Improving Indicators of College Readiness: Methods for Optimally Placing Students Into Multiple Levels of Postsecondary Coursework

Daniel M. Leeds[1]
Research Analyst
CNA
3003 Washington Blvd
Arlington, VA 22201
Leedsd@cna.org

Christine G. Mokher[2]
Associate Professor
Florida State University
1114 W. Call Street
Tallahassee, FL 32306
cmokher@fsu.edu

Manuscript submitted to *Educational Evaluation and Policy Analysis*
Initial Submission on August 21, 2018
First Revise & Resubmit on March 18, 2019
Second Revise & Resubmit on August 20, 2019
Accepted on October 4, 2019

Acknowledgements: The Institute of Education Sciences, U.S. Department of Education, supported this research through Grant R305E120010 to [INSTITUTION NAME]. The report represents the best opinion of the authors at the time of issue and does not represent the views of the Institute or the U.S. Department of Education. We would like to thank the staff of the Florida Department of Education for their support of the project and for providing the data required for our analyses. This study was conducted under an Institutional Review Board (IRB) approval from Western IRB (WIRB protocol # 20121640).

Title page

Manuscript submitted to Educational Evaluation and Policy Analysis on August 21, 2018 and resubmitted on March 18, 2019 and August 20, 2019. Accepted on October 4, 2019.

Title: Improving Indicators of College Readiness: Methods for Optimally Placing Students Into Multiple Levels of Postsecondary Coursework

Abstract

Over half of community college students place into developmental education, resulting in significant financial costs. We extend previous research demonstrating that using placement tests to assign students into developmental courses results in frequent misplacement. We use Florida data to explore the extent to which students are misplaced into their first college course by more than one level. Results suggest moving away from placement tests and toward other metrics (like high school GPA) may not be as beneficial in Florida as was demonstrated in prior studies. Rather, it may be preferable to choose cutoffs that minimize misplacement than to use new metrics. States should consider their own unique contexts and examine whether they can improve placement accuracy by changing cut scores.

(Word count: 120/120)

Keywords: educational economics, efficiency, course placement, postsecondary education

Author Bios

[1] Daniel M. Leeds is a Research Analyst in the Resources and Force Readiness division at CNA. His research interests include economics of education, labor economics, and demography.

[2] Christine G. Mokher is an Associate Professor of Higher Education at Florida State University and a Senior Research Scientist at CNA. Her research examines state and local policies focused on college- and career-readiness and success, with a particular emphasis on student transitions from secondary to postsecondary education.

Improving Indicators of College Readiness: Methods for Optimally Placing Students Into Multiple Levels of Postsecondary Coursework

High school students are often unaware of the skills needed for college. Roughly 90% of high school freshmen expect to complete some postsecondary education (Kirst, 2005), but many are unprepared to do so. Students in low-performing high schools and first-generation college students are particularly unaware of how the knowledge and skills needed to graduate from high school differ from those needed in college (Boser & Burd, 2009; National Commission on the High School Senior Year, 2001). Students' underpreparation for postsecondary education has financial implications for both students and institutions, as developmental education (DE) courses among first-time, degree-seeking fall enrollees are estimated to cost $7 billion annually (Scott-Clayton, Crosta, & Belfield, 2014).

Despite the high costs of DE and the evidence questioning its effectiveness (Valentine, Konstantopoulos, & Goldrick-Rab, 2017), DE enrollment rates remain high. A series of studies from the Community College Research Center (Belfield & Crosta, 2012; Scott-Clayton, 2012; Scott-Clayton et al., 2014) found the use of college placement tests to assign students to DE led to frequent placement errors. Some students are *overplaced* into for-credit courses in which they are predicted to fail; likewise, there are also a substantial number of students who are *underplaced* into DE courses when they likely could have passed for-credit courses. Although the studies from the Community College Research Center were conducted in limited contexts that may not be generalizable to other states or placement exams, the findings have been influential in policy recommendations, encouraging the nationwide use of multiple measures for college course placement (i.e., Education Commission of the States, 2016). There is a need for further research on placement accuracy to determine if the findings replicate in other settings and in different policy contexts, such as one where colleges offer multiple levels of DE courses.

Responding to the need for further research on placement accuracy, we examine how well placement exams predict college success using data from the State of Florida. Several unique factors differentiate Florida from the contexts of prior studies. First, Florida Department of Education staff worked with McCann Associates to develop Florida's Postsecondary Education Readiness Test (PERT), which was aligned with both the state K-12 standards and Postsecondary Readiness Competencies. This state-specific alignment of the placement test to the standards could potentially lead to improvements in predictive power over national placement tests that may not be as closely aligned with any local context. Second, the state required that the PERT be used by all public state colleges, using identical cutoffs, for placement into initial courses. This requirement resulted in a large statewide sample with a diverse set of institutions. Third, during the timeframe of this study, the state was implementing the Florida College and Career Readiness Initiative (FCCRI), which required all mid-performing high school students to take the PERT in Grade 11. Students who scored below college-ready were required to take math or English college readiness courses in Grade 12. This policy allows us to further examine how well early assessment in high school can predict future college success.

Our approach builds on the work of Scott-Clayton et al. (2014) who study the severe error rate (SER). The SER considers the share of students predicted to earn a B or better in for-credit courses but placed into DE (underplacement) and the share of students placed into for-credit courses but predicted to fail (overplacement). This error rate is referred to as "severe" because it represents clear error in placement. In our study, we also examine how often students are misplaced by more than one level, as Florida places students into lower and upper level DE.

Consideration of multiple levels of DE is important because policy implications depend on where both cutoffs are relative to one another. Either cutoff can act as a constraint on the other since the cutoff for college-level courses cannot be set below the cutoff for upper DE courses.

We find that using exam scores alone, students tend to be overplaced in math but underplaced in English. In both subjects, eliminating placement into upper DE would minimize misplacement. This suggests that very few Florida students would benefit from taking a single DE course. Instead, most students are either so far behind that they likely need two courses to catch up, or they are prepared enough to go directly into college-level courses. Using high school transcript data can reduce misplacement, but not as much as adjusting test cutoffs values to minimize overplacement and underplacement rates (e.g., reducing the current cutoff of 85 for upper level DE English to 65).

Our study makes several important new contributions to the literature. First, we provide a way to conceptualize calculating error within a system with two placement cutoffs by incorporating the extent to which students are doubly under or overplaced (e.g., when a test score places a student in a for-credit course, but that student would be likely to fail not only that course but even a course one level less challenging). Many colleges offer multiple levels of DE courses, so it is important to look beyond assessing error at only one cutoff. Second, our results suggest that moving away from placement tests and toward other metrics (like high school GPA) may not be as clearly beneficial in Florida as it is in the context of prior studies. Instead, our results suggest that it may be better to adjust placement cutoffs so as to minimize misplacement error than to place students using new metrics, particularly in math. As a result, states should consider their own unique contexts and examine whether they can improve placement accuracy by changing cut scores before rushing to make more dramatic changes to placement policies like eliminating placement tests or developing complicated placement algorithms with multiple measures. Context is important because placement accuracy may depend on multiple factors including the characteristics of students being tested, the reliability and validity of the placement test, the content of the placement test, and how well this content aligns with college courses taken by students. Third, our findings indicate that SERs are dependent on the testing context, including the test type and the cutoffs. This has implications for policymakers considering abandoning placement testing or DE courses, as they may be better off examining the accuracy and error of their current placement policies before making significant changes. Lastly, our study examines how taking the placement test in high school (rather than upon college entry) may affect the accuracy of placement predictions. These results should be interpreted with caution since many students in our sample did not retake the PERT upon college entry. Nevertheless, we provide preliminary evidence on an important placement issue with implications for policymakers as states move toward early assessments and other college-readiness interventions.

Literature Review

It is important to accurately assign students to college courses because placement decisions will likely impact students' future college success. Valentine et al. (2017) examined the effects of placement into DE courses at community colleges and 4-year universities. They conducted a meta-analysis of DE regression discontinuity studies and found students narrowly assigned to DE performed significantly worse than their nondevelopmental peers on multiple outcomes, including pass rates of gateway courses and degree completion rates. Other research, though, suggests that, even though overall effects of DE tend to be negative, there may be some positive effects for students with lower levels of academic preparation (Boatman & Long, 2018).

A review of the literature on college entrance exams indicated high school GPA was a stronger predictor of college grades and 4-year graduation rates, even though grading standards varied by school (Atkinson & Geiser, 2009). Another review of the literature on grading found that high school grades more consistently predicted postsecondary enrollment, persistence, and degree completion than standardized tests (Brookhart et al., 2016). Despite promising evidence on the predictive validity of high school grades, there are some concerns that high school grades are subject to grade inflation and may not reflect the skills needed for college-level work. For example, McCormick and Lucas (2011) found that most secondary math teachers think they are covering appropriate material to prepare students for college, whereas the majority of college professors believe students come to college unprepared for college-level math courses. Compared to high school grades, a placement test may provide a more objective measure of college readiness that is more closely aligned with the skills needed for college-level courses.

Although most assessment studies have focused on the ACT or SAT, several have examined the accuracy of community college placement tests. Belfield and Crosta (2012) used data from a statewide community college system to examine the association between COMPASS and ACCUPLACER placement scores and DE course grades, college GPA, credits earned, and success in gatekeeper math and English courses. The study resulted in weak associations between placement scores and most outcomes, although high school GPA had stronger associations, particularly with college GPA and credits earned. Other high school transcript data (i.e., number of high school math and English courses, number of honors courses, number of F grades, and number of credits) did not improve predictive power.

In a related working paper, Scott-Clayton (2012) evaluated the predictive validity of COMPASS in a large, urban community college system. Results from this study suggested that placement test scores were likely to misplace significant numbers of students, particularly in English. Yet, in this study, the addition of indicators for high school achievement and student background characteristics were more likely to reduce severe misplacement.

Scott-Clayton et al. (2014) extended Belfield and Crosta's (2012) and Scott-Clayton's (2012) analyses by examining the accuracy of students placed into DE and for-credit courses, using COMPASS and ACCUPLACER data. They calculated the SER as the proportion of students predicted to earn a B or better in for-credit courses but placed into DE (underplacement) plus the proportion placed into for-credit courses but predicted to fail (overplacement). Approximately 1 in 4 students were severely misplaced in math and 1 in 3 in English. By using high school transcript data, it is predicted colleges could reduce misplacement by up to 30%, with little benefit from adding test scores to transcript data. The authors posited that transcript data might be more accurate because placement tests tend to be short in duration, provide noisy measures, and may not cover all skills needed for college success (Scott-Clayton et al., 2014).

Different types of placement tests may have different accuracy. Ngo and Melguizo (2015) used the same methods as Scott-Clayton et al. (2014) to calculate the SER before three colleges in a California community college district switched from a diagnostic assessment to a computer-adaptive test for math placement and after they switched assessments. The researchers found the SER was higher with the computer-adaptive test and concluded that diagnostic tests may improve placement accuracy relative to more commonly used computer-adaptive tests. However, computer-adaptive tests tend to be more cost effective and take less time to administer.

