies find some evidence that RD estimates are more negative when cutoffs fall lower in the ability distribution; Scott-Clayton and Rodriguez (2012) also use pre-existing characteristics to examine impacts for high- and low-academic risk students who all score around the same test score cutoff.

of null to negative effects suggests remediation may be overprescribed as a treatment.[5] The contribution of our study is to use a rich predictive model of college grades to simulate the prevalence of mis-assignment using common cutoff rules with the two most commonly-used remedial screening tests, to explore whether high school transcript information might be a more valuable screening device, and to examine empirically how institutions trade-off the costs of assigning either too many or too few students to remediation. We also test whether the choice of remedial screening device has disparate impacts by race or gender. Our analysis uses administrative data including high school transcripts, remedial test scores, and college grades for tens of thousands of students in two large but otherwise distinct community college systems. One is a large urban community college system (LUCCS) with six affiliated campuses; the other is a state-wide system of over 50 community colleges (SWCCS).[6]

To preview our results, we find that roughly one in four test-takers in math and one in three test-takers in English are severely mis-assigned, with severe under-placements in remediation much more common than severe over-placements in college-level coursework. Holding the remediation rate fixed, we find that using high school transcript information for remedial assignment—either instead of or in addition to test scores—could significantly reduce the prevalence of these assignment errors. Further, the choice of screening device has significant implications for the racial and gender composition of both remedial and college-level courses. Finally, we find that if institutions took account of students' high school performance, they could remediate substantially fewer students without lowering success rates in college-level courses.

---

[5] While the medical treatment analogy is useful, it is also important to note that remediation may serve other important institutional functions beyond just treating underprepared students (Scott-Clayton & Rodriguez, 2012). Beyond its "developmental" purpose, being assigned to remedial education may provide students with an early informative signal about their likelihood of college success, and/or it may serve as a means of rationing access to already-crowded college courses. We believe the accuracy of the assignment process is no less important under these alternative models of remediation.

[6] Both systems requested confidentiality in exchange for permission to freely analyze and report on the data.

The paper proceeds as follows: Section II provides background on remedial testing and summarizes the relevant research on test validity. Section III describes the methodology, including our institutional context and data. Section IV presents our results, and Section V concludes with a discussion of policy implications.

II. Background on Remedial Testing and Test Validity

At non-selective, "open-access" two- and four-year institutions, many students' first stop on campus will be to a testing center to be screened for remediation in reading/writing and math. In practice, institutional decisions about which screening tools to use and where to establish cutoffs for college-level coursework appear to be somewhat ad-hoc (Bettinger and Long 2009).[7] The affordability and efficiency of the screening tool itself are clearly important, particularly for large institutions that may need to process thousands of entrants within a matter of weeks.

Currently, two remedial placement exams dominate the market: ACCUPLACER®, developed by the College Board, is used at 62 percent of community colleges, and COMPASS®, developed by ACT, Inc., is used at 46 percent (Primary Research Group, 2008). Both testing suites offer a written essay exam as well as computer-adaptive tests in reading comprehension, writing/sentence skills, and several modules of math (of which pre-algebra and algebra are most common). The tests are not timed, but on average each test component takes less than 30 minutes to complete, such that an entire battery of placement exams may be completed in under two hours (College Board, 2007; ACT, Inc., 2006).[8] Typically, colleges waive the placement test for students with high ACT or SAT scores. Those who fail the test(s) are assigned to remedial coursework, which may stretch from one to several courses depending upon the student's score.

[7] However these decisions are made, they are increasingly made at a system- or even state-wide level (Hodara, 2012).
[8] Scores on the COMPASS® algebra exam may be determined by as few as eight questions (ACT, Inc., 2006).

Unlike the SAT and ACT exams used for college admissions, no significant test preparation market has sprung up around placement exams, perhaps because many students are not even aware of these exams and their consequences until after admission. One recent qualitative study found that students were generally uninformed about remedial assessments, with some students even believing it would be "cheating" to prepare (Venezia, Bracco, & Nodine, 2010).

A. Related Literature on Test Validity

Perhaps the simplest approach to evaluating the validity of a screening test is to identify the key outcome of interest and regress it on the predictor(s) of interest, either alone or in conjunction with other available predictors.[9] The researcher then examines goodness-of-fit statistics (R-squareds or correlation coefficients) as well as the size and significance of the resulting regression coefficients. This method has been used, for example, to examine the predictive validity of the SAT and ACT (Bowen & Bok, 1998; Bettinger, Evans, & Pope, 2011).

With respect to remedial placement exams, the College Board has published correlation coefficients relating each of the ACCUPLACER® modules to measures of success in the relevant college credit-bearing course, with correlations ranging from 0.23 to 0.29 for the math exams and from 0.10 to 0.19 in reading/writing (Mattern & Packman, 2009). In two working papers related to this study, Scott-Clayton (2012) finds comparable correlation coefficients for the COMPASS® in a large urban community college system (ranging from 0.19 to 0.35 in math and 0.06 to 0.15 in English), while Belfield and Crosta (2012) find much lower correlations for both COMPASS® and ACCUPLACER® at a state-wide system of community colleges.

[9] We recognize that a test per se cannot be validated: it is its use in a given context which is validated (Brennan, 2006). We focus here on screening devices for course placement in math and English, under the hypothesis that if the tests are not valid for placement in their own subject they are unlikely to be valid for placement in other subjects less directly related to the material on the exams.

Goodness-of-fit analyses, however, necessitate several caveats. Linearity and distributional assumptions may be violated in the case of dichotomous or ordinal outcomes. In addition, these statistics may be biased downward because of the restricted range of variation over which they must be computed (ACT, Inc., 2006).[10] More fundamentally, these measures provide no tangible estimates of how many students are correctly or incorrectly assigned under different screening devices, nor any practical guidance for policymakers wondering whether test cutoffs are set in the right place.

A second approach is to examine success rates in the college-level course for students selected on the basis of different screening devices and assignment thresholds. Bettinger, Evans, & Pope (2012) perform this type of analysis with respect to the ACT, simulating the college dropout rates that would result depending upon how ACT subtest scores are weighted in a college admissions process with a fixed number of spots. Examining test validity in a different context, Autor & Scarborough (2008) observe how the productivity of job hires (as measured by length of employment) changes when employment tests are introduced into the applicant screening process. These types of analyses are useful but focus on only one side of the assignment process. In the case of remediation, policymakers may worry not only about unprepared students being assigned to college-level work, but also about adequately prepared students being assigned to remediation. As discussed above, both types of mistakes have potentially significant costs.

[10] This range restriction occurs because the relationship between test scores and college grades can only be computed for those whose scores allow them directly into college-level courses. Statistical corrections that are sometimes employed in an effort to address this restriction-of-range may themselves rely on implausible assumptions (Rothstein, 2004). While in theory one could examine the relationship between test scores and college grades for any student who ever makes it to college coursework, for students initially assigned to remediation the treatment may confound the relationship between initial scores and future performance.

A third approach, which we develop for our primary analyses, is to analyze measures of diagnostic accuracy, or "the ability to correctly classify subjects into clinically relevant subgroups" (Zweig & Campbell, 1993). This approach has a long history in the medical screening literature and a more recent history in educational measurement, but has not been widely applied in economics or education policy research.[11] Such analyses may utilize a variety of metrics, but all aim to quantify the frequencies of accurate diagnoses, false-positive diagnoses, and false-negative diagnoses using a given test and classification threshold.[12] If decision-makers also have information on the costs and benefits of each type of event (as well as the cost of testing itself), the event frequencies can be weighted accordingly and combined into a welfare function (or loss function) that can guide the selection of the optimal screening tool and cutoff.

Sawyer (1996) is the first to apply this type of decision theory framework to the choice of remedial screening tests. He notes that no assignment rule can avoid making errors—some students who could have succeeded in the college-level course will be assigned to remediation (an under-placement error), while some students who cannot succeed at the college level will be placed there anyway (an over-placement error). Figure 1 summarizes the four potential events that result from an assignment decision by cross-tabulating potential outcomes in the college level course against actual treatment assignments.

The assignment accuracy rate, which adds the proportions of students in cells (1) and (4) of Figure 1, derives from an implied welfare function in which the decision-maker gives equal

---

[11] This could be due to a longstanding focus on identifying average treatment effects: as long as such effects are constant, then the matter of identifying whom to treat is less important. But given an increasing interest in the potential heterogeneity of treatment effects, it will become increasingly important to develop assignment tools to more accurately target interventions.

[12] For example, researchers studying the accuracy of an automated Pap smear test in the 1950s analyzed rates of false-positive and false-negative classifications for a range of possible diagnostic thresholds, then used this information to determine the optimal threshold (Lusted, 1984). The automated Pap smear test was analyzed using something similar to receiver-operating characteristic (ROC) plots, which for any given diagnostic threshold, plot what proportion of the healthy are falsely identified as sick against what proportion of the sick are correctly identified as such.

weight to students placed accurately into remediation or college-level coursework, and zero weight to under- and over-placement errors. Publishers of the two most commonly used remedial placement exams now provide estimated placement accuracy rates, ranging from 60 to 80 percent, to help support their validity (ACT, Inc., 2006; Mattern & Packman, 2009). In related working papers using the same data utilized here, Scott-Clayton (2012) and Belfield and Crosta (2012) also find accuracy rates in this range, at least when "success" in college coursework is defined as earning a B or better.

But accuracy rates may vary depending upon how success is defined: this can be seen in Figure 2, which provides a schematic plot of college math success rates against placement test scores. Among students scoring at the hypothetical cutoff, 45% earn a B or better in college-level math (bottom line), 62% earn a C or better (middle line), and 74% can at least pass (top line). Thus, if placed in remediation 45% of these students at the cutoff (as well as the proportion indicated by the B-or-better line for students with scores below the cutoff) will be under-placed by any criterion; if placed in college-level then 26% of those at the cutoff (as well as the proportion indicated by one minus the passing percentage for student with scores above the cutoff) will be over-placed by any criterion. The remaining proportion who would earn a C or D are ambiguously classified; placing them into the college-level course is correct under a passing criterion for success, but is a mistake under the B-or-better success criterion. Prior research consistently finds that remedial tests are more accurate at classifying students based on the B-or-better criterion than on lower success criteria (ACT, Inc., 2006; Mattern & Packman, 2009; Scott-Clayton, 2012; Belfield & Crosta, 2012). Scott-Clayton (2012) and Belfield and Crosta

(2012) find that when the goal is simply identifying who will pass versus fail, accuracy rates range between just 36 and 50 percent.[13]

Our analysis (described in detail below) will focus on error rates rather than accuracy rates, for two reasons. First, Sawyer's (1996) study demonstrates how policy conclusions based on accuracy rates can shift dramatically depending upon the definition of success. He compares accuracy rates using ACT math subtest scores versus using a locally-developed test for math placement at a large public institution in the Midwest. He finds that if success is defined as earning a B or better, using the ACT math subscore with a relatively high cutoff generates the best accuracy rates, while if success is defined as earning only a C or better, using the locally-developed test with a relatively low cutoff generates the best accuracy rates. Second, his results indicate that a wide range of potential cutoffs can generate similar accuracy rates, even as the mix of over-placement and under-placement errors changes substantially. Since these errors may have different costs (and will fall on different students), it is useful to consider them separately.