Placement policies also need to define how test scores are used in determining college readiness. Score cutoffs are often set by college or state agency employees with little knowledge about which tests most effectively place students, how to evaluate cutoffs, and which measures can address tests' shortcomings (Melguizo, Kosiewicz, Prather, & Bos, 2014). Students are therefore placed into courses that do not align with their actual level of readiness.

Given that decisions about setting placement score cutoffs are usually ill-informed, there is likely much variation across colleges in how well cutoffs assign students to the appropriate course levels. Most research to date on DE placement accuracy has been limited to small samples of community colleges, and replication studies are needed to understand whether their results are generalizable to other settings. Further, these studies have all examined placement policies where there is a single level of DE, even though many community colleges provide two or more levels of DE classes in each subject. Our study provides a unique opportunity to use data from an entire college system in a large and diverse state with multiple levels of DE to inform the debate about how to improve placement accuracy in different contexts.

Context of this Study

In Florida, DE is provided almost exclusively by 28 public community colleges, which are referred to as state colleges. Florida's placement policies have evolved over time. To provide context for the study, we describe the placement policies affecting students in our sample who were enrolled in Grade 11 in 2011/12, Grade 12 in 2012/13, and any state college in Fall 2013.

Florida state colleges use the Postsecondary Education Readiness Test (PERT), a computer-adaptive test of reading, writing, and math with 30 items on each section. The PERT determines college course placement using scale scores from 50 to 150. Students scoring below a fixed cutoff are required to take lower and upper DE, those scoring above this cutoff but below another cutoff are required to take upper DE, and those scoring above both cutoffs are labeled "college ready" and placed into for-credit courses.

Under the FCCRI, the PERT became mandatory in 2011/12 for Grade 11 students with midrange scores on the Grade 10 Florida Comprehensive Achievement Test (FCAT) in math or reading. Students not meeting college-ready cutoffs on the PERT were required to enroll in a Grade 12 CRS course in the corresponding subject area. Students who completed CRS courses in high school had to retake the PERT and earn college-ready scores to enroll in for-credit college courses. Retesting could occur any time prior to college matriculation; some high schools offered PERT retesting at the end of CRS courses, but they were not required to do so. Students were given information about the college-ready cut scores on the PERT and their performance, which may have provided motivation to retest.

Research Questions

We build on previous studies examining how well college placement exams predict college success, using the State of Florida as a new context. We extend prior research by considering Florida's use of separate cutoffs for upper DE and for-credit courses, as well as the timing of test scores taken by students in high school and college. We also assess the extent to which placement cutoffs in our data reflected consistent policy preferences. Specifically, we address the following four questions:

1. How accurately were Florida state college students placed into DE versus for-credit courses?
2. Could placement accuracy in the first college course have been improved by using different cutoffs, different metrics, or different sets of test scores (for students who retake the exam)?
3. How does placement accuracy change when taking into account multiple DE levels?
4. What do current cutoffs suggest about policymakers' preferences for overplacement versus underplacement?

Data

Our data consist of student records in Florida's K–20 Education Data Warehouse. We restricted our analysis to students who first enrolled in Grade 11 during 2011/12 and seamlessly enrolled in a state college in 2013/14. This was the first cohort widely exposed to the PERT in high school. Students who took the PERT in Grade 11 were assigned to CRS courses in Grade 12 if they scored below 113 in math or 104 in English. In college, students were assigned to lower DE in math if they scored below 96 and upper DE if they scored below 113; they were assigned to lower DE in English if they scored below 85 and to upper DE if they scored below 104. We omit students missing FCAT or PERT scores, high school GPAs, or demographic data.

We use both first PERT scores and highest PERT scores in our probits. High PERT scores earned early in high school may underestimate true ability at college enrollment. Including an indicator for when high scores were achieved should mitigate this source of bias. If scores on retests do not reflect student ability, including first scores will account for this potential bias. Not all students retested; approximately 28% of students with math and reading PERT scores had only tested once in each subject.

Cumulative high school GPAs (overall and by subject) were computed on a 4.0 scale. Due to limitations in the way high schools report transcript data to the state, less than 3% of grades included pluses or minuses. Because of this, we ignored plus and minus values, which may lower the predictive power of GPA. Outcome values are grades from students' first postsecondary courses in math and English.

Table 1 shows summary statistics of student characteristics. There were 151,391 students with demographic and GPA data who began Grade 11 in Florida public high schools during the 2011/12 school year. Of these students, 26.6% seamlessly enrolled in a 2-year college in Fall

2013. They were disproportionately female, non-White, free or reduced-price lunch (FRPL) eligible, and/or current or former English language learner (ELL) students with slightly higher grades than the overall cohort. Seamless enrollees who took both PERT subjects were like those who took either, as few took the test in just one subject. However, the share of math PERT takers placed in DE math was greater than the share of reading PERT takers placed in DE English. Mokher, Leeds, and Harris (2018) found no evidence that assignment to take the PERT or enroll in a CRS course affected college enrollment or performance for students near the college readiness cutoff.

[Insert Table 1]

Some students were dropped from the final samples because covariates perfectly predicted success or failure. Two small colleges that listed almost no students in DE math courses were also dropped from all analyses because their inclusion could harm the interpretability of our results. Final sample sizes were 29,924 for math and 29,018 for English.

Analytical Models

To address the research questions, we set up logit models to simulate placement policy and predict misplacement rates. We first used a model with one cutoff between DE and for-credit courses (following Scott-Clayton et al., 2014) to predict misplacement rates at each cutoff value. Then, to reflect Florida's two DE levels, we used a model with one cutoff between lower DE and upper DE and another between upper DE and for-credit courses. This model was designed to predict both the direction and degree of misplacement at each combination of cutoffs.

Cutoffs should minimize some combination of overplacement and underplacement. Scott-Clayton et al. (2014) proposed minimizing their unweighted sum, which they refer to as the SER. Lowering a placement cutoff will reduce underplacement but increase overplacement, while raising the cutoff will do the opposite. Minimizing the SER therefore means setting marginal overplacement equal to marginal underplacement.[1]

Focusing only on the difference between DE and for-credit courses ignores some types of misplacement. Since students in lower DE must complete more requirements than those in upper DE before taking for-credit courses, they are at greater risk of dropping out. Although some students may require the extra support provided in lower DE, those who do not require the additional support should not face the time and financial costs associated with taking an extra DE course. There are multiple ways in which students may experience greater harm from assignment to lower level DE courses instead of a single upper level DE course. For example, assignment to lower level DE courses can increase both cost and time to completion for students and can lead to greater discouragement, which may increase the risk of dropout or negatively impact course performance. Because of this, we examined cutoffs for both upper DE and for-credit courses.

Using a metric such as GPA percentile instead of the PERT may also lower the SER. More accurate placement metrics may reduce both overplacement and underplacement, a Pareto improvement over the current system. We considered both the SER with the DE placement rate held fixed and the minimum SER under each metric.[2] Doing this allowed us to separate the effects of a new assignment mechanism from the effects of using each optimally.

Double Cutoff Model

We proceed with this section first by briefly introducing the model that underpins our data. We then explain the numeric approach that we used to apply this model to our data. Figure 1 illustrates a hypothetical double cutoff model whereby vertical lines represent the math cutoffs used in Florida at the time of placement for the students in our sample. One cutoff was between lower DE and upper DE, and the second cutoff was between upper DE and for-credit courses.

The four functions plotted on the graph represent the probabilities of particular outcomes in particular course levels. For ease of exposition, Figure 1 treats these as continuous functions of students' PERT scores; we relaxed the continuous function assumption for our numeric estimates. Here, $f_1(x)$ represents the share of students at score x expected to earn an A or B in a for-credit course, $f_2(x)$ represents the share of students expected to earn an A or B in upper DE, $f_3(x)$ represents the share of students expected to pass a for-credit course, and $f_4(x)$ represents the share of students expected to pass upper DE.[3]

[Insert Figure 1]

Six regions in Figure 1 represent misplacement:

- Region C contains students placed in lower DE but predicted to get a B or better in for-credit courses. These students are *double underplaced*, as they are misplaced by two course levels.
- Region B contains students placed in lower DE but predicted to get a B or better in upper DE. These students are *single underplaced*.
- Region F contains students placed in upper DE but predicted to get a B or better in for-credit courses. These students are *single underplaced*.
- Region D contains students placed in upper DE predicted to fail those courses. These students are *single overplaced*.
- Region H contains students placed in for-credit courses predicted to fail those courses. These students are *single overplaced*.
- Region G contains students placed in for-credit courses predicted to fail upper DE. These students are *double overplaced*.

The SER is minimized by minimizing these 6 regions (see Appendix A for derivation); students in regions A, E, and I are accurately placed. The key insight provided by this model is when minimizing the SER—that is, if all misplacement is equally bad, regardless of degree or direction—then *marginal single underplacement should equal marginal single overplacement at each cutoff*. This is because, for example, lowering the cutoff for upper DE moves nearby students from region A into region D (increasing the SER) and from region B into region E (lowering the SER); however, those in region C would move to region F, where they would still be misplaced. Since some forms of misplacement may be worse than others, we expand the model later in this section; by doing so, we hope to illustrate both the implications of the current cutoffs in each subject and how they might be adjusted to fit a given set of policy preferences.

Placement accuracy. Because the probability of getting a particular grade in a certain course level may not be a smooth function of one's PERT score, we used numeric methods to compute the SER at each possible cutoff value. Although we ran regression analyses to do so, we were less interested in specific parameter estimates than in the predicted performance of students at each possible test score, GPA value, or other possible cutoff. By computing the expected SER at each possible cutoff value, we were able to select the one that minimized the SER.

We used a probit (following Scott-Clayton et al., 2014) to estimate the probability that a student received a B or better in or failed a given course; those who received a B or better clearly did not need additional preparation, while those who failed did.[4] Students predicted to receive Cs did not count toward either outcome. Unlike Scott-Clayton et al. (2014), we treated Ds as failure, as they rarely satisfy prerequisites or receive credit. We believe the decision to treat Ds as failure was more appropriate for the policy context in Florida, given that students must receive a minimum grade of C in the gateway math and English courses to fulfill the requirements for an associate degree. However, appendix D includes a sensitivity analysis of placement accuracy using a grade of D of passing instead of C.[5]

Probabilities were estimated as:

$$\Pr(Fail) = \Phi\big(\alpha_1 + (\boldsymbol{Test\ Scores})\beta_{1,1} + (\boldsymbol{Transcript})\beta_{2,1} + \boldsymbol{X}\beta_{3,1} + \varepsilon_1\big) \quad (1)$$

$$\Pr(B\ or\ Better) = \Phi\big(\alpha_2 + (\boldsymbol{Test\ Scores})\beta_{1,2} + (\boldsymbol{Transcript})\beta_{2,2} + \boldsymbol{X}\beta_{3,2} + \varepsilon_2\big) \quad (2)$$

$\boldsymbol{Test\ Scores}$ is a vector containing FCAT and PERT scores for both subjects and squared PERT scores in for-credit estimates, along with interactions of each score with race and gender.[6] Since students could retake the PERT, our preferred specification took full advantage of available information by using students' first and highest test scores. The preferred specification also included indicators for whether a student tested only once in either subject, the number of tests taken, an interaction between testing once and being below the for-credit cutoff, and whether students assigned to CRS courses complied with their assignment. Our analyses also explored how the SER varies depending on which placement test scores are used for students who retest. We compared results when using all test score variables to those including a single variable for: (a) first PERT score, (b) highest PERT score, or (c) last PERT score. If a student took the PERT only once, the same score was included in all three specifications. This allowed all students to remain in all specifications so that sample composition did not affect our results.