B. The Potential Value of High School Transcript Data

Even the test publishers themselves emphasize that test scores should not be used as the sole factor in placement decisions (see e.g. *Accuplacer Coordinator's Guide*, College Board, 2007). One potentially rich source of additional information is a student's high school transcript, used either in conjunction with or as an alternative to placement tests for deciding on remedial assignment. Transcripts are readily accessible, as most students submit their high school transcripts as part of the admissions process, and may yield a wealth of information on cognitive skills, subject-specific knowledge, as well as student effort and motivation. Moreover, because they are accumulated over time across a range of courses and instructors, high school grade point

[13] This may be because (to paraphrase Tolstoy) all good students are alike, while struggling students may struggle for a multitude of reasons—only some of which are related to aptitude per se.

averages (GPAs) and courses completed may simply be less noisy than brief, "one-off" exams. Yet to the best of our knowledge, high school grades and coursework have not been widely utilized or even studied as potential screening tools for assignment into remediation.

This is surprising given their demonstrated explanatory power for college outcomes and beyond. Studies have found strong associations between high school GPA and freshman GPA (Rothstein, 2004), as well as between high school efforts and college enrollment (on high school algebra, see Gamoran & Hannigan, 2000; on high school coursework, see Long et al., 2012; and on curricular intensity in high school, see Attewell & Domina, 2008). A related study by Long et al. (2009) looks at the influence of high school transcripts on the need for math remediation in Florida. However, remediation is identified as failing the Florida Common Placement Test, which presupposes the validity of the placement test. Nevertheless, the results from Long et al. (2009) suggest a strong influence of high school curriculum: remediation need varies inversely with 8th grade math scores and with the level of math taken in high school. Plausibly, information from high school appears to be predictive of performance in college.

The optimal decision rule may be a combination of placement tests and transcripts (Noble & Sawyer, 2004). A major contribution of our study is to compare the usefulness of high school transcript information either instead of or in addition to remedial test scores, and to explore whether the choice of screening device has disparate impacts by race or gender.

III. Methodology

We use a rich predictive model of college grades to examine several validity metrics under alternative policy simulations, focusing on three questions. First, how well do remedial screening tests identify students who are likely or unlikely to succeed in college-level coursework? Second, what is the incremental value of such tests above and beyond the information provided by high school transcripts generally and HS GPA in particular? We examine these questions for the full sample and for subgroups by race/ethnicity and gender. Finally, what are the trade-offs involved in establishing higher versus lower screening thresholds for remedial "treatment," and what does the chosen threshold reveal about institutional preferences?

A. Validity Metrics and Alternative Screening Policies

To address the potential oversimplification of examining a single placement accuracy rate, the simple two-by-two chart in Figure 1 could be expanded to include multiple gradations of success, and policymakers could assign separate weights to every possible outcome. But it would be presumptuous for researchers to attempt to completely specify the weights in a highly intricate welfare function. Instead, we propose a simple alternative to the accuracy rate: a loss function that we call the severe error rate (SER). Specifically, the SER combines the proportion of students predicted to earn a B or better in college-level but instead placed into remediation (the severe under-placement rate, or region D in Figure 2) with the proportion of students placed into college-level but predicted to fail there (the severe over-placement rate, or region E in Figure 2).

We see at least two advantages of the SER relative to placement accuracy rates. First, it focuses attention on the most severe assignment errors, which may be associated with the highest costs. While there may be disagreement about the "correct" placement for a student predicted to earn only a C or D in a college-level course, it seems uncontroversial that a student likely to earn an A or B should be placed directly into college-level and a student likely to fail should not. Second, by breaking the SER into its two components, we allow for severe over-placements and severe under-placements to have different weights in a welfare analysis.

Finally, to acknowledge that policymakers may care about factors beyond mis-assignment rates, we show two additional metrics for each policy simulation: the predicted success rate among those placed directly into the college-level course (using the C-or-better criterion) and the remediation rate. For example, given two different assignment systems with the same overall error rates, policymakers may prefer the system that has a higher success rate in the college-level course. And even when we hold the remediation rate fixed overall, alternative screening devices may differentially affect remediation rates within race or gender subgroups, something that we examine below.

We examine these metrics under the current test-score cutoff-based policies in place in each system (using pre-algebra and algebra test scores to screen for remedial math, and reading/writing test scores to screen for remedial English). We then compare the results to those obtained with two alternative screening devices, holding the proportion assigned to remediation fixed: 1) using an index of high school achievement alone, using information from high school transcripts, and 2) using an index that combines both test scores and high school achievement. Later, we examine how these metrics vary as we vary the proportion assigned to remediation, holding the choice of screening device fixed.

B. Estimating Severe Under- and Over-placement Rates

The SER combines the proportion of students predicted to earn a B or better in college-level but instead placed into remediation (the severe under-placement rate) with the proportion of students placed into college-level but predicted to fail there (the severe over-placement rate). The first step in calculating severe error rates is thus to estimate rich predictive models of students' probability of failing the college-level course as well as the probability of earning a B or better.[14] To do this, we restrict the sample to those who ever enrolled in a college-level course in the relevant subject (math or English) without taking a remedial course in that subject first.[15] We refer to this as the math or English estimation sample. Separately for college-level math and English courses, we run the following two probit regressions:

$$\text{(1a)} \quad \Pr(Fail = 1) = \alpha + (TestScores)\beta_1 + (HSAch)\beta_2 + X\beta_3 + \varepsilon$$

$$\text{(1b)} \quad \Pr(BorBetter = 1) = \alpha + (TestScores)\beta_1 + (HSAch)\beta_2 + X\beta_3 + \varepsilon$$

where ***TestScores*** is a vector of pre-algebra and algebra test scores for college math outcomes, and reading/writing test scores for college English outcomes, ***HSAch*** is a vector of high school achievement measures including cumulative GPA and credits accumulated (the precise measures, described in the data section below, vary somewhat across our two systems), and ***X*** is a vector of other demographic variables that have predictive value. For the LUCCS analysis, ***X*** includes race/ethnicity, gender, age, ESL status, years since high school graduation, and an indicator of whether or not the individual previously attended a local high school. For the SWCCS analysis, the model includes race/ethnicity and gender. For both systems we also include interactions of test scores and high school achievement with race/gender.[16] Even though these demographic variables cannot be used in the assignment process, they help improve the predictions that underlie our estimated error rates.[17]

[14] We group withdrawals and incompletes as failures given evidence that these outcomes are grade related (Ang & Noble, 1993).
[15] Analyzing the relationship between pre-treatment predictors and grades for those who took remediation could confound the estimates for two reasons: 1) the remedial treatment may effectively eliminate skill deficiencies, or 2) the only remediated students who make it to college-level courses may have high levels of unobserved motivation.
[16] We do not use reading/writing test scores in our predictive model for college math grades or vice versa because this would require limiting the sample to students who took tests in both subjects, and the incremental predictive power of the cross-subject test scores was comparatively small.
[17] Because we are ultimately interested in estimating overall error rates and not in predicting individual outcomes per se, the inclusion of these demographic variables turns out to make virtually no difference to our full-sample estimates of our validity metrics. Full regression results are available upon request.

After running these two regressions for the estimation sample, we then compute predicted probabilities of failing or earning a B-or-better for all students with available data, including those scoring below the cutoff (we call this larger group the prediction sample). The following equations describe how these predicted probabilities are used to compute the probability of severe under-placement or over-placement for each individual under a given assignment rule:

$$(2) \qquad \Pr(SeverelyUnderplaced = 1) = \Pr(BorBetter = 1) \quad if \;\; remediated, \; 0 \; otherwise$$

$$(3) \qquad \Pr(SeverelyOverplaced = 1) = \Pr(Fail = 1) \quad if \;\; NOT \; remediated, \; 0 \; otherwise$$

An individual's probability of being severely misplaced is simply the sum of over-placement and under-placement probabilities from (2) and (3). The SER for the sample as a whole, or for a given subgroup, is simply the average of these individual probabilities.

When we simulate severe error rates using alternative screening devices, the underlying probabilities of success from (1a) and (1b) remain fixed and we simply vary the assignment rule. When comparing across screening devices we initially choose cutoffs that ensure the proportions assigned to remediation remain roughly constant. If the alternative device were a single measure, such as cumulative high school GPA, we could simply set the cutoff at the percentile corresponding to the current test-score based cutoff. But since we are simulating alternative *sets* of predictors, we first combine these multiple measures into a single regression-based index.[18]

C. Addressing extrapolation concerns

A limitation of this type of analysis is that it requires extrapolation of relationships that are observed only for those placing directly into college-level to those who score below the current test cutoff. There is no way to be sure that the observed relationship between scores and

[18] So, for example, to select the cutoff in math using high school information, we regress college-level math grades (among only those assigned directly to college-level) on the set of high school achievement variables described above, and establish the cutoff as the 75th percentile on this index of predicted grades.

outcomes for high-scorers is equally applicable to very low-scorers.[19] For at least two reasons, however, it may be reasonable to extrapolate within a limited range below the cutoff. First, the test scores themselves are quite noisy; the COMPASS algebra module for example has a standard error of measurement of 8 points, meaning a score of 30 (LUCCS cutoff for the most recent cohorts) is not distinguishable with 95% confidence even from the lowest possible score of 15 (ACT, Inc., 2006, p. 92). Second, the earlier cohorts in LUCCS were subject to lower cutoffs (27 for the two math modules, 65 for the reading module) meaning that we do have some observations below the current cutoffs that do not rely upon extrapolation. To address extrapolation concerns, we perform a sensitivity analysis in which we exclude at the outset all students with test scores substantially below the current cutoffs.[20]

D. Institutional Context and Data

We analyze two very large, but distinct community college systems in order to improve the generalizability of our results. The datasets for this analysis were provided under restricted-use agreements with a large, urban community college system (LUCCS) including six individual institutions, and a state-wide community college system (SWCCS) comprising 50 separate institutions.[21] For additional detail on institutional context, see Scott-Clayton (2012) for LUCCS and Belfield and Crosta (2012) for SWCCS.

During our study period, LUCCS was using the COMPASS® test, with modules for numerical skills/pre-algebra, algebra, and reading, as well as a writing exam adapted slightly from the standard COMPASS® writing module (each writing exam is graded in a double-blind

[19] It is worth noting, however, that restriction-of-range in our initial predictive model does not *necessarily* lead to biased accuracy and error rates. In contrast, goodness-of-fit statistics may be biased by range restrictions even when regression coefficients based upon the restricted sample are unbiased (Rothstein, 2004).
[20] It also bears emphasis that simulations of success rates above the current cutoffs do not rely upon extrapolation and are of policy relevance on their own since many institutions, including LUCCS, have increased cutoffs recently.
[21] The LUCCS data come from four cohorts of nearly 70,000 first-time degree-seekers who entered one of the system's colleges in the fall of 2004 through 2007. The SWCCS data is from two cohorts of 49,000 students who enrolled in the academic years 2008-2010, almost all of whom are in degree programs.

system by two LUCCS readers at a central location). The SWCCS permits a range of placement tests, although the majority of students took either ACCUPLACER® or COMPASS® tests (we analyze the ACCUPLACER® and COMPASS® samples separately at SWCCS). In both systems, test cutoffs are established centrally, and students' compliance with course assignment decisions is high: while some students may not enroll in the required remedial course immediately, relatively few circumvent remediation to enroll directly in a college-level course. Re-testing is not allowed at LUCCS until after remedial coursework has been completed; at SWCCS approximately 10-15 percent of students retest prior to initial enrollment. In both cases, we use the maximum test score (prior to enrollment) for our simulations since this is what is actually used for placement in practice.

Table 1 provides demographic information on the full sample and main subsamples for the predictive validity analysis. The first column describes the overall populations. Subsequent columns are limited to students who took a placement exam in the respective subjects and then further restricted to those with high school information available. Note that these students tend to be younger and are more likely to have entered college directly from high school. Table 1 also shows the percentages assigned to remedial coursework in each subject as a result of their placement exam scores.

For LUCCS, as at higher education institutions generally, nearly six out of ten entrants are female. While more than half of LUCCS entrants are age 19 or under and come directly from high school, nearly one-quarter are 22 or older, and on average entrants are 2.6 years out of high school. Finally, LUCCS is highly diverse (over one-third of students are Hispanic, over one-quarter are black, and over ten percent are of Asian descent). Across these four cohorts of LUCCS entrants, more than three-quarters were assigned to remediation in at least one subject:

63 percent in math, 59 percent in writing or reading. The proportions among those who actually take the placement exams is necessarily higher, with 78 percent of math test takers assigned to math remediation, 76 percent of reading/writing test takers assigned to writing remediation.

For SWCCS, a slight majority of students are female and the typical entrant is a couple of years out of high school. In contrast to LUCCS, only one-third of the students are minorities. But SWCCS shows similarly high rates of remedial assignment: 70% in math, 58% in English, and three-quarters overall. These rates are slightly higher for our math and English testing samples.

Our measures of high school achievement differ somewhat between LUCCS and SWCCS. For LUCCS, the high school data comes from transcripts that are submitted as part of a system-wide college application process.[22] Staff at the system's central office identify "college-preparatory" courses in key subjects from the transcripts, and record the total number of college-preparatory units and average grades earned within each subject and overall. Thus our high school measures for LUCCS include cumulative grade point averages both overall and in the relevant subject; cumulative numbers of college-preparatory units completed, both overall and in the relevant subject; and indicators of whether any college-preparatory units were completed, both overall and in the relevant subject.

For SWCCS, our high school data come from an administrative data match to statewide K-12 public school records (and thus are only available for students who attended a public school).[23] The high school measures we use for SWCCS are: unweighted high school GPA, and from 11th and 12th grade transcripts: the total number of courses taken, the number of

[22] Students who simply show up on a given campus are known as "direct admits" and typically have much more limited background information available in the system-wide database.

[23] Though most of these students had both GPA and detailed transcript data, for some we only had GPA information. Differences between our sample and students without HS GPAs were not large.

honors/advanced courses, the number of math courses, the number of English courses, the number of F grades received, and the total number of credits taken.

IV. Results

A. Severe Error Rates and Other Validity Metrics

Table 2 reports severe error rates and other validity metrics using alternative screening devices for remedial placement. Focusing first on the "test scores" column, which simulates current policy at LUCCS and SWCCS, we see that one-quarter to one-third of tested students are severely misplaced depending upon the sample and subject. Recall that this does not imply that the remainder are all accurately placed, just that they are not *severely* misplaced. With the exception of the ACCUPLACER® math sample at SWCCS, severe under-placements are two to six times more prevalent than severe over-placements. In LUCCS, for example, nearly one in five students who take a math test, and more than one in four students who take the English tests, are placed into remediation even though they could have earned a B or better in the college-level course. This implies that nearly a quarter of remediated students in math (=18.5/76.1), and one-third of remediated students in English (=28.9/80.5), are students who probably do not need to be there.

In all of our samples for both subjects, holding the remediation rate fixed but using high school achievement instead of test scores to assign students results in both lower severe error rates and higher success rates among those assigned to college level. The reduction in severe error rates comes from reductions in both under-placements and over-placements, so unlike debates about where cutoffs should be optimally set, there is no tradeoff here between these two types of errors. With the exception of math placement in LUCCS, the reductions are substantial, suggesting that out of 100 students tested, 4 to 8 fewer students would be severely misplaced, representing up to a 30 percent reduction in severe errors compared to test-based placements. Also with the exception of math placement in LUCCS, for which improvements are more modest, using high school achievement instead of test scores results improves the success rate among those placed in college level by roughly 10 percentage points. For example, among students assigned directly to college-level, the percentage earning at least a C or better increases from 76 percent to 89 percent in the SWCCS COMPASS® sample, even though the same number of students are admitted.

Utilizing both test scores and high school transcript data for assignment generates the best placement outcomes at LUCCS, although the incremental improvement beyond using high school data alone is small. At SWCCS, the combination yields no additional improvement beyond using high school information alone.[24]

Holding remediation rates fixed as we compare alternative screening tools is a useful benchmark, but it also limits the potential for major improvements particularly with respect to the severe under-placement rate. With remediation rates of 60 to 80 percent, it is possible that many students might be under-placed regardless of what screening device is used to select them. (Note that as the remediation rate approaches either 0 or 100 percent, the choice of screening device is irrelevant.) In an extension below, we examine our validity metrics across the full range of possible diagnostic thresholds for remediation.

B. Sensitivity analysis: excluding low-scoring students

As noted above, one concern is that our underlying predictive models (expressed in equations 1a and 1b) may not extrapolate to students far below the current test score cut-offs. To address this concern, we re-run the entire analysis with very low-scoring students excluded from

[24] In some cases, the combination actually appears to do marginally worse than using high school data alone, which can result if test scores are extremely noisy.

the sample.[25] These restrictions exclude approximately 25 to 50 percent of test takers depending upon the sample and subject.

The results are presented in Table 3. We first note that there are some level shifts in these validity metrics between Tables 2 and 3. For example, because we have explicitly excluded very low-scorers, the remediation rates under current policy for these restricted samples are uniformly lower than those in Table 2. For the same reason, over-placement rates are higher and under-placement rates generally lower after low-scorers are excluded, although the overall severe error rates remain very similar.

Overall, throwing out these low-scorers does little to alter the pattern of findings from Table 2. Using high school achievement measures instead of test scores still improves both overall error rates and college-level success rates. And it is still the case that combining these two types of measures generates the best results in math at LUCCS, but for all other samples and subjects the combination provides little added value above using high school achievement alone.

C. Do Alternative Screening Tools Have Disparate Impacts by Gender or Race?

Even if high school transcript-based assignments are more accurate than test-based assignments on average, one may worry that using high school transcripts might disadvantage some students relative to others, and we may be particularly concerned if performance on these alternative measures varies systematically by race/ethnicity and/or gender. Autor and Scarborough (2008) examine the question of disparate impact in the context of job screening tests: while such screening tests may more accurately identify productive potential employees,

[25] For LUCCS: the math analysis excludes students scoring more than 10 points below the current test score cutoff on either of the two math test modules. English analysis excludes students scoring more than 3 points below the current writing test score cutoff or 10 points below the current reading test score cutoff. For SWCCS: the math and English analysis excludes students scoring more than 10 points below the current test score cutoff on either of the math or English test modules, respectively.

they may also alter the racial composition of resulting hires. They demonstrate theoretically that groups with lower average test performance are not necessarily disadvantaged by the introduction of a test, if the alternative screening practices (e.g., managerial discretion) already take these average group differences into account. They then show that the introduction of job screening tests at a large retail firm resulted in more productive hires without changing the proportion of minority hires.

In the spirit of Autor and Scarborough, we examine our validity metrics by gender and racial/ethnic identity for evidence of disparate impacts under alternative assignment rules. As with job screening tests, there is potentially an equality-efficiency trade-off in the choice of remedial screening tools if one tool is more accurate but as a result more minorities and/or females are placed in remediation.[26] Note that while we include gender and race/ethnicity in the underlying model predicting college-level outcomes (described above in equations 1a and 1b), we assume that these demographic factors cannot be used in any assignment rule. Thus, while we establish our cutoffs for the high school index and test-plus-high-school index at levels that keep the overall remediation rate fixed, the rate among any particular subgroup may change.

We present the results by gender in Table 4.[27] The first thing to note is that the pattern we found in Tables 2 and 3 holds within each gender subgroup as well: using high school transcript data instead of test scores for placement would reduce the severe error rate and increase college-level success rates for all subjects and samples; combining test scores and high school information would lead to additional incremental improvements in LUCCS math placement.

[26] There are two differences with our context, however: first, in our setting, the test-score-based policy is the default already in place, and we examine replacing or augmenting this with additional quantitative, externally verifiable measures (as opposed to a version of managerial discretion). Second, since 85 percent of LUCCS testers are minorities (with roughly 30% black, 34% Hispanic, and 10% Asian), any disparate impacts are likely to be between minority groups rather than between minorities and non-Hispanic whites.

[27] For brevity, we show only the LUCCS COMPASS® and SWCCS ACCUPLACER® results to demonstrate the consistency across samples/exams. The patterns for the SWCCS COMPASS® sample are very similar.

Nonetheless, there is some evidence of disparate impacts, in the direction that one might anticipate. Using high school information instead of test scores has the effect of decreasing the remediation rate for women but increasing it for men, for both SWCCS and LUCCS samples. This reinforces findings from prior research that men tend to do better on standardized tests while women tend to earn higher grades (see Hedges and Nowell, 1995).