Given that Florida pursued a strategy of early assessment and implemented CRS courses to help more students test college-ready prior to college enrollment, the timing of the PERT scores may have important implications. It is possible that CRS courses and the opportunity to retest may have reduced certain types of placement errors that otherwise would have occurred if placement were based only on the first PERT score. Furthermore, score usage policy itself may have affected how students approached the PERT. For example, if students knew there was an option to retake the PERT, they may have been less concerned about performing well the first time. This means the first PERT scores that we observed in the data may have been lower than they would have been if students could not retest. Therefore, even though we could have used existing data to explore differences in predictive validity under alternative testing policies, we did not use that data and therefore our results did not perfectly account for behavioral effects of different policies on retesting and placement.

$\boldsymbol{Transcript}$ contains cumulative GPA, credits attempted, and number of courses failed, both overall and by subject. $\boldsymbol{X}$ contains indicators for race, gender, free or reduced-price lunch (FRPL) status, disability status, English language learner (ELL) status, native English speaker status, and district attended.[6]

Following our research questions, we sought to examine the extent to which students were accurately placed into DE courses based on their PERT scores and whether accuracy could have been improved by using different cutoffs, different metrics, or different sets of test scores for re-testers. We first computed the optimal SER using a single-cutoff model to show how our context differed from the one in Scott-Clayton et al. (2014) and to illustrate the implications of switching to a double-cutoff model. Scott-Clayton et al. computed the SER for a single cutoff between DE and for-credit (FC) courses as:

$$\begin{aligned} SER_1 = {} & [\Pr(B\ or\ better\ FC | Placed\ in\ DE) * \mathbf{1}(Placed\ in\ DE)] \\ & + [Pr(Fail\ FC | Placed\ in\ FC) * \mathbf{1}(Placed\ in\ FC)] \end{aligned} \tag{3}$$

We expanded on the Scott-Clayton et al. (2014) analysis by computing the SER when cutoffs separate lower DE, upper DE, and for-credit courses, since Florida used cutoffs for each level. We examined how the results changed when accounting for these cutoffs by computing:

$$
\begin{aligned}
SER_2 = {} & [\Pr(B\ or\ better\ in\ FC|Placed\ in\ UDE) * \mathbf{1}(Placed\ in\ UDE)] \\
& + [\Pr(B\ or\ better\ in\ FC|Placed\ in\ LDE) * \mathbf{1}(Placed\ in\ LDE)] \\
& + [\Pr(B\ or\ better\ in\ UDE|Placed\ in\ LDE) * \mathbf{1}(Placed\ in\ LDE)] \\
& + [Pr(Fail\ FC|Placed\ in\ FC) * \mathbf{1}(Placed\ in\ FC)] \\
& + [Pr(Fail\ UDE|Placed\ in\ FC) * \mathbf{1}(Placed\ in\ FC)] \\
& + [Pr(Fail\ UDE|Placed\ in\ UDE) * \mathbf{1}(Placed\ in\ UDE)]
\end{aligned} \tag{4}
$$

We used predicted values from probit regressions to compute SER_1 and SER_2. First, we assumed two hypothetical placement cutoffs, which sorted students into three course levels based on their placement metric. Placement into a specific course level determined whether the indicator terms in each of the six terms in Equation 4 were equal to 0 or 1. Our probit regressions then enabled us to estimate the probability that each student would get a B or better in each course level or a D or lower in each course level, corresponding to the probability in the first half of each term in Equation 4. The combination of implied placement and predicted values therefore enabled us to compute the SER across all students. By looking at every possible pair of cutoffs (such that the upper DE cutoff is no higher than the for-credit coursework cutoff), we computed every possible SER and identified the pair of cutoff values that minimized it.

We computed probabilities of success in each course level using students placed in that level, as compliers and noncompliers may be systematically different.[7] In both single and double cutoff models, probit regressions for performance in for-credit courses therefore contained only students who were both placed in and took those courses; parameter estimates were used to predict probabilities for all other students. In the double cutoff model, we estimated parameters for upper DE students using those who were placed in and took upper DE and then extrapolated to all other students. We focused on compliers for three reasons. First, most students complied with placement. Table 2 sorts seamless 2-year enrollees (those who enrolled in college by the fall after high school graduation) by placement and first-year enrollment. In both subjects, compliance was highest for students placed into for-credit courses (over 90%). Although fewer students took math than English, those who took math were more likely to comply with course assignment at each level.

[Insert Table 2]

Second, many students who did not comply with course placement likely had concordance scores on the SAT or ACT (not included in our data) that allowed them to enroll directly into for-credit courses regardless of their PERT scores. Concordance scores were subject to separate policies, which should be adjusted independently. However, the predictive validity of our model may have been weaker since we did not have the data to adjust for these scores. Third, we cannot say which observed compliers would have complied with new policies.

Relaxing model assumptions based on policymaker preferences. Policymakers should not minimize the SER if some forms of misplacement are perceived to be more harmful than others. Being double misplaced may result in greater discouragement and less learning than being single misplaced, and being overplaced could be worse than being underplaced (e.g., since failing a course has negative repercussions) or vice versa (e.g., if DE placement leads to lower persistence levels). Once these harm levels are known, placement cutoffs should be set to reflect them. A more flexible model can illustrate both how relative harms should affect policy and whether the placement cutoffs in our data reflected consistent policy preferences.

We therefore consider a model (formally presented in Appendix A.2) in which double misplacement (in either direction) is worse than single misplacement by amount $\omega \geq 0$ and overplacement is better or worse than underplacement by amount $\tau \gtrless 0$. Allowing for different levels of harm by misplacement type means setting marginal *harm from misplacement* on either side of each cutoff as equal rather than setting marginal misplacement as equal. Under existing cutoffs, we can use probabilities $f_1(\cdot)$ through $f_4(\cdot)$ to impute the preferences of a perfectly rational policymaker. We also show how cutoffs, remediation rates, and misplacement rates would respond to a range of parameter values.

Limitations

Our findings are subject to several limitations. First, our sample is limited to high school students who seamlessly enrolled in college. Many community colleges have large populations of non-traditional students who have been out of school for several years or longer before enrolling in college. The results for misplacement might be quite different depending on student characteristics such as age and number of years since high school graduation. Second, our sample is limited to students who attended a Florida state college. The results may not be generalizable to 4-year universities or other states, particularly if different placement tests are used that are not as closely aligned to state standards. Third, the analysis relies on out-of-sample predictions of course outcomes using data on students who complied with the placement policy and enrolled in college-level courses (for either model) or upper DE courses (for the double cutoff model). These models will be invalid if compliers poorly predict counterfactual course outcomes for students with other limitations. This is a common issue among other prior studies on this topic area of placement test accuracy.

In particular, extrapolation out of the sample could lead to biased estimates for students with other placements and result in their placement into courses too demanding or insufficiently challenging for their ability levels. We addressed this in two ways, following Scott-Clayton et al. (2014). First, although we did not have psychometric data, the standard deviation of PERT scores showed that at an exam reliability of 0.9 (on par with the SAT or ACT), the standard error of measurement was 4.6 in math and 5.1 in reading. Thus, if a student received a score of 104 in math, we could not put her with 95% confidence in *any* course level; similarly, if a student received a score of 95 in English, we could not with statistical confidence put him in any course level. If reliability was 0.7 (on par with a well-designed classroom exam), we would barely be confident that a student at the college readiness cutoff in either subject does not belong in lower DE. Although extrapolation far from each subject's course level cutoffs may therefore require a degree of caution, extrapolation near either cutoff is unlikely to introduce noticeable bias. Even if extrapolation far from course-level cutoffs were completely unreliable, results close to the cutoffs would be useful in telling us the direction in which each cutoff should be shifted. Therefore, to both show that results are not driven by extreme outliers and illustrate that the directions of our findings are unchanged, we also conducted analyses omitting the top and bottom 1% of PERT scores among seamless enrollees. Results of these analyses (available on request) were similar to those presented here.[8]

Results

We begin by graphing probabilities of success by course level and subject. Next, we discuss SER-minimizing metrics and cutoffs under a single cutoff model. We then present analogous results for a double cutoff model. We finish by discussing what actual cutoffs would reveal about perfectly rational policymakers and how policy preferences might affect outcomes.

Predicted Probabilities of Success and Failure by Course Level

Figure 2 shows predicted probabilities of course outcomes by PERT scores. In both subjects, probabilities increase noisily with PERT scores. Receiving a B or better in upper DE is generally more likely than in a for-credit course; the same holds for passing. However, some students are more likely to pass a for-credit course than get a B or better in upper DE, and very low-scoring students are more likely to pass a for-credit course than upper DE. These findings largely support our assumption that a given performance level will be harder to achieve in a for-credit course than in an upper DE course.

[Insert Figure 2]

Students with low math scores were unlikely to pass upper either DE or for-credit courses. Probabilities of success in upper DE increase rapidly and plateau at a probability of one, while probabilities in for-credit courses increase more gradually. Students barely assigned to upper DE in math had less than a 40% probability of passing; however, those who narrowly missed assignment to for-credit courses had nearly a 90% probability of passing upper DE and a 70% probability of receiving an A or B. Students narrowly assigned to for-credit courses had a 60% probability of passing and a 35% probability of receiving an A or B. In English, there were only two PERT scores at which students were more likely to fail for-credit courses than to pass. Nearly three quarters of students at the upper DE cutoff would have passed a for-credit English course, and half would have earned an A or B. Thus, many students were underplaced in English.

Single Cutoff Model

We first present results for a single cutoff model, in which policymakers do not differentiate between upper and lower DE. This both provides a baseline for the double cutoff model and allows us to determine whether differences between our results and the Scott-Clayton et al. (2014) study are attributable to the model or to policy contexts and data sets.

In Table 3, the columns show six placement metrics: highest PERT score, highest FCAT score, high school GPA within subject, overall high school GPA, all high school data (GPA, credits attempted, number of honors or AP courses taken, and number of courses failed, computed overall and by subject), and all high school data plus PERT scores. For the last two measures, grades in for-credit courses (on a 4.0 scale) were regressed on components of that measure and predicted out-of-sample, and students were sorted into percentiles.