Thus even while high school transcript information may be more accurate for students of both genders, some may object to a policy change that impacts men and women differentially. At least at LUCCS, using the combined test-plus-high-school index for remedial assignment appears to be a win-win situation for both genders relative to the current test-score based policy. Using the combined index for assignment would not raise the remediation rate for either subgroup relative to current policy, but would lower both over- and under-placements for both genders in both subjects, and would noticeably increase success rates for those placed directly into college-level work.[28] At SWCCS, using the combined measure moderates, but does not eliminate, the disparate impact on remediation rates by gender.

Table 5 performs the same analysis by race/ethnicity, focusing on LUCCS which has sufficiently large sample sizes within each subgroup. Again, we find that using high school information in combination with test scores maintains or reduces severe error rates and increases college-level success rates for all racial groups across all subjects. Even using high school information alone would reduce severe error rates for all groups and subjects except blacks in English and Asians in math. Again, however, we find that these improvements in accuracy must be weighed against some disparate impacts on remediation rates. In math, using high school information alone versus test scores alone lowers the remediation rate for Hispanic students by 7

[28] This slight decline in the remediation rate when using alternative assignment rules is also reflected in the full sample results in Table 3; it reflects the fact that we cannot set the cutoff at a point that will precisely preserve the original 76.1 percent remediation rate in math and 80.5 rate in English.

percentage points and increases it for Asian students by 10 percentage points, though these changes are moderated by using the combined measure for placement. In English, using high school information alone would increase the remediation rate by 11 percentage points for black students and reduce it for Asian students by nearly 25 percentage points relative to the current test-based policy. These changes in English could be moderated by using the combined test-plus-high-school measure, but would still remain large.

Table 6 applies the subgroup results from Tables 4 and 5 to simulate class compositions at LUCCS under our alternative screening devices. If high school information were used for screening instead of test scores, college-level math classes would have substantially higher proportions of female and Hispanic students; on the other hand, representation of black and Asian students would fall. In college-level English, switching to high school achievement would not change the gender composition, but representation of black students would fall by half (from 31 to 15 percent) and Asian students' representation would more than double (from 8 to 23 percent). These compositional changes are moderated, but not eliminated, when a combined measure of test scores and high school achievement is used for placement.

D. Optimal Cutoffs: Trading Off Under-placement and Over-placement

So far, we have presented results that compare alternative screening devices while holding the overall percentage of students remediated fixed at current levels. But in considering the optimal screening policy, the diagnostic threshold can be allowed to vary along with the instrument used, allowing for greater potential improvements in accuracy. For a given instrument, if policymakers weight over-placement and under-placement errors equally, then the optimal instrument and cutoff can be chosen to minimize the overall severe error rate.

Figure 3 (panel A) shows the overall severe error rates, under-placement and over-placement rates for math using alternative screening instruments with the LUCCS data. As the percentile cutoff increases, increasing proportions of students are assigned to remediation and so under-placement rates grow sharply and over-placement rates fall. Error rates for alternative instruments must converge at both the high and low end of the potential cutoff range when either no students or all students are assigned to remediation (in our figures this is complicated by the fact that the current test-only placement rule is actually based upon two sub-scores, only one of which we allow to vary here—we hold the easier pre-algebra test cutoff fixed at its current level, which matters only at the very low range of algebra cutoff scores).[29]

Except for very low cutoffs, the SERs using test scores alone are higher than under the alternative instruments we simulate. Using high school achievement alone or in addition to test scores reduces SERs most noticeably for cutoffs between the 50th and 80th percentiles. If policymakers cared only about the SER, the optimal policy would be to assign students based on the combined test-plus-high-school-achievement index with a cutoff at the 65th percentile. This policy would reduce the SER by 3.1 percentage points (13 percent) while slightly improving the success rate among those placed into college-level (shown in panel B), but perhaps the most notable difference is that it would achieve these outcomes with a remediation rate 10 percentage points lower than the rate under current policy.

Interestingly, current policy at LUCCS—indicated by the large gray circular marker on the test-only line—appears to be at an SER-minimizing level for the test-only instrument; however, the test-only SER line is relatively flat around the current cutoff, with cutoffs from the 55th to the 85th percentile all generating SERs in the range of 24 to 25 percent. Since these

[29] Because we hold the pre-algebra cutoff fixed, even with a very low algebra cutoff high proportions of students will be assigned to remediation, which tends to increase under-placements but limits over-placements, as reflected in Figure 3.

percentile cutoffs roughly correspond to remediation rates (this correspondence is exact for our two alternative instruments, which are one-dimensional indices), this implies that a very wide range of remediation rates can generate similar severe error rates.

We perform a similar analysis in Figure 4 for LUCCS English, examining error rates for a range of cutoffs on the current writing test (holding the reading test cutoff fixed) and for corresponding cutoffs on alternative instruments. Here, utilizing the high school achievement index with a cutoff at just the 35th percentile could reduce remediation rates from nearly 80 percent to 35 percent while also reducing the SER by 9 percentage points and holding the college-level success rate essentially flat. Moreover, the figure indicates that given the current test-score based instrument, the SER minimizing policy would be to admit virtually everyone to college level (though the SER is flat between the 5th and 35th percentiles).

We find similar patterns using the SWCCS data applied to both tests, illustrated in Figures 5 and 6. The SERs using a high school achievement measures alone are always lower than those using test scores alone; at SWCCS the combined assignment rule yields SERs that are virtually identical to those just using high school achievement. Given that schools are using test scores alone, the current cutoff in math appears near an SER-minimizing point (though as in LUCCS, the SER curve is nearly flat for a wide range of cutoffs) while in English the SER could be lowered relative to current levels by simply admitting everyone to college-level.

Institutions' choice of cutoff can reveal information about how they perceive the costs of different types of assignment errors. In math, the test-score-only SER is flat across a wide range of cutoffs, but both systems choose a cutoff near the top of this range; in English, both systems choose a cutoff higher than the SER-minimizing level. This suggests that institutions perceive the costs of over-placement to be significantly higher than the costs of under-placement.

V. Discussion

Our results underscore the reality that it is difficult to predict who will succeed in college by any means: regardless of the screening tool we examine, one-fifth to one-third of students are likely to be severely mis-placed. Yet among a set of feasible, if imperfect screening devices, high school transcript information is at least as useful as and often superior to placement test scores. In both math and English, using high school GPA/units alone as a placement screen results in fewer severe placement mistakes than using test scores alone (with error reductions of 12 to 30 percent relative to test scores, in all samples/subjects except LUCCS math). There is no assignment trade-off: both under-placement and over-placement errors can be reduced, and the success rate in college-level courses increased, without changing the proportion of students assigned to remediation. At LUCCS, these errors are further reduced when placement tests and high school information are used in combination, while at SWCCS we find that placement tests have little incremental value if high school information is already available. Our results are not driven by the predicted outcomes for very low scoring students (for whom our model relies more heavily on extrapolation); the pattern of findings holds even when these students are excluded.

One potential explanation for the limited utility of placement exams is that they are simply quite short (taking just 20-30 minutes per module) and thus very noisy, as noted above. Another possible factor may be a disconnect between the limited range of material tested on the exam and the material required to succeed in the typical first college-level course (Jaggars & Hodara, 2011). For example, ACT, Inc.'s own (2006) analysis suggests that the COMPASS algebra exam is more accurate for predictions of success in "college algebra" versus "intermediate algebra," but many students meet their college-level math requirement by taking courses that are not primarily algebra-based, such as introductory statistics. In comparison, high school transcript information may be both less noisy (because it is accumulated over years instead of minutes), and may capture broader dimensions of college readiness, such as student effort and motivation.

Compared with current test-score based policies, using high school information for remedial assignment not only reduces severe placement errors overall but also within each racial/ethnic and gender subgroup we examine. Despite these universal improvements in accuracy, some subgroups in some subjects do better on the tests while others do better on a high school achievement index—meaning that the choice of screening device has implications for the gender and racial/ethnic composition of college-level courses. For example, if the remediation rate is held fixed, then switching to assignment based on high school information only would increase the representation of women and Hispanics in college-level math at the expense of men, black and Asian students; while in college-level math the switch would dramatically increase the representation of Asian students while lowering representation of black students. Using a combined measure for placement could moderate the disparate impacts of this potential policy shift. An alternative approach to addressing these disparate impacts would be to use high school information but lower the cutoffs such that no subgroup would face a higher remediation rate.

Our findings provide new insights regarding how institutions weigh over-placement errors versus under-placement errors. While the over-placement problem—students admitted to college-level courses even though they end up failing there—is well known and much discussed, we find that severe under-placements are actually far more common. Our estimates suggest that one-quarter to one-third of students assigned to remediation could have earned a B or better had they been admitted directly to college-level work. Moreover, we find evidence that institutions could substantially lower their remediation rates without increasing the severe error rate. That they have not done so—in fact LUCCS has increased its cutoffs recently—suggests that institutions are more concerned about minimizing over-placements than under-placements.

This may be because the costs of over-placement fall not just on the over-placed student (who may be discouraged and/or risk losing financial aid eligibility), but also on faculty members who dislike having to fail students, as well as on other students in the college-level course who may experience negative peer effects. The costs of under-placement, in contrast, fall primarily on the institution and the under-placed student. Moreover, over-placements may simply be easier to observe: it is straightforward to document how many students are placed into a college-level course fails there, while under-placements can only be estimated statistically.

The apparently greater weight given to over-placements also appears consistent with the financial incentives of colleges. These incentives depend on the cross-subsidy (revenues minus costs) between remedial and college-level courses. In most states, revenues through state aid formulas are equal across remedial and college-level courses, although for six states the funding formula is more generous for remedial courses (in only three states it is less generous). Very few states provide data on the costs of remedial courses specifically, although these courses are more often taught by lower paid faculty and use limited technology. However, data for Ohio's two-year colleges shows that remedial courses cost 9% less than college-level courses. It thus seems quite possible that remedial courses subsidize college-level courses, giving colleges an implicit incentive to under-place students.[30] If so, colleges may face a financial constraint if remediation rates are reduced without any additional resources provided.

Finally, our findings have implications for the interpretation of prior estimates of the impact of remedial assignment, which are largely based upon regression discontinuity designs.

[30] For funding formulae, see http://faccc.org/research/FTEspending_bystate.pdf. For costs of remediation in Ohio, see http://regents.ohio.gov/perfrpt/special_reports/Remediation_Consequences_2006.pdf.

First, the relatively low predictive validity of placement exam scores (the running variable in RD studies) suggests that RD estimates may generalize beyond just students scoring near the cutoffs. This is an important conclusion, since a common critique of prior null-to-negative impact estimates has been that these estimates are local to students scoring near the cutoff, and that students well below the cutoff may experience more positive effects. On the other hand, even if test scores were as good as random—meaning that the existing null-to-negative RD estimates could be interpreted as global average treatment effects—this would not rule out the possibility of heterogeneity in treatment effects. It may simply be that treatment effects vary along some dimension other than test scores. Indeed, Scott-Clayton & Rodriguez (2012) provide evidence using LUCCS data that RD estimates of the impact of remediation are more negative for subgroups identified as low-risk on the basis of high school transcript data. It is possible that there are positive impacts of remediation for some subset of students who are underprepared, but that current test-based assignment policies simply catch too many prepared students in a widely-cast remedial net.

References

ACT, Inc. (2006). *COMPASS/ESL reference manual*. Iowa City, IA: ACT, Inc.

Ang, Cheng H., & Julie P. Noble, *The Effects of Alternative Interpretations of Incomplete and Withdrawal Grades on Course Placement Validity Indices*, Research Report No. 93-3, (Iowa City, IA: American College Testing, 1993).

Attewell, P. & T. Domina. (2008). Raising the bar: Curricular intensity and academic performance. *Educational Evaluation and Policy Analysis, 30*, 51-71.

Autor, D.H. & D. Scarborough. (2008). Does job testing harm minority workers? Evidence from retail establishments. *Quarterly Journal of Economics*, 123(1): 219-277.

Bailey, T., Jeong, D. W., & Cho, S.-W. (2010). Referral, enrollment and completion in developmental education sequences in community colleges. *Economics of Education Review, 29*(2), 255–270.

Belfield, Clive & Peter Crosta (2012). Predicting Success in College: The Importance of Placement Tests and High School Transcripts. CCRC Working Paper No. 42. New York: Community College Research Center.

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.

Bettinger, E.P., Evans, B.J., & D.G. Pope. (2011). Improving college performance and retention the easy way: Unpacking the ACT exam. NBER Working Paper 17119.

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

Bowen, William G. and Derek Bok (1998). *The Shape of the River: Long-Term Consequences of considering Race in College and University Admissions.* Princeton, NJ: Princeton University Press.

Brennan, R. L. (Ed.). (2006). *Educational measurement* (4th ed.). Westport, CT: ACE/Praeger Publishers.

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.

Carrell, S., Fullerton, R., & West, J.E. (2009). Does your Cohort Matter? Measuring Peer Effects in College Achievement. *Journal of Labor Economics*, 27(3), 439-464.

College Board. (2007). *ACCUPLACER coordinator's guide*. New York, NY: College Board.

Dadgar, M. (2012). Essays on the Economics of Community College Students' Academic and Labor Market Success. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (Accession Order No. [3506175]).

Gamoran A., & Hannigan E. C. (2000). Algebra for everyone? Benefits of college-preparatory mathematics for students with diverse abilities in early secondary school. *Educational Evaluation and Policy Analysis,* 22(3), 241–254.

Greene, Jay P., and Greg Forster (2003). *Public High School Graduation and College Readiness Rates in the United States.* Manhattan Institute Education Working Paper No. 3. New York: Manhattan Institute for Policy Research, Center for Civic Innovation.

Hedges, L.V. & A. Nowell. (1995). Sex Differences in Mental Test Scores, Variability, and Numbers of High-Scoring Individuals. *Science*, 269: 41-45.

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]).

Horn, L., & Nevill, S. (2006). *Profile of undergraduates in U.S. postsecondary education institutions: 2003-04: With a special analysis of community college students* (NCES 2006-184). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Jaggars, Shanna Smith & Michelle Hodara (2011). The Opposing Forces that Shape Developmental Education: Assessment, Placement, and Progression at CUNY Community Colleges. CCRC Working Paper No. 36. New York: Community College Research Center.

Long, M. C., Iatarola, P., & D. Conger. 2009, Explaining Gaps in Readiness for College-Level Math: The Role of High School Courses, *Education Finance and Policy*, 4: 1, 1-33.

Long, M.C., Conger, D. & P. Iatarola. 2012. Effects of High School Course-Taking on Secondary and Postsecondary Success. *American Education Research Journal*, 49, 2, 285-322.

Lusted, Lee B. "ROC Recollected [Editorial]," *Medical Decision Making* 4 (1984): 131-135.

Martorell, P., & McFarlin, I. J. (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.

Mattern, K. D., & Packman, S. (2009). *Predictive validity of ACCUPLACER scores for course placement: A meta-analysis* (Research Report No. 2009-2). New York, NY: College Board.

Noble, J.P. & R.L. Sawyer. (2004). Is high school GPA better than admissions test scores for predicting academic success in college? *College and University Journal, 79,* 17-23.

Parsad, B., Lewis, L., & Greene, B. (2003). *Remedial education at degree-granting postsecondary institutions in fall 2000: Statistical analysis report* (NCES 2004-101). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Primary Research Group, Inc. (2008). *Survey of assessment practices in higher education.* New York, NY: Author.

Rothstein, J.M. (2004). College performance predictions and the SAT. *Journal of Econometrics,* 121: 297-317.

Sawyer, R. (1996). Decision theory models for validating course placement tests. *Journal of Educational Measurement, 33*(3), 271–290.

Sacerdote, B. (2001). Peer Effects with Random Assignment: Results from Dartmouth Roommates. *Quarterly Journal of Economics*, 116(2).

Scott-Clayton, Judith (2012). Do High Stakes Placement Exams Predict College Success? CCRC Working Paper No. 41. New York: Community College Research Center.

Scott-Clayton, Judith & Olga Rodriguez. (2012). Development, Discouragement, or Diversion? New Evidence on the Effects of College Remediation National Bureau of Economic Research Working Paper No. 18328. Cambridge, MA: NBER .Venezia, A., Bracco, K. R., & Nodine, T. (2010). *One shot deal? Students' perceptions of assessment and course placement in California's community colleges*. San Francisco, CA: WestEd.

Venezia, Andrea, Bracco, K. R. & Nodine, T. (2010). *One Shot Deal? Students' Perceptions of Assessment and Course Placement in California's Community Colleges.* San Francisco, CA: WestEd.

Winston, G.C. & Zimmerman, D.J. (2004). Peer Effects in Higher Education. In *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It,* edited by C. Hoxby. Chicago: National Bureau of Economic Research and University of Chicago Press.

Zimmerman, D.J. 2003. Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *Review of Economics and Statistics*, 85(1), 9-23.

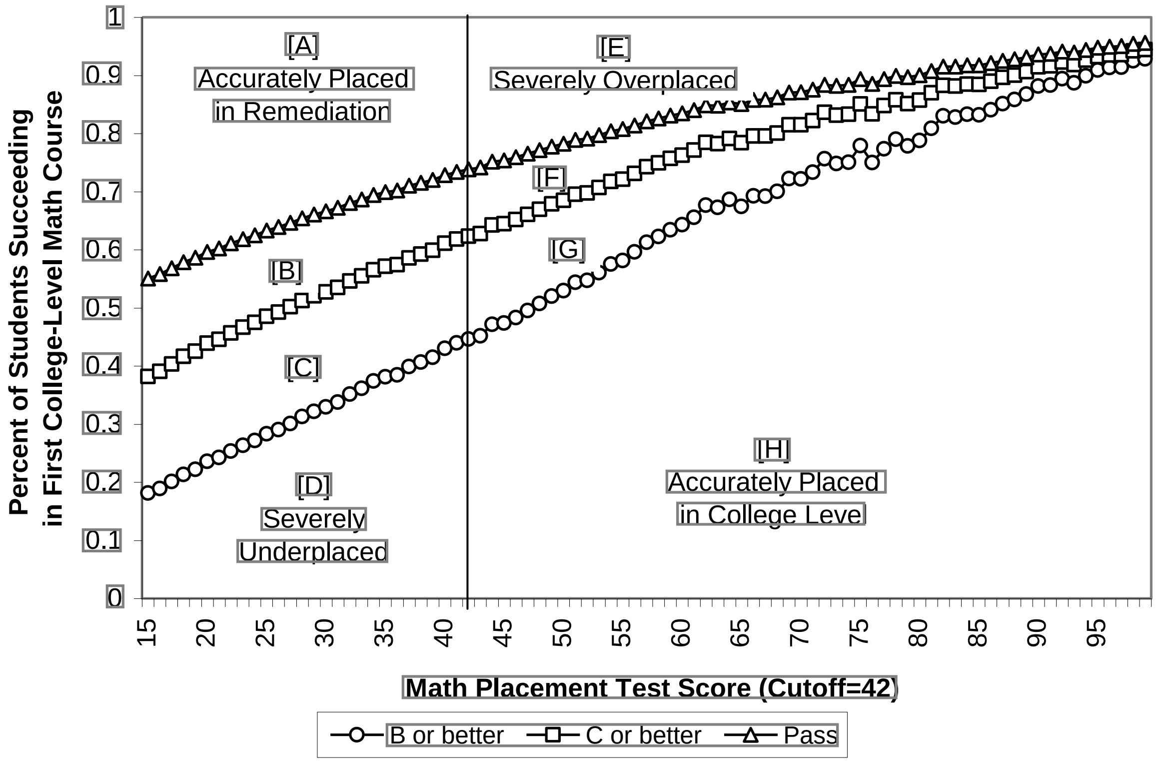
Zweig, Mark H. & Gregory Campbell (1993). Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine. *Clinical Chemistry* 39 (4): 561-577.

Figure 1

Classifications Based on Predicted Outcomes and Treatment Assignment

	Predicted to Succeed in College-Level Course?	
Treatment assignment	No	Yes
Assigned to remediation	(1) accurately placed (true positive)	(2) Under-placed (false positive)
Assigned to college-level	(3) Over-placed (false negative)	(4) accurately placed (true negative)

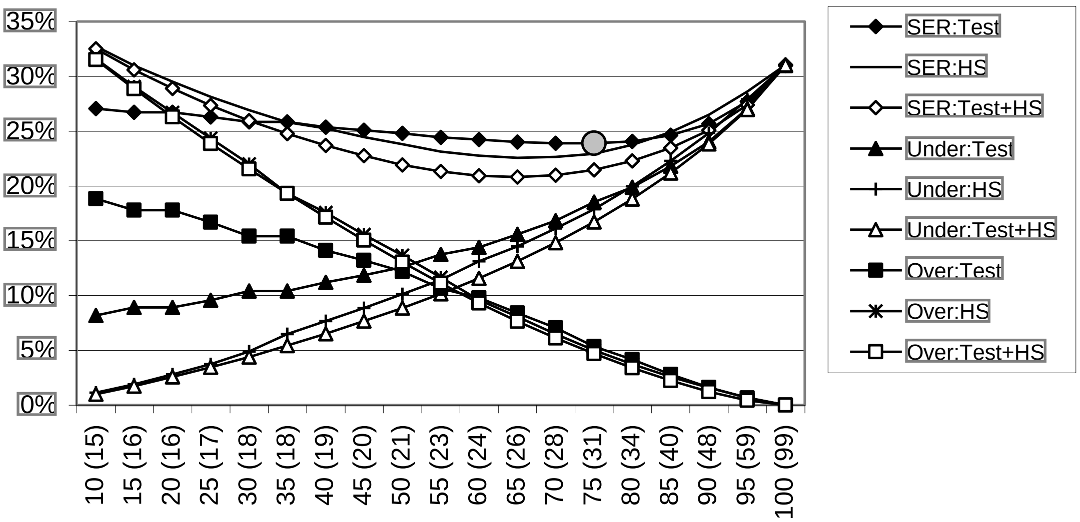
Figure 2 (Schematic). Percent Succeeding in College-Level Math, by Math Test Score



Notes: This schematic diagram illustrates the concept of accuracy and error rates using alternative definitions of success in the college-level course. The vertical line indicates a hypothetical cutoff for remedial assignment. Students scoring at this hypothetical cutoff have a 45% chance of earning a B or better in college-level math, 62% chance of earning a C or better, and 74% chance of passing. Thus, if placed in remediation 45% of these students will be severely underplaced; if placed in college-level then 1-74=26% of students with this score will be severely overplaced. The region of the chart that is unlabeled, lying between the "B or better" line and the "Passed" line, represent ambiguous classifications (i.e., the proportion likely to earn only a C or D at college level, meaning their classification will depend upon the standard of success chosen).