[Insert Table 3]

Rows show the cutoff score or percentile, the percentage of students placed in DE, the percentage underplaced and overplaced, and the total SER. Underplacement is the probability of both getting a B or better if placed in a for-credit course and being placed in DE. Overplacement is the probability of both failing a for-credit course if placed in one and being placed in one. Total SER is the sum of the probabilities for underplacement and overplacement. Table 3 is divided into four panels. The top panel shows placement accuracy in math, holding the remediation rate constant; column (1) represents the status quo.[9] This is how FLDOE first set PERT cutoffs; it also allows us to separate the effect of switching placement metrics from that of optimizing cutoff locations. However, we rarely perfectly match the DE placement rate using discrete scores and percentiles. For example, 57.8% of our sample was placed in DE math based on PERT scores, but the closest rate using FCAT scores is 58.4%. The second panel shows outcomes at SER-minimizing cutoffs under each metric; if students are distributed differently over each metric, keeping the same placement rates will not make sense. The third and fourth panels are analogous to the first two but show outcomes for English.

In all four panels, the SER falls consistently from column (2) to column (6), with most of the changes due to a decrease in both overplacement and underplacement. There are consistent patterns across all metrics. Holding DE placement rates fixed, students are more likely to be overplaced in math (between 11.0 and 15.0% depending on the metric) and underplaced in English (between 13.4 and 18.6%). Minimizing the SER means raising the math cutoff and lowering the reading cutoff. In math, the optimal cutoff improves the SER by 1 percentage point or less, while in English there are improvements up to 6 percentage points. With optimal cutoffs, students are more likely to be underplaced in math (between 7.6 and 10.1%) and overplaced in English (between 13.9 and 25.1%).

In both subjects, the optimal PERT cutoff is less accurate than a suboptimal GPA cutoff. The optimal PERT cutoff in math has a total SER of 23.9%, which is approximately 2 percentage points higher than the SER of 22.0 for overall GPA holding DE rates fixed. Results are similar in English, where the optimal PERT cutoff has a total SER of 26.2 compared to 23.6% for overall GPA at the fixed DE rate. However, accuracy gains from switching to high school GPA are much greater in English than in math. This could be because students take a wider range of core math courses than English courses, even controlling for honors status. Therefore, while all metrics can predict postsecondary performance more accurately in math than in English, there may be particular gains from using classroom-based measures rather than standardized assessments in English. We show these results graphically in Appendix Figure B.1.

In both our context and that of Scott-Clayton et al. (2014), approximately a quarter of students were misplaced in math and a third were misplaced in English. However, DE placement rates for their samples were frequently much larger than ours. Since our sample was enrolled in DE at much lower rates than theirs, optimal adjustments might be quite different across contexts.

Switching from test-based placement to GPA-based placement may have smaller effects on misplacement in our study for several reasons. First, our GPA variables were limited to categorical grade variables (e.g., A, B, or C), since the Florida transcript data were missing plus or minus values (e.g., A-, B+) for most records. This means there was less variation available in grades used to predict student outcomes relative to the Scott-Clayton et al. study (which had high school grade data on a 0 to 100 grading scale). Second, all Florida colleges used the PERT for the college placement test, while the colleges in the Scott-Clayton et al. sample used Accuplacer and Compass. The PERT was created specifically to align with Florida's state standards, so it may also be better aligned with college courses in Florida than national placement tests. Third, our sample is limited to high school students who seamlessly enrolled in community college, while the Scott-Clayton et al. sample includes a mix of recent high school graduates and older students. This is important because there is some evidence that the predictive validity of Accuplacer may vary by student age (Cole, Muenz, & Bates, 1998), so the results may not be comparable among students with different characteristics like age.

However, the primary reason why our results differ from Scott-Clayton et al. (2014) is likely the double cutoff model itself. As an extreme example, consider a single-cutoff model in which the cutoff between any DE and for-credit coursework is properly set, but in which all students below the cutoff are incorrectly assigned to upper DE instead of lower DE. A single cutoff model will see no misplacement and will instead assume that improvement can be made only through changing the placement metric. However, in a double cutoff model, changing the placement metric while keeping placement proportions constant will still misplace many students; the only solution to misplacement within DE would be to adjust the cutoffs themselves.

Double Cutoff Model—Minimizing the SER

Results using two cutoffs are shown in Table 4 (math) and Table 5 (English). Each is divided into two panels; the top panel shows the effect of switching metrics while keeping upper and lower DE rates fixed, and the bottom panel shows the effect of minimizing the SER under each metric. Minimizing the SER treats all misplacement as equally bad, regardless of level (single or double) or direction (underplacement or overplacement).

[Insert Table 4]

Because Table 4 has two margins for misplacement, there is more misplacement than in a single cutoff model. Single overplacement is substantially larger than the other three forms of misplacement combined with values up to 27.6% in math and 16.1% in English. These values are greater than the total SER for each metric in Table 3, which only go up to 15.0% in math and 13.4% in English holding DE rates fixed. Similarly, the total SER falls from column (2) to column (6). Holding DE placement rates constant, switching to overall GPA would reduce the SER only slightly (from 42.1 to 40.6 in math). Incorporating high school transcript data further improves placement accuracy to 37.9%.

In the bottom of Table 4, the SER is minimized by increasing the upper DE cutoff and lowering the for-credit cutoff, often setting the two equal. For example, the optimal cutoff score for both levels would be 107 for PERT math and 325 for FCAT math. Optimal cutoffs using all high school data plus PERT scores are not equal but still reduce upper DE enrollment by over 90%. Thus, a placement system designed to minimize the SER would either eliminate placement into upper DE or come very close to doing so—while the course level itself would not be eliminated, completing lower DE would be a prerequisite for upper DE. While some students would certainly be best placed in upper DE, too many students at the low end of its placement range are expected to fail and too many at the high end of its placement range could perform well in for-credit math. The average student at every possible score in five of six metrics would therefore be more accurately placed elsewhere. While PERT scores for our sample are not perfectly normal distributions, there is no evidence of a bimodal distribution that would naturally sort students out of upper DE. It is possible to envision scenarios in which this is attributable to course sequencing (e.g., if the upper DE course is nearly as challenging as the for-credit course) or staffing decisions (e.g., if the best teachers are assigned to the most able students and the most needy students, with students in upper DE falling into neither category); however, these scenarios are purely hypothetical and their analysis is beyond the scope of this work.

When holding the DE placement rate fixed in English in Table 5, misplacement is balanced between single overplacement and single underplacement. Again, the total SER falls consistently by column from 38.9% in column (2) to 30.9% in column (6). Unlike in math, switching to overall GPA greatly reduces the SER (from 39.6% to 32.6).

[Insert Table 5]

The SER-minimizing upper DE and for-credit cutoffs are again equal across all six metrics, but lower DE placement rates vary from 0.6% under PERT to 12.0% using high school data plus PERT. The total SER falls by only 2 percentage points across columns; however, unlike in a single-cutoff model, the optimal SER in English is lower than in math under each metric. Appendix Figure B2 depicts SER values in a two-cutoff framework using contour graphs.

Using the full range of values (available upon request), we can consider other counterfactual policies. For example, recent changes under Florida's Senate Bill 1720 both made DE optional for all recent high school graduates and eliminated requirements for PERT testing upon college entry. Laws pushing the vast majority of students into for-credit courses are predicted to contribute to higher misplacement rates in math (rising to 47.2% if no students enrolled in DE) but predicted to greatly reduce misplacement in English (to 27.1% if no students enrolled in DE, only 0.1% from the minimum SER value). We cannot directly compare our simulation results to actual trends under the reform because DE was optional rather than completely eliminated, and our model does not account for complications such as selection bias. However, our findings are consistent with observed trends in performance under the DE reform, as described by Hu et al. (2016). Our simulations indicate that pass rates would decline in for-credit courses, particularly in math. Following the reform, course-based passing rates in for-credit courses did fall, conditional on enrollment; the magnitude was negligible in English and more pronounced in math, as predicted by our model. Yet the cohort-based passing rate for the percentage of students in the full cohort who both enrolled in and passed for-credit courses in the first semester increased after the reform, which indicates that some students who would have been underplaced into DE courses were able to succeed in for-credit courses.

We also find—unlike our single-cutoff results—that switching to GPA as a placement metric without adjusting cutoff levels is not as effective as setting optimal cutoffs for the PERT. For example, the total SER for the optimal PERT cutoff is 27.0%, which is considerably lower than the total SER of 32.6% for overall GPA when DE rates are held fixed. Therefore, policymakers may wish to investigate whether cutoffs on placement assessments are optimal before embarking on the more challenging task of switching placement metrics.

Our analyses also explored how the results of the predictive models differed depending on the timing of when the PERT was taken. Table 6 shows placement cutoffs and total SERs under a double cutoff model (with no additional covariates for high school performance) for models that include: (a) all PERT scores (from our preferred specification), (b) first PERT score only, (c) last PERT score only, and (d) highest PERT score only. We find that the total SER using all PERT scores is 33.0% in math and 27.0% in English. The total SER is very similar (within 2 percentage points) across all sets of results using different timing of PERT scores in each subject area. This suggests that using placement scores at different points in time has little impact on the accuracy of our predictions. When looking at the placement cutoffs for upper DE and for-credit courses, results are almost identical among the models using all PERT scores and highest PERT scores with cut scores of 107 in math and 65 in English. This suggests that once we know a student's highest test score, there is not much influence of other factors such as when the test was taken or the number of attempts that it took for the student to achieve the score. However, there are some differences in the results when first PERT scores are used. One possible explanation is that some students who take the PERT for the first time will receive scores that underestimate their true ability levels and will therefore benefit from retesting. This makes it more difficult to differentiate high-ability and low-ability students with the same PERT score. However, once students have retested, higher ability students tend to achieve higher scores, which leads to an increase in the optimal PERT cutoffs.

[Insert Table 6]

Double Cutoff—Implications for Policymakers

Minimizing the SER will not be optimal if misplacement types are perceived to be differentially harmful. We now assume that policymakers base their decision-making on these harm levels. Specifically, double misplacement causes an additional $\omega > 0$ harm relative to single misplacement, and overplacement causes an additional $\tau \gtreqless 0$ harm relative to underplacement. We can use the modified double cutoff model in Appendix A.2 to solve for these values using the actual cutoffs.

The cutoffs in our data show that if policymakers were indeed minimizing a weighted version of the SER, they would regard double misplacement as similarly harmful for each subject. Double misplacement would be over 2.5 times more harmful than single misplacement in math ($\omega_M = 1.6070$) and over 2.3 times more harmful in English ($\omega_E = 1.3325$). They would consider overplacement to be less than half as harmful than underplacement in math

($\tau_M = -0.5643$) but nearly 2.5 times as harmful in English ($\tau_E = 1.4620$). This would mean that it would be worse to place a student one level too high in English than two levels too low.