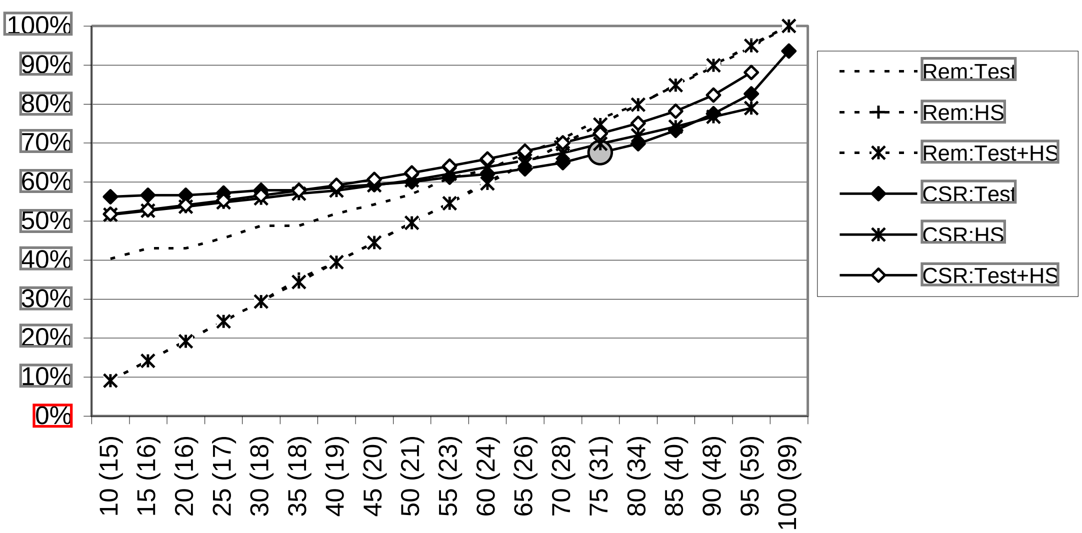
Figure 3. Assignment Outcomes by Simulated Cutoff (LUCCS, Math)

A. Severe Error Rates (SERs) and Over/Under-Placement Rates



Percentile Cutoff (Corresponding Algebra Score)

B. College-Level Success Rates (CSRs) and Remediation Rates



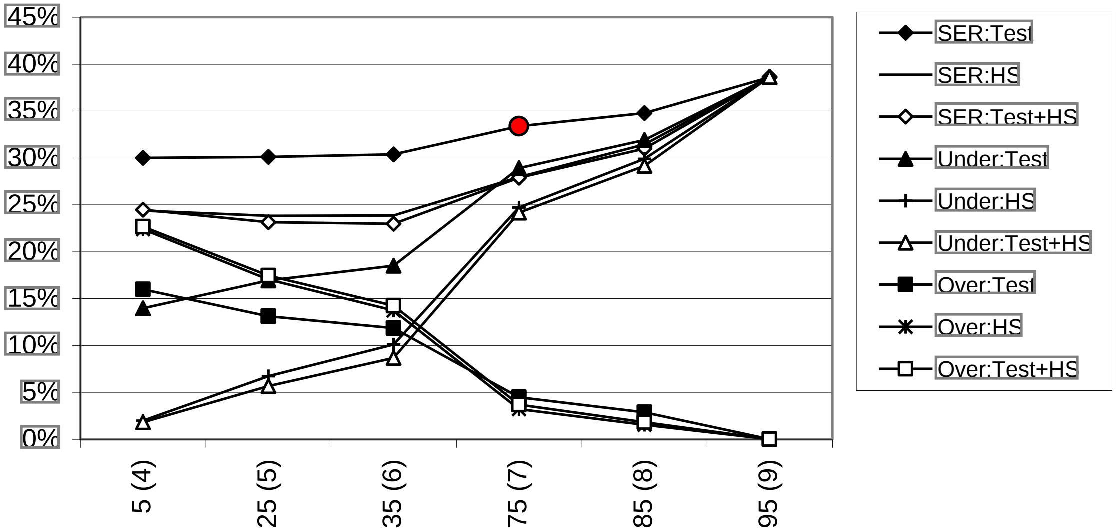
Percentile Cutoff (Corresponding Algebra Score)

Source: Administrative data from LUCCS (2004-2007 entrants).

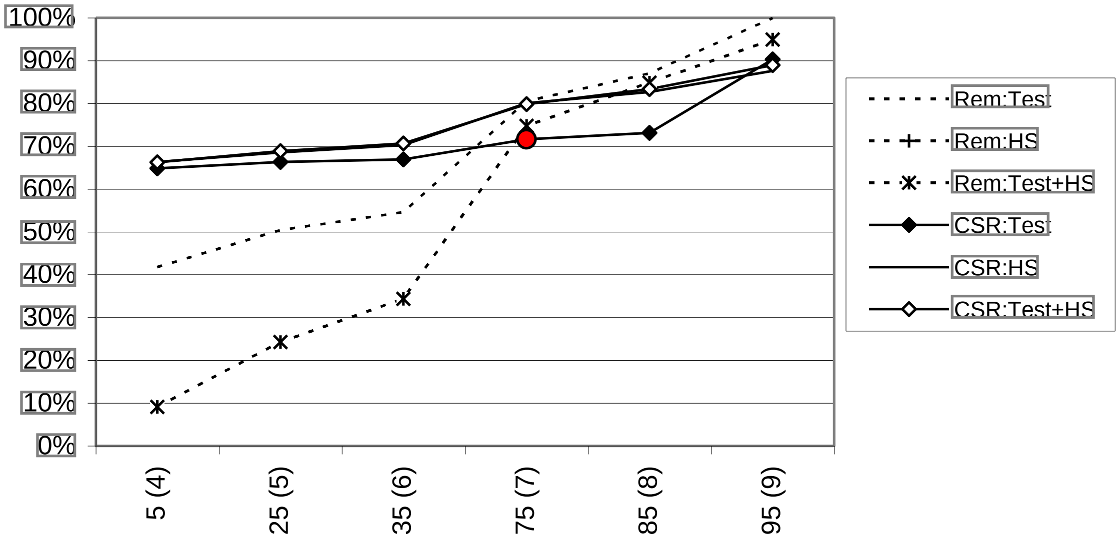
Notes: Test-only results are based on varying the algebra test cutoff while the pre-algebra cutoff is fixed at the current cuoff of 30. Allowing the pre-algebra cutoff to vary as well makes little difference for these results except for algebra cutoffs below the 50th percentile. The fixed pre-algebra cutoff explains why the test-only results begin to diverge sharply from the HS-only and Test+HS results for lower simulated algebra cutoffs: even when the algebra cutoff is very low, the fixed pre-algebra cutoff will continue to assign students to remediation, increasing underplacements but decreasing overplacements relative to HS-only and Test+HS models with similarly low cutoffs. *Gray dot indicates simulated current policy.

Figure 4. Assignment Outcomes by Simulated Cutoff (LUCCS, English)

A. Severe Error Rates (SERs) and Over/Under-Placement Rates



B. College-Level Success Rates (CSRs) and Remediation Rates

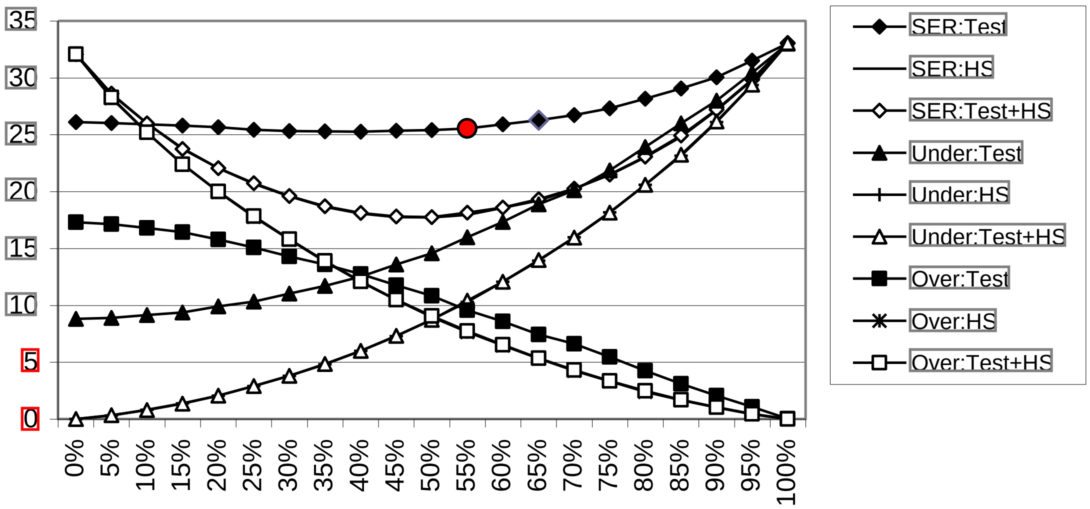


Source: Administrative data from LUCCS (2004-2007 entrants).