While optimal values for τ and ω are beyond the scope of this paper, we must still consider what factors might lead to such different values of τ for each subject.[10] Three possibilities seem likely. First, rational policymakers might consider it more crucial to guarantee a baseline ability in English than in math; for example, if they believe skills learned in English courses will be more valuable to students in future courses or careers. Second, they may be optimizing criteria other than misplacement rates. Third, they may not be optimizing properly. Here, the second and third hypotheses are most likely, as cutoffs on the PERT were set to match remediation rates on the prior placement exam, rather than to achieve a particular outcome.

But even if policymakers know their preferences for prioritizing different forms of misplacement, they may not know how to put those preferences into effect. To provide guidance on how this might work, Figure 3 and 4 show how the PERT cutoffs and placement rates respond to a range of τ and ω values in math and English. In both figures, τ ranges from -1 (single overplacement is harmless; double overplacement is only a problem due to ω) to 1 (single overplacement is twice as bad as single underplacement). We examine two cases for ω – one in which ω equals zero (double misplacement is no worse than single misplacement) and one in which ω equals 1.5 (double misplacement is 2.5 times worse).

[Insert Figures 3 and 4]

For low enough values of τ, both DE cutoffs in each subject will be equal to 50—if single overplacement is harmless, eliminating one type of underplacement is rational. As τ rises (and overplacement becomes more harmful), both cutoffs rise. In math, both cutoffs increase sharply as τ first begins to increase, but then increase only gradually. This dovetails with evidence from

Figure 2 that students with low math PERT scores have extremely low probabilities of passing upper DE, and even lower of passing for-credit math. Even if any single instance of overplacement caused little harm, the amount of overplacement would make low cutoff scores untenable. In English, however, both cutoffs stay at 50 substantially longer, as many students with extremely low English PERT scores are capable of passing either course level. As ω rises, the cutoffs move farther apart; to take an extreme case, if we had ω = 10,000, cutoffs would be at 50 and 150 because double misplacement is impossible in upper DE. See Appendix C, Table C1 for sample values of τ and ω and how they affect cutoffs, placement, and misplacement rates.

Conclusion

The Florida College and Career Readiness Initiative mandated that students take the Postsecondary Education Readiness Test (PERT) to determine placement into College Readiness and Success courses in Grade 12. The PERT was also used in college to sort students into lower level DE, upper level DE, and for-credit courses. In principle, the FCCRI could have signaled and promoted college readiness; however, it relied on the PERT's accuracy in course placement. We build on an existing body of work examining the extent to which students are accurately placed into DE courses based on their placement test scores. As with prior studies, these results are contingent upon the extent to which out-of-sample predictions using compliers can be used to predict counterfactual course outcomes for students with another placement.

Like Scott-Clayton et al. (2014), we find that using high school data improves placement accuracy. Unlike in the Scott-Clayton et al. study, setting optimal cutoffs on the PERT improves accuracy more than selecting a new metric, while holding remediation rates fixed. This finding suggests that policymakers should lend caution to switching placement metrics to improve placement decisions. For example, California's recent legislation under AB 705 requires community colleges to make placement decisions primarily using indicators of high school performance. This policy was based on evidence suggesting that too many students were being underplaced into DE courses based on placement test scores (California Community Colleges, 2018). However, policymakers in states considering these types of changes should also consider that adjusting existing placement score cutoffs may lead to greater placement accuracy than using suboptimal cutoffs on a theoretically better metric.

In addition to considering the accuracy of placement metrics, policymakers should also consider how feasible it would be to implement different placement procedures. Further optimizing cutoff scores could be easier to implement since placement policies using multiple measures may face challenges in getting students to provide transcripts in a standardized manner and may pose substantial administrative burdens on both students and college staff. Using multiple measures could also create confusing signals of college readiness for students if placement is based on numerous indicators rather than a single test score cutoff and would likely make advising more complicated. Although it may be difficult for institutional researchers or state agency staff to replicate our procedures for adjusting cutoffs to minimize placement error if they lack the statistical skills or institutional bandwidth to implement predictive analytics.

In Florida the optimal cutoffs also imply that few, if any, students should be sorted into upper DE based on placement metrics; most are either already prepared to succeed in for-credit courses or require more intensive remediation through two levels of DE. These findings tend to be consistent regardless of whether students' first PERT scores, last scores, highest scores, or a combination of scores is used to predict the likelihood of course success. While implementing this policy is beyond the scope of our work, using an opt-in system for upper DE might prevent misplacement while allowing students freedom of course selection.

Recent legislative changes no longer require that students take the PERT in high school or upon college enrollment; instead, college advisors are instructed to take a holistic view of high school performance. The legislation also made DE optional for recent high school graduates, while requiring that colleges provide academic support to mitigate overplacement. While exempt students (those who enrolled in a public high school during or after 2003/04, enlisted military personnel, and veterans) are no longer required to take the PERT upon college enrollment, they are still advised to do so to inform their course selections. Additionally, all Florida state colleges still require that nonexempt students take the PERT to determine course placement. The reform also requires colleges to redesign the delivery methods for DE courses, but these courses still do not count for college credit. If implemented properly, these policies could improve student outcomes by reducing underplacement; but at worst, they could exacerbate both forms of misplacement. Early descriptive research indicates that more students are passing for credit courses in the first semester, but passing rates are declining within for-credit courses, indicating that not all students are prepared to succeed (Hu et al., 2016). Longer term implications of these changes remain unknown.

One direction for future research is to consider the lifetime costs of misplacement. Overplaced students have to retake courses and may lose financial aid. Underplaced students pay for courses that they do not need, and financial aid often will not cover DE. Either form of misplacement may increase the time to degree or probability of dropping out of college. Information about the costs of misplacement could be used to inform priorities for the direction and degree of misplacement. Additional analyses might also explore whether adjusting cutoffs would affect enrollment, employment, tuition, or revenue at 2-year colleges.

Another direction for future research is to identify additional predictors of student success. For example, though high school GPA does not depend on specific math or English courses taken, a student with a B average in math through AP Calculus will likely be better prepared than one with an A average through Algebra II. Diagnostic tests may also place students more accurately than computer-adaptive tests, although they may be more expensive and time-consuming to administer (e.g., Ngo & Melguizo, 2015). Additionally, the optimal cutoffs based on the SER could be compared to the cutoffs determined using a regression discontinuity framework (as described by Melguizo, Bos, Ngo, Mills, & Prather, 2016) to determine which method works best. However, many students will remain misplaced even under optimal metrics and cutoffs, as academic mismatch is only one reason why students do not succeed in college. Some may not want to put in the effort to complete assignments or attend class. Others may fail due to outside factors such as work commitments or financial difficulties. Future research may also examine the extent to which students fail due to lack of preparation, lack of effort, or external circumstances, and examine policies for each of these causes.

Endnotes

1. Optimal cutoffs may be different ex ante and ex post or from year to year, a challenge for stable placement systems. However, in a state as large as Florida, optimal cutoffs are unlikely to vary substantially from year to year.
2. If other metrics cannot perfectly match the current DE rate, we either use the closest DE rate (in a one-cutoff model) or minimize the sum of squared differences from each level of DE (in a two-cutoff model).
3. See Appendix A for further details on the model that underpins our data.
4. We use a probit rather than a linear probability model because we use the entire domain of each chosen metric to compute the SER. A linear probability model may be preferable when the goal is to estimate a specific parameter, but may also return probabilities outside of [0,1] at extreme values of the metric being evaluated, making interpretation of the resulting SER extremely challenging.
5. If we instead treated a D as a passing grade, it would be harder to overplace students, particularly in English. As a result, the SER would fall and cutoffs would be set at lower scores—if it is harder for overplacement to occur, then curbing underplacement would become a greater priority.
6. Single cutoff models included indicators for college attended, but in double cutoff models these perfectly predicted too many successes or failures to be used.
7. We did impute grades for some noncompliers. Students who received an A or B in a higher course level than assigned were assumed to have received an A or B in their assigned level, and students who failed in a lower level than assigned were assumed to have failed in the assigned level.

8. Other attempted solutions did not produce empirically meaningful results. Attempting a standard Heckman correction with a linear probability model in the second stage, where a series of dummy variables indicating placement in CRS courses and compliance with that placement supplied the exclusion restriction, resulted in negative predicted pass rates at low PERT scores. Re-running the Heckman correction with probits in both the first and second stages solved this issue but results of this model implied that incorporating additional high school data on top of PERT scores reduced placement accuracy in math, which seems difficult to believe. It may be that clustering of the dependent variable therefore interferes with Heckman correction. Using students with below-average PERT scores in each course level as our estimation sample resulted in negative single underplacement values, which has no meaningful interpretation.
9. If multiple cutoff values matched the observed remediation rate, we used the highest one.
10. For a policymaker, determining optimal weights would involve computing the effect of each (mis)placement type on the probability of earning a degree, time to degree, lifetime benefits of a degree, college costs, and other factors, then computing lifetime values. Policymakers may have additional concerns—for example, universities may care about effects on their rankings, while states may want to meet enrollment or placement benchmarks. Researchers would therefore have to determine the optimal Nash equilibrium under competing priorities.

References

Atkinson, R. C., & Geiser, S. (2009). Reflections on a century of college admissions tests. *Educational Researcher, 38*(9), 665-676.

Mokher, C.G., Leeds, D., & Harris J. (2018). Adding it up: How the Florida College and Career Readiness Initiative impacted developmental education. *Educational Evaluation & Policy Analysis, 40*(2):219-242. doi.org/10.3102/0162373717744891

Belfield, C., & Crosta, P. (2012). *Predicting success in college: The importance of placement tests and high school transcripts* (CCRC Working Paper No. 42). New York, NY: Columbia University, Teachers College, Community College Research Center.

Boatman, A., & Long, B. T. (2018). Does remediation work for all students? How the effects of postsecondary remedial and developmental courses vary by level of academic preparation. *Educational Evaluation and Policy Analysis*, *40*(1), 29-58.

Boser, U., & Burd, S. (2009). *Bridging the gap: How to strengthen the PK-16 pipeline to improve college readiness.* Washington, DC: New America Foundation.

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M.T., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research, 86*(4), 803-848.

California Community Colleges (2018). *What is AB 705?* Retrieved from https://assessment.cccco.edu/ab-705-implementation/

Cole, J. C., Muenz, T. A., & Bates, H. G. (1998). Age in correlations between ACCUPLACER's reading comprehension subtest and GPA. *Perceptual and Motor Skills, 86*(3), 1251-1256.

Education Commission of the States. (2016). *Moving from single to multiple measures for college course placement.* Retrieved from https://www.ecs.org/moving-from-single-to-multiple-measures-for-college-course-placement/

Hu, S., Park, T. J., Woods, C. S., Tandberg, D. A., Richard, K., & Hankerson, D. (2016). *Investigating developmental and college-level course enrollment and passing before and after Florida's developmental education reform* (REL 2017–203). Retrieved from the Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance website: https://ies.ed.gov/ncee/edlabs/regions/southeast/pdf/REL_2017203.pdf

Kirst, M. W. (2005). *Separation of K-12 and postsecondary education governance and policymaking: Evolution and impact* (Report No. 17). Stanford, CA: Stanford University, The Bridge Project.

McCormick, N., & Lucas M. (2011). Exploring mathematics college readiness in the United States. *Current Issues in Education, 14*(1), 3-28.

Melguizo, T., Bos, J. M., Ngo, F., Mills, N., & Prather, G. (2016). Using a regression discontinuity design to estimate the impact of placement decisions in developmental math. *Research in Higher Education, 57*(2), 123-151.

Melguizo, T., Kosiewicz, H., Prather, G., & Bos, J. (2014). How are community college students assessed and placed in developmental math? Grounding our understanding in reality. *Journal of Higher Education, 85*(5), 691-722.

National Commission on the High School Senior Year. (2001, January). *The lost opportunity of senior year: Finding a better way.* Princeton, NJ: The Woodrow Wilson National Fellowship Foundation.

Ngo, F., & Melguizo, T. (2015). How can placement policy improve math remediation outcomes? Evidence from experimentation in community colleges. *Educational Evaluation and Policy Analysis, 38*(1), 171-196.

Scott-Clayton, J. (2012). *Do high-stakes placement exams predict college success?* Retrieved from the Columbia University, Community College Research Center website: https://ccrc.tc.columbia.edu/publications/high-stakes-placement-exams-predict.html

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

Valentine, J. C., Konstantopoulos, S., & Goldrick-Rab, S. (2017). What happens to students placed into developmental education? A meta-analysis of regression discontinuity studies. *Review of Educational Research, 87*(4) 806-833.

Tables

Table 1. Summary statistics of student characteristics by sample.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sample	Entire cohort	Seamless enrollees	Math PERT takers	Reading PERT takers	Both PERT takers	Math sample	Reading sample
Percent female	50.9%	55.2%	55.5%	54.3%	54.7%	54.7%	54.6%
Percent non-White	53.4%	57.5%	57.9%	58.5%	58.9%	60.0%	60.9%
Percent FRPL	58.8%	60.6%	61.0%	61.5%	62.1%	62.7%	63.1%
Percent ever ELL	19.8%	23.6%	23.7%	24.1%	24.3%	25.4%	26.0%
Percent special education	11.3%	10.4%	10.5%	10.9%	11.0%	11.0%	11.2%
Cumulative GPA	2.78 (0.67)	2.98 (0.44)	2.97 (0.44)	2.96 (0.44)	2.95 (0.44)	2.94 (0.44)	2.93 (0.43)
Math PERT score	-	-	109.64 (11.74)	-	109.37 (11.75)	109.32 (11.75)	109.12 (11.74)
Reading PERT Score	-	-	-	104.94 (13.48)	104.85 (13.47)	104.76 (13.45)	104.51 (13.38)
Percent placed in lower DE	-	-	11.2%	6.2%	-	11.7%	6.4%
Percent placed in upper DE	-	-	45.5%	36.3%	-	46.1%	37.4%
Percent placed in for-credit	-	-	43.2%	57.5%	-	42.2%	56.2%
Percent placed in any DE in either subject	-	-	-	-	69.8%	70.1%	70.9%
Percent placed in any DE in both subjects	-	-	-	-	30.4%	30.8%	31.4%
N	151,391	40,227	38,927	33,688	32,389	29,924	29,018

Table 2. Student course placement compared to actual courses taken in math (top) and English (bottom).

Math placement versus enrollment

		Total Placed	Any Course	Among Course Takers		
				Lower DE	Upper DE	For-Credit
Course Placement	Lower DE	N = 4,337	85.58% (N = 3,746)	72.21% (N = 2,705)	7.53% (N = 282)	20.26% (N = 759)
	Upper DE	N = 17,723	90.99% (N = 16,126)	8.40% (N = 1,355)	53.32% (N = 8,599)	38.27% (N = 6,172)
	For-Credit	N = 16,827	90.72% (N = 15,265)	0.33% (N = 51)	2.18% (N = 333)	97.48% (N = 14,881)
	No Score	N = 1,300	89.77% (N = 1,167)	1.03% (N = 12)	2.06% (N = 24)	96.92% (N = 1,131)
	Total	N = 40,227	90.25% (N = 36,304)	11.36% (N = 4,123)	25.45% (N = 9,238)	63.20% (N = 22,943)

English placement versus enrollment

		Total Placed	Any Course	Among Course Takers		
				Lower DE	Upper DE	For-Credit
Course Placement	Lower DE	N = 2,079	89.13% (N = 1,853)	53.91% (N = 999)	14.79% (N = 274)	31.30% (N = 580)
	Upper DE	N = 12,222	91.99% (N = 11,243)	10.12% (N = 1,138)	46.18% (N = 5,192)	43.70% (N = 4,913)
	For-Credit	N = 19,387	93.49% (N = 18,125)	0.94% (N = 171)	6.05% (N = 1,096)	93.01% (N = 16,858)
	No Score	N = 6,539	92.23% (N = 6,031)	5.02% (N = 303)	7.86% (N = 474)	87.12% (N = 5,254)
	Total	N = 40,227	92.60% (N = 37,252)	7.01% (N = 2,611)	18.89% (N = 7,036)	74.10% (N = 27,605)

Table 3. Results for minimizing the SER using different placement metrics under a single cutoff model in math (top) and English (bottom).

	(1)	(2)	(3)	(4)	(5)	(6)
Placement metric	PERT	FCAT	Subject GPA	Overall GPA	All HS Data	All HS + PERT
			Math–holding DE rate fixed			
Score/percentile on metric	113	337	59	59	58	58
% placed in DE	57.8%	58.4%	57.5%	57.8%	58.0%	58.0%
Underplacement rate	10.6%	12.6%	10.4%	10.0%	9.2%	8.7%
Overplacement rate	13.5%	15.0%	13.2%	12.0%	11.4%	11.0%
Total SER	24.2%	27.6%	23.5%	22.0%	20.6%	19.7%
			Math–optimal cutoffs			
Score/percentile on metric	115	348	71	70	67	66
% placed in DE	66.7%	76.5%	70.7%	69.5%	67.0%	66.0%
Underplacement rate	13.8%	18.9%	14.9%	13.9%	12.2%	11.4%
Overplacement rate	10.1%	7.6%	7.9%	7.6%	7.9%	7.9%
Total SER	23.9%	26.5%	22.8%	21.5%	20.2%	19.3%
			English–holding DE rate fixed			
Score/percentile on metric	104	312	45	43	43	43
% placed in DE	43.2%	43.2%	43.6%	42.9%	43.0%	43.0%
Underplacement rate	18.5%	18.6%	15.3%	14.1%	13.5%	13.4%
Overplacement rate	13.4%	13.4%	10.0%	9.5%	8.9%	8.9%
Total SER	31.4%	32.0%	25.3%	23.6%	22.4%	22.3%
			English–optimal cutoffs			
Score/percentile on metric	82	262	25	25	25	27
% placed in DE	4.3%	5.6%	21.1%	24.5%	25.0%	27.0%
Underplacement rate	1.1%	1.7%	5.9%	6.5%	6.1%	6.7%
Overplacement rate	25.1%	24.6%	17.2%	15.2%	14.6%	13.9%
Total SER	26.2%	26.3%	23.0%	21.7%	20.7%	20.6%

Note: Results are based on probit models estimating the probability that a student received a B or better in or failed a given course using different placement metrics. Panels holding the DE rate fixed use the score or percentile cutoff that keeps the DE placement rate as close to the current rate as possible. Panels using optimal cutoffs use the score or percentile cutoff values under each placement metric that minimize the SER.

Table 4. Minimizing the math SER using different placement metrics under a double cutoff model while holding DE rates fixed (top) and estimating optimal cutoffs (bottom).

	(1)	(2)	(3)	(4)	(5)	(6)
Placement metric	PERT	FCAT	Subject GPA	Overall GPA	All HS data	All HS + PERT
	Holding DE rates fixed					
Upper DE cutoff score	96	306	13	12	12	12
For-credit cutoff score	113	337	59	59	58	58
Lower DE placement rate	11.7%	12.2%	11.5%	12.0%	12.0%	12.0%
Upper DE placement rate	46.1%	46.2%	46.0%	45.9%	46.0%	46.0%
Double underplacement rate	0.7%	1.4%	0.9%	0.8%	0.7%	0.5%
Single underplacement rate	10.4%	12.0%	10.9%	10.5%	9.8%	8.9%
Single overplacement rate	27.6%	26.9%	25.3%	24.9%	25.3%	25.5%
Double overplacement rate	3.3%	4.9%	5.2%	4.4%	3.8%	3.0%
Total SER	42.1%	45.2%	42.3%	40.6%	39.5%	37.9%
	Optimal Cutoffs					
Upper DE cutoff score	107	325	44	42	42	40
For-credit cutoff score	107	325	44	42	42	44
Lower DE placement rate	37.9%	37.9%	40.7%	41.6%	42.0%	40.0%
Upper DE placement rate	0.0%	0.0%	0.0%	0.0%	0.0%	4.0%
Double underplacement rate	5.0%	6.8%	5.8%	5.8%	5.0%	4.2%
Single underplacement rate	5.5%	5.9%	8.6%	8.6%	8.5%	7.9%
Single overplacement rate	14.2%	14.0%	11.4%	11.4%	11.1%	12.4%
Double overplacement rate	8.3%	10.6%	9.8%	9.8%	7.9%	6.2%
Total SER	33.0%	37.3%	35.7%	35.7%	32.5%	30.7%

Note: Results are based on probit models estimating the probability that a student received a B or better in or failed a given course using different placement metrics. Panels holding DE rates fixed use the scores or percentile cutoffs that keeps both DE placement rates as close to the current rates as possible. Panels using optimal cutoffs use scores or percentile cutoff values under each placement metric that minimize the SER.

Table 5. Minimizing the English SER using different placement metrics under a double cutoff model while holding DE rates fixed (top) and estimating optimal cutoffs (bottom).