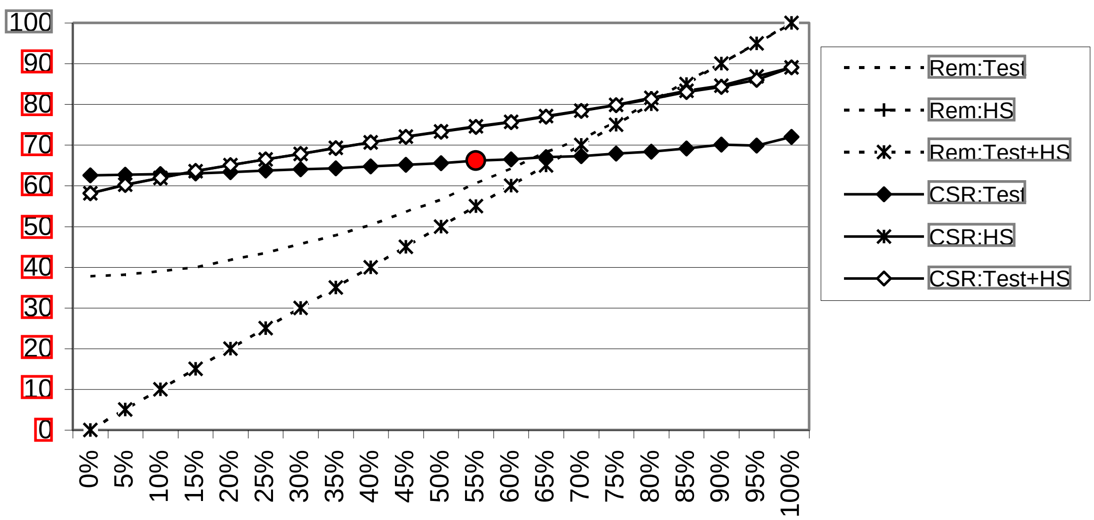
Notes: Test-only results are based on varying the writing test cutoff while the reading cutoff is fixed at the current cuoff of 70. The fixed reading cutoff explains why the test-only results begin to diverge sharply from the HS-only and Test+HS results for lower simulated writing cutoffs: even when the writing cutoff is very low, the fixed reading cutoff will continue to assign students to remediation, increasing underplacements but decreasing overplacements relative to HS-only and Test+HS models with similarly low cutoffs. *Gray dot indicates simulated outcomes of current policy.

Figure 5. Assignment Outcomes by Simulated Cutoff (SWCCS, Math)

A. Severe Error Rates (SERs) and Over/Under-Placement Rates



B. College-Level Success Rates (CSRs) and Remediation Rates

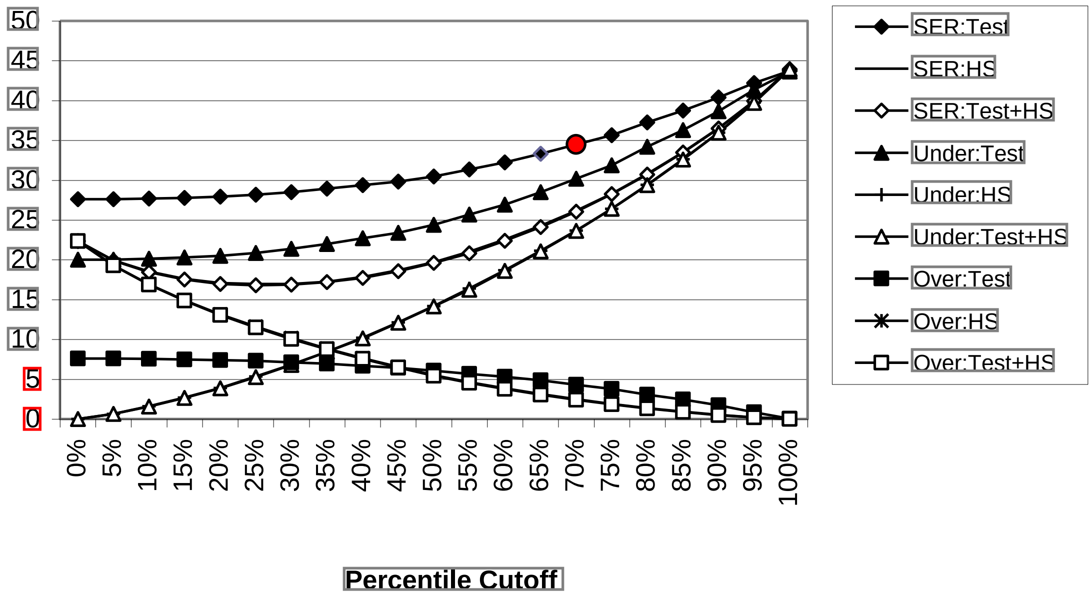


Source: Administrative data from SWCCS (2008-2009 entrants).

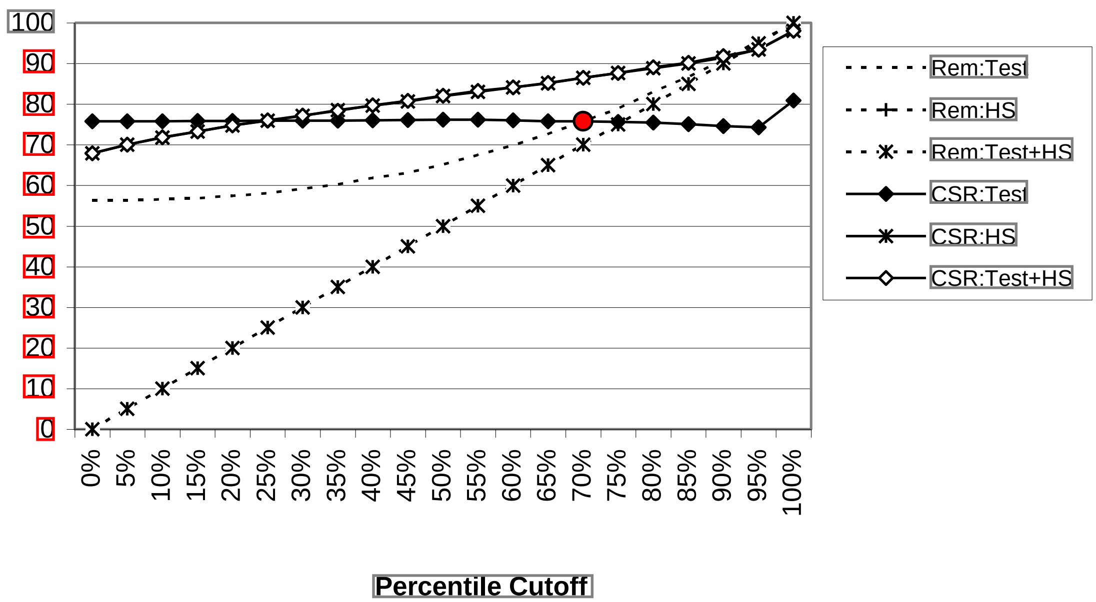
Notes: Test-only results are based on varying the algebra test cutoff while the arithmetic cutoff is fixed at the current cuoff of 55.. *RED DOT indicates simulated current policy.

Figure 6. Assignment Outcomes by Simulated Cutoff (SWCCS, English)

A. Severe Error Rates (SERs) and Over/Under-Placement Rates



B. College-Level Success Rates (CSRs) and Remediation Rates



Source: Administrative data from SWCCS (2008-2009 entrants).

Notes: Test-only results are based on varying the reading test cutoff while the sentence skills cutoff is fixed at the current cuoff of 86. *RED DOT indicates simulated current policy.

Table 1. Selected Demographics

	Math Sample			English Sample	
	All Degree-Seeking Entrants	Math Test-Takers	Math Test-Takers with High School Achievement Data	Reading/ Writing Test-Takers	Reading/ Writing Test-Takers with High School Achievement Data
A. LUCCS Sample					
% female	56.8	57.3	58.2	56.7	57.2
% minority	85.4	85.5	84.5	86.8	86.0
Age (years)	21.0	21.1	20.8	21.5	21.2
Years since high school graduation	2.6	2.7	2.2	3.0	2.4
% entering <1 year after high school	55.0	53.4	62.8	50.1	59.3
Average cumulative HS GPA	70.3	69.7	72.6	69.5	72.5
Average COMPASS algebra score	27.0	26.9	27.5	26.5	27.1
Average COMPASS reading score	70.8	71.2	71.1	70.9	70.9
% assigned to remediation					
…in math	63.0	78.9	77.8	70.1	68.5
…in English (reading or writing)	59.4	63.8	61.8	76.1	75.4
…in either subject	81.5	91.4	90.8	92.2	#N/A
Sample size	68,220	54,412	37,860	50,576	34,808
B. SWCCS Sample					
% female	53.7	51.9	49.9	53.8	51.0
% minority	33.1	29.9	27.0	33.3	28.8
Age (years)	22.5	22.0	18.7	22.5	18.7
Average cumulative HS GPA	2.5	2.5	2.6	2.5	2.5
Average COMPASS algebra score	34.8	34.5	39.7	34.4	38.9
Average COMPASS reading score	79.9	82.6	81.9	79.8	78.9
% assigned to remediation					
…in math	70.2	70.2	60.7	70.9	61.5
…in English (reading or writing)	58.4	55.1	59.5	58.4	62.3
…in either subject	74.8	80.6	76.7	75.5	75.5
Sample size	48,735	31,587	10,897	47,230	14,789

Source: Administrative data from LUCCS (2004-2007 entrants) and SWCCS (2008-2009 entrants).

Notes: At LUCCS, Approximately 30 percent of test takers do not have high school achievement data available because they enrolled directly at an institution instead of via a centralized application system. For SWCCS, full transcript data from the public school system within the state was matched to college enrollees. Thus, data are available only for those matriculating from the public schools.