	(1)	(2)	(3)	(4)	(5)	(6)
Placement metric	PERT	FCAT	Subject GPA	Overall GPA	All HS Data	All HS + PERT
	Holding DE rates fixed					
Upper DE cutoff score	85	265	7	7	6	6
For-credit cutoff score	104	312	46	45	43	43
Lower DE placement rate	6.4%	6.3%	7.0%	7.0%	6.0%	6.0%
Upper DE placement rate	37.4%	37.5%	37.2%	37.8%	37.0%	37.0%
Double underplacement rate	1.8%	1.9%	1.5%	1.3%	0.9%	0.9%
Single underplacement rate	17.9%	17.3%	15.9%	15.0%	13.5%	13.4%
Single overplacement rate	16.1%	15.9%	13.5%	13.8%	14.3%	14.4%
Double overplacement rate	3.8%	3.8%	3.3%	2.4%	2.3%	2.2%
Total SER	39.6%	38.9%	34.3%	32.6%	31.1%	30.9%
	Optimal Cutoffs					
Upper DE cutoff score	65	241	6	10	12	12
For-credit cutoff score	65	241	6	10	12	12
Lower DE placement rate	0.6%	2.5%	3.4%	9.7%	12.0%	12.0%
Upper DE placement rate	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Double underplacement rate	0.1%	0.7%	0.6%	2.0%	2.2%	2.2%
Single underplacement rate	0.1%	0.1%	0.9%	2.1%	2.6%	2.4%
Single overplacement rate	14.7%	14.9%	13.9%	13.0%	12.5%	12.7%
Double overplacement rate	12.2%	11.2%	11.2%	8.7%	7.8%	7.7%
Total SER	27.0%	26.8%	26.6%	25.7%	25.2%	25.0%

Note: Results are based on probit models estimating the probability that a student received a B or better in or failed a given course using different placement metrics. Panels holding DE rates fixed use the scores or percentile cutoffs that keeps both DE placement rates as close to the current rates as possible. Panels using optimal cutoffs use scores or percentile cutoff values under each placement metric that minimize the SER.

Table 6. Minimizing the SER under a double cutoff model with optimal cutoffs using different timing of PERT exams, by subject area.

	(1)	(2)	(3)	(4)
Timing of PERT exam	All PERT	First PERT	Last PERT	Highest PERT
	Math			
Upper DE cutoff score	107	100	107	107
For-credit cutoff score	107	104	107	107
Total SER	33.0%	32.4%	31.8%	32.3%
	English			
Upper DE cutoff score	65	62	76	65
For-credit cutoff score	65	62	76	65
Total SER	27.0%	24.5%	26.5%	26.7%

Note: Results are based on probit models estimating the probability that a student received a B or better in or failed a given course using different timing of PERT exams

Figure Captions

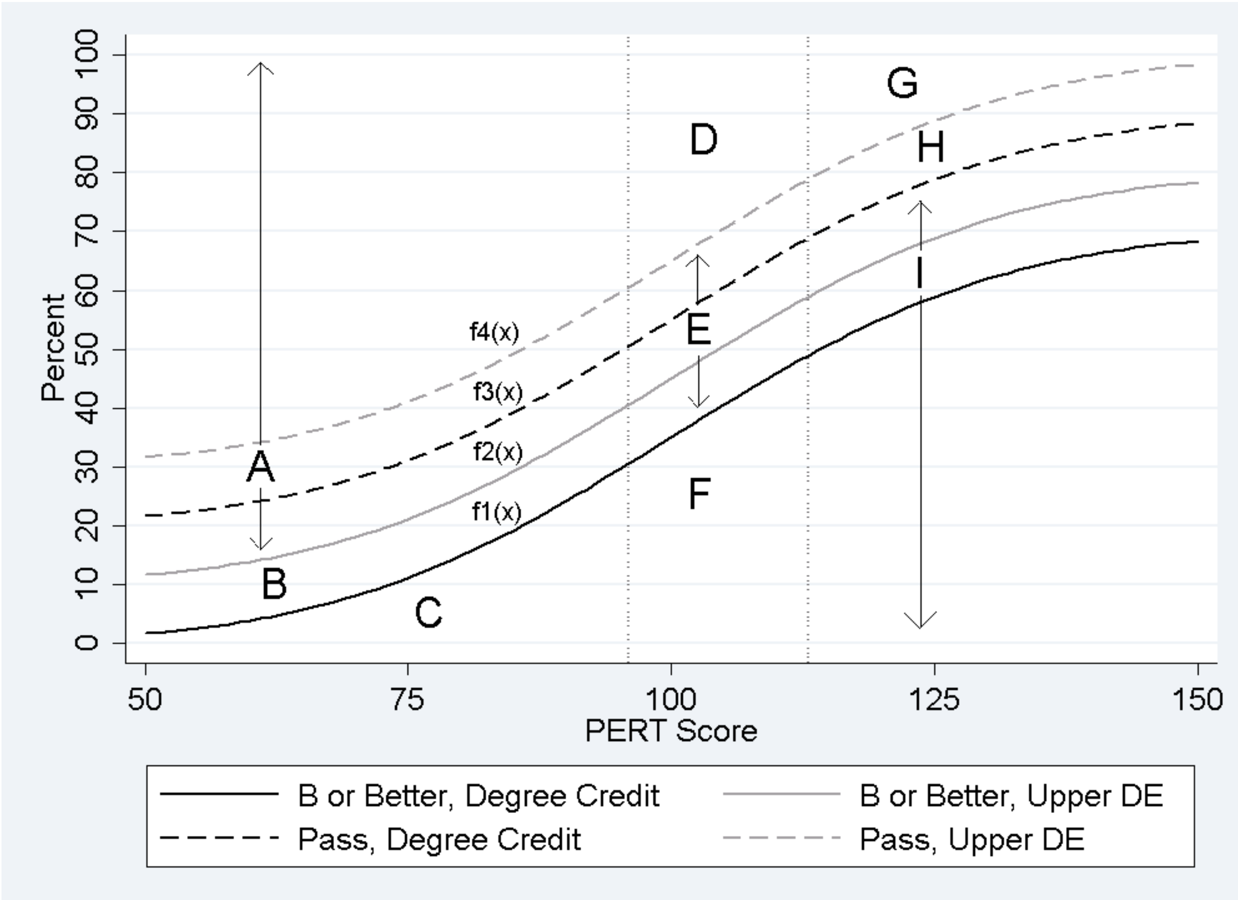
Figure 1. Hypothetical example of placement accuracy graph

Figure 2. Results for the predicted probabilities of course performance by PERT scores in math (top) and reading (bottom)

Figure 3. Developmental education placement and severe error rates for sample omega and tau values (math)

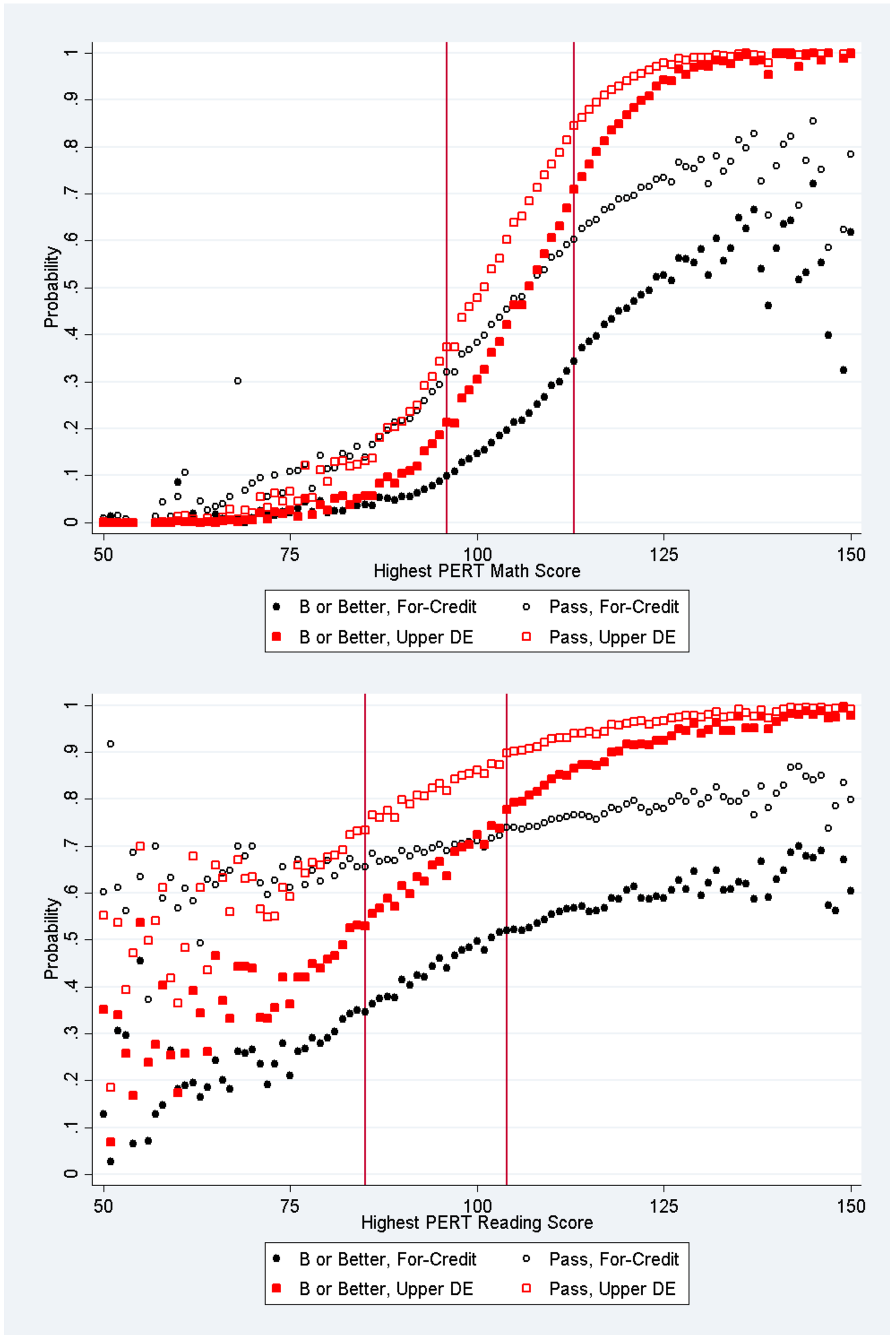
Figure 4. Developmental education placement and severe error rates for sample omega and tau values (English)

Figure 1.



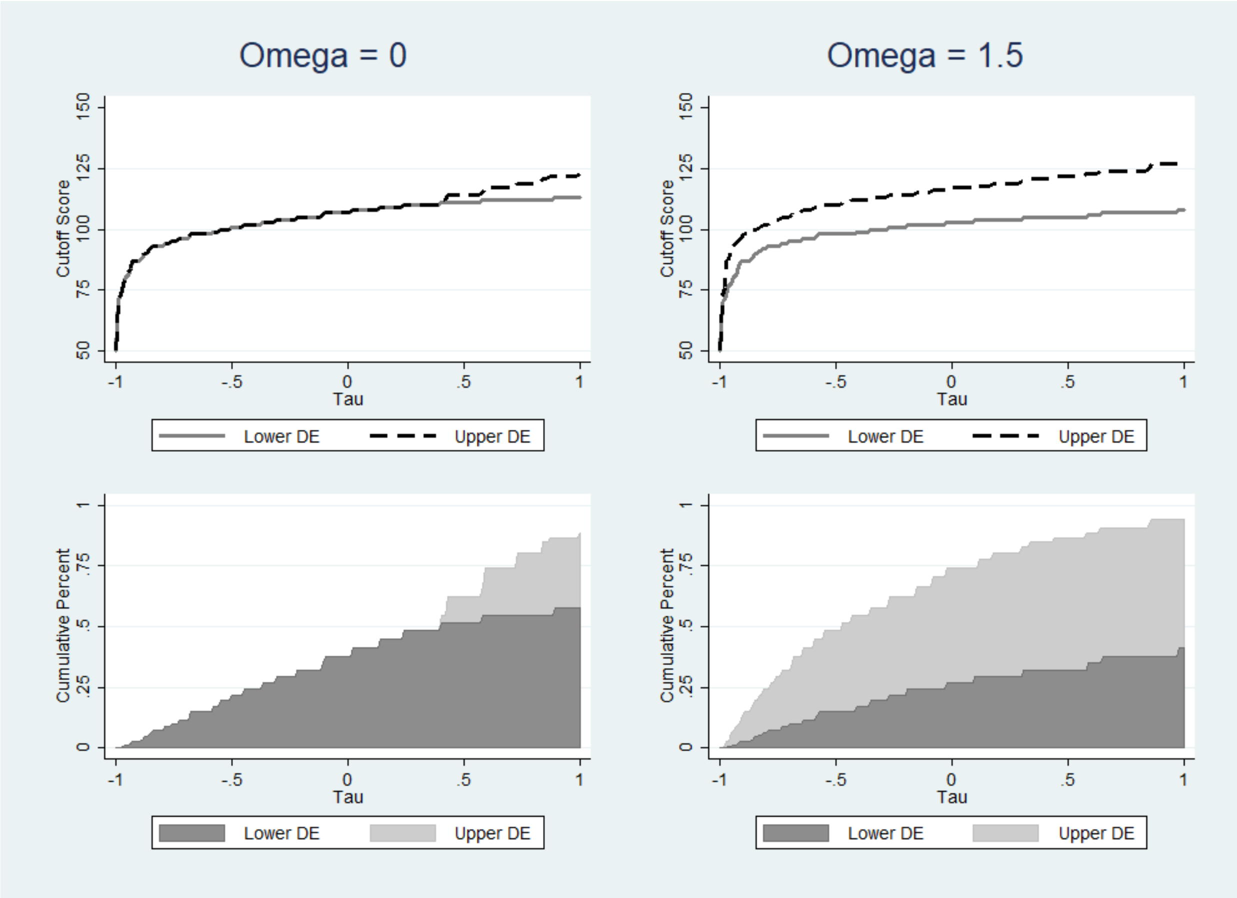
Note: The dotted vertical lines indicate cut scores where students were assigned to lower DE in math if they scored below 96 and upper DE if they scored below 113. In this figure, $f_1(x)$ represents the probability that a student with a given PERT score will receive a B or better in a for-credit course, $f_2(x)$ represents the probability that a student with a given PERT score will receive a B or better in upper DE, $f_3(x)$ represents the probability that a student with a given PERT score will pass a for-credit course, and $f_4(x)$ represents the probability that a student with a given PERT score will pass upper DE. Vertical dotted lines represent the cutoffs used to place students into upper DE and for-credit math courses.

Figure 2.



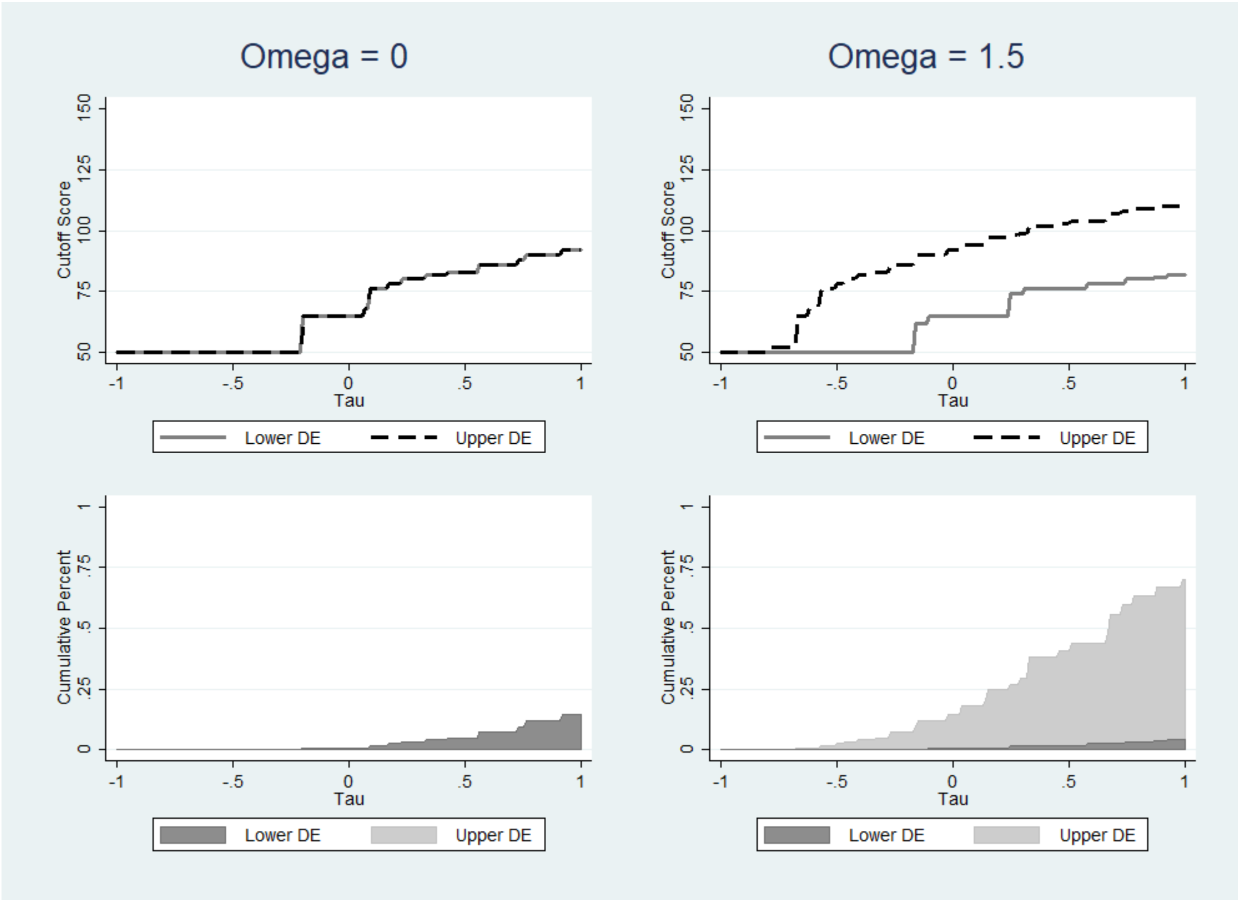
Note: Markers show the average predicted probability (y axis) of a particular outcome for students with a given score (x axis) on the PERT, which scales its scores from 50 to 150. Lack of support explains some of the variation at the extremes, while differences in student characteristics explain most of the remainder. Filled-in black circles show the probability of receiving a B or better in a for-credit course, hollow black circles show the probability of passing a for-credit course, filled-in red squares show the probability of receiving a B or better in upper DE, and hollow red squares show the probability of passing upper DE. Vertical lines show the cutoffs for upper DE and for-credit placement.

Figure 3.



Note: The top two graphs show the cutoffs for enrollment in lower DE and upper DE in math, respectively, as the penalty for overplacement increases from -1 to 1. Shaded areas in the bottom two graphs show the percent of students enrolled in lower DE and upper DE math, respectively, as the penalty for overplacement increases from -1 to 1. The total height of the bottom two graphs shows the number of students in either level of DE math. The two graphs on the left use an omega value of 0, which means that double misplacement is no worse than single misplacement. The two graphs on the right use an omega value of 1.5, which means that double misplacement is 2.5 times worse than single misplacement.

Figure 4.



Note: The top two graphs show the cutoffs for enrollment in lower DE and upper DE in English, respectively, as the penalty for overplacement increases from -1 to 1. Shaded areas in the bottom two graphs show the percent of students enrolled in lower DE and upper DE English, respectively, as the penalty for overplacement increases from -1 to 1. The total height of the bottom two graphs shows the number of students in either level of DE English. The two graphs on the left use an omega value of 0, which means that double misplacement is no worse than single misplacement. The two graphs on the right use an omega value of 1.5, which means that double misplacement is 2.5 times worse than single misplacement.

ONLINE APPENDICES

Appendix A: Solving First-Order Conditions

A.1: Minimizing SER

This section refers to Figure 1 in the main body of the paper.

By definition, $f_3(\cdot) \geq f_1(\cdot)$ and $f_4(\cdot) \geq f_2(\cdot)$—the probability of an A or B cannot be greater than that of an A, B, or C. Cutoffs are given by α and β, respectively. For ease of interpretation, we assume that $f_2(\cdot) \geq f_1(\cdot)$ and $f_4(\cdot) \geq f_3(\cdot)$—it is easier to succeed in upper DE than in a for-credit course. This may not apply to all students and could fail even in the aggregate due to self-selection, quality of instruction, or other factors. However, this assumption allows us to define each misplacement type more cleanly.

When minimizing the SER, we attempt to solve the Lagrangian:

$$\begin{aligned}\mathcal{L} = \int_{50}^{\alpha} f_2(x)\,dx + \int_{\alpha}^{\beta} [f_1(x) + 1 - f_4(x)]\,dx + \int_{\beta}^{150} [1 - f_3(x)]\,dx + \lambda_1(\alpha - 50) \\ + \lambda_2(\beta - \alpha) + \lambda_3(150 - \beta)\end{aligned}$$

Taking partial derivatives with respect to α and β (using Leibniz's Rule) gives:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\alpha} = f_2(\alpha)\frac{\partial(\alpha)}{\partial\alpha} - f_2(50)\frac{\partial(50)}{\partial\alpha} + \int_{50}^{\alpha}\frac{\partial f_2(x)}{\partial\alpha}dx + [f_1(\beta) + 1 - f_4(\beta)]\frac{\partial(\beta)}{\partial\alpha} \\ - [f_1(\alpha) + 1 - f_4(\alpha)]\frac{\partial(\alpha)}{\partial\alpha} + \int_{\alpha}^{\beta}\frac{\partial[f_1(x) + 1 - f_4(x)]}{\partial\alpha}dx \\ + [1 - f_3(150)]\frac{\partial(150)}{\partial\alpha} - [1 - f_3(\beta)]\frac{\partial(\beta)}{\partial\alpha} + \int_{\beta}^{150}\frac{\partial[1 - f_3(x)]}{\partial\alpha}dx + \lambda_1 - \lambda_2 \\ = 0\end{aligned}$$

and

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\beta} = f_2(\alpha)\frac{\partial(\alpha)}{\partial