Table 2. Predicted Severe Error Rates and Other Validity Metrics Using Alternative Measures for Remedial Assignment

	Measures Used for Remedial Assignment					
	Test Scores	HS GPA/ Units	Test+HS Combined	Test Scores	HS GPA/ Units	Test+HS Combined
A. LUCCS Sample	COMPASS® Sample					
Math	*N=37,813*					
Severe error rate	23.9	22.9	21.4	-	-	-
Severe overplacement rate	5.3	5.0	4.7	-	-	-
Severe underplacement rate	18.5	17.9	16.7	-	-	-
CL success rate (>=C), if assigned to CL*	67.5	69.8	72.4	-	-	-
Remediation rate	76.1	74.7	74.7	-	-	-
English	*N=34,697*					
Severe error rate	33.4	29.4	29.3	-	-	-
Severe overplacement rate	4.5	2.2	2.7	-	-	-
Severe underplacement rate	28.9	27.2	26.6	-	-	-
CL success rate (>=C), if assigned to CL*	71.6	81.8	81.4	-	-	-
Remediation rate	80.5	79.8	79.8	-	-	-
B. SWCCS Sample	COMPASS® Sample			ACCUPLACER® Sample		
Math	*N=4,881*			*N=6,061*		
Severe error rate	34.2	26.9	27.2	26.6	18.9	18.9
Severe overplacement rate	5.8	2.5	2.7	12.3	8.2	8.2
Severe underplacement rate	28.4	24.4	24.5	14.3	10.7	10.7
CL success rate (>=C), if assigned to CL*	76.4	88.5	88.1	65.1	74.5	74.4
Remediation rate	68.5	70.0	70.0	54.0	55.0	55.0
English	*N=8,307*			*N=6,573*		
Severe error rate	26.2	19.6	19.6	33.5	26.9	26.8
Severe overplacement rate	8.8	4.9	5.0	5.6	2.7	2.6
Severe underplacement rate	17.3	14.7	14.6	27.8	24.3	24.2
CL success rate (>=C), if assigned to CL*	72.6	82.4	82.4	76.0	86.4	86.5
Remediation rate	57.6	60.0	60.0	70.2	70.0	70.0

Source: Administrative data from LUCCS (2004-2007 entrants) and SWCCS (2008-2009 entrants).
Notes: The severe error rate is the sum of the proportion of students 1) placed into college level and predicted to fail there and 2) placed into remediation although they were predicted to earn a B in the college level. The remediation rate is the percentage of all students assigned to remediation. *The CL success rate is the proportion of students assigned directly to college-level coursework in the relevant subject who are predicted to earn at least a C grade or better. Note that SWCCS institutions use one of two types of tests (and in rarer cases, both): Accuplacer and/or COMPASS. The sample of students with Accuplacer data is held constant across columns 1, 3, and 5 as we simulate using alternative measures for placement; similarly the sample of students with COMPASS data is held constant across columns 2, 4, and 6.

Table 3. Predicted Severe Error Rates and Other Validity Metrics, Restricting Analysis to Exclude Low-Scoring Students

	Measures Used for Remedial Assignment					
	Test Scores	HS GPA/ Units	Test+HS Combined	Test Scores	HS GPA/ Units	Test+HS Combined
A. LUCCS Sample	COMPASS® Sample					
Math	*N=21,894*					
Severe error rate	25.6	23.9	21.9	-	-	-
Severe overplacement rate	10.0	9.1	8.8	-	-	-
Severe underplacement rate	15.6	14.7	13.0	-	-	-
CL success rate (>=C), if assigned to CL*	66.9	69.0	71.0	-	-	-
Remediation rate	56.4	54.7	54.5	-	-	-
English	*N=26,246*					
Severe error rate	33.5	29.4	29.2	-	-	-
Severe overplacement rate	5.9	3.5	3.8	-	-	-
Severe underplacement rate	27.5	25.8	25.4	-	-	-
CL success rate (>=C), if assigned to CL*	71.6	79.9	80.1	-	-	-
Remediation rate	74.2	74.7	74.7	-	-	-
B. SWCCS Sample	COMPASS® Sample			ACCUPLACER® Sample		
Math	*N=2,431*			*N=3,461*		
Severe error rate	28.7	17.6	17.8	27.6	20.5	20.5
Severe overplacement rate	11.7	7.5	7.6	21.6	17.6	17.6
Severe underplacement rate	17.0	10.1	10.1	6.0	2.9	2.9
CL success rate (>=C), if assigned to CL*	76.4	84.6	84.3	65.1	70.1	70.1
Remediation rate	36.7	35.0	35.0	19.4	20.0	20.0
English	*N=4,780*			*N=3,333*		
Severe error rate	25.2	17.3	17.4	29.8	20.6	20.6
Severe overplacement rate	15.3	12.0	12.1	11.7	6.7	6.7
Severe underplacement rate	9.9	5.3	5.3	18.1	13.9	13.9
CL success rate (>=C), if assigned to CL*	72.6	78.2	78.2	76.1	84.5	84.6
Remediation rate	26.4	25.0	25.0	38.1	40.0	40.0

Source: Administrative data from LUCCS (2004-2007 entrants) and SWCCS (2008-2009 entrants).
Notes: The severe error rate is the sum of the proportion of students 1) placed into college level and predicted to fail there and 2) placed into remediation although they were predicted to earn a B in the college level. The remediation rate is the percentage of all students assigned to remediation. *The CL success rate is the proportion of students predicted to earn at least a C grade or better, conditional upon being assigned directly to college-level coursework. LUCCS: Math analysis excludes students scoring more than 10 points below the current test score cutoff on either of the two math test modules. English analysis excludes students scoring more than 3 points below the current writing test score cutoff or 10 points below the current reading test score cutoff. SWCCS: Math and English analysis excludes students scoring more than 10 points below the current test score cutoff on either of the math or English test modules, respectively.

Table 4. Predicted Severe Error Rates
Using Alternative Measures for Remedial Assignment, BY GENDER

	Men			Women		
	Test Scores	HS GPA/ Units	Test+HS Combined	Test Scores	HS GPA/ Units	Test+HS Combined
A. LUCCS (COMPASS®) Sample						
Math		*N=15,814*			*N=22,046*	
Severe error rate	22.6	21.7	20.0	24.8	23.8	22.5
Severe overplacement rate	7.0	5.2	6.0	4.2	4.9	3.7
Severe underplacement rate	15.6	16.5	14.0	20.6	18.9	18.6
CL success rate (>=C), if assigned to CL*	62.2	66.6	67.8	72.2	72.0	76.3
Remediation rate	73.4	76.2	72.7	78.1	73.7	76.2
English		*N=14,884*			*N=19,924*	
Severe error rate	29.5	26.3	25.8	36.2	31.8	31.9
Severe overplacement rate	4.5	2.2	2.7	4.4	2.2	2.7
Severe underplacement rate	25.0	24.1	23.0	31.8	29.5	29.2
CL success rate (>=C), if assigned to CL*	67.1	79.7	78.5	74.3	83.0	83.2
Remediation rate	82.7	82.5	82.3	78.8	77.8	77.9
B. SWCCS (ACCUPLACER®) Sample						
Math		*N=2,975*			*N=3,086*	
Severe error rate	27.0	19.3	19.3	26.2	18.4	18.5
Severe overplacement rate	14.7	7.8	8.0	10.0	8.6	8.5
Severe underplacement rate	12.3	11.5	11.3	16.2	9.8	10.0
CL success rate (>=C), if assigned to CL*	59.9	71.3	71.1	70.7	76.7	76.9
Remediation rate	51.7	61.8	61.2	56.2	48.4	49.1
English		*N=3,220*			*N=3,353*	
Severe error rate	32.7	27.6	27.8	34.2	26.3	25.9
Severe overplacement rate	7.4	2.9	2.9	3.9	2.4	2.4
Severe underplacement rate	25.3	24.7	24.9	30.3	23.9	23.5
CL success rate (>=C), if assigned to CL*	69.0	82.9	82.8	83.1	88.7	88.8
Remediation rate	69.4	75.7	76.1	71.1	64.5	64.1

Source: Administrative data from LUCCS (2004-2007 entrants) and SWCCS (2008-2009 entrants).

Table 5. Predicted Severe Error Rates and Other Validity Metrics
Using Alternative Measures for Remedial Assignment, by Race/Ethnicity (LUCCS only)

	White, non-hispanic			Black, non-hispanic			Hispanic		
	Test Scores	HS GPA/ Units	Test+HS Combined	Test Scores	HS GPA/ Units	Test+HS Combined	Test Scores	HS GPA/ Units	Test+HS Combined
Math		*N=5,609*			*N=10,901*			*N=12,932*	
Severe error rate	26.1	25.8	23.7	25.9	25.0	24.1	20.3	18.0	17.5
Severe overplacement rate	4.8	5.2	4.7	5.1	3.9	3.9	5.1	5.9	4.9
Severe underplacement rate	21.3	20.5	19.0	20.8	20.9	20.1	15.3	12.1	12.6
CL success rate (>=C), if assigned to CL*	76.9	76.6	79.1	61.9	67.8	68.7	54.2	60.4	62.9
Remediation rate	69.6	67.0	66.8	80.3	81.9	81.4	84.4	77.3	80.8
English		*N=4,655*			*N=9,793*			*N=12,169*	
Severe error rate	41.2	34.7	33.4	28.8	30.3	28.9	31.7	26.7	27.2
Severe overplacement rate	5.1	1.9	3.6	4.6	1.0	1.9	4.7	2.6	2.9
Severe underplacement rate	36.0	32.8	29.8	24.1	29.3	27.0	27.0	24.1	24.3
CL success rate (>=C), if assigned to CL*	76.9	91.0	86.4	72.7	83.6	81.2	65.0	74.4	76.0
Remediation rate	73.9	72.4	68.4	78.4	89.4	85.9	83.3	80.5	81.7

	Asian			All other (incl. unknown)		
	Test Scores	HS GPA/ Units	Test+HS Combined	Test Scores	HS GPA/ Units	Test+HS Combined
Math		*N=3,944*			*N=4,474*	
Severe error rate	24.9	28.7	22.5	25.6	23.6	22.5
Severe overplacement rate	7.2	5.0	6.1	5.9	4.8	4.7
Severe underplacement rate	17.6	23.6	16.4	19.7	18.7	17.8
CL success rate (>=C), if assigned to CL*	79.9	81.7	82.4	65.1	69.3	71.0
Remediation rate	47.9	58.1	48.5	74.8	74.2	74.3
English		*N=4,188*			*N=4,003*	
Severe error rate	37.7	26.9	29.3	36.2	32.3	31.9
Severe overplacement rate	2.6	5.4	3.4	4.6	1.2	2.5
Severe underplacement rate	35.1	21.3	25.8	31.6	31.1	29.4
CL success rate (>=C), if assigned to CL*	75.6	81.9	84.1	74.7	86.3	84.2
Remediation rate	86.7	61.8	73.4	78.5	81.8	79.0

Source: Administrative data from LUCCS (2004-2007 entrants) and SWCCS (2008-2009

Table 6. Simulated Composition of College-Level Courses, Using Alternative Measures for Remedial Assignment (LUCCS only)

		College-Level Students		
	Tested Students	Test Scores	HS GPA/ Units	Test+HS Combined
Math				
Female	58.2	53.4	60.7	54.9
White	14.8	18.9	19.3	19.5
Black	28.8	23.7	20.6	21.3
Hispanic	34.2	22.3	30.7	26.0
Asian	10.4	22.7	17.3	21.3
Other/unknown race/ethnicity	7.3	6.9	6.9	6.7
Sample size	37,860	9,041	9,560	9,560
English				
Female	57.2	62.1	62.8	62.6
White	13.4	17.9	18.3	21.0
Black	28.1	31.2	14.8	19.6
Hispanic	35.0	30.0	33.8	31.6
Asian	12.0	8.2	22.8	15.9
Other/unknown race/ethnicity	7.2	7.5	5.7	6.5
Sample size	34,808	6,787	7,024	7,031

Source: Administrative data from LUCCS (2004-2007 entrants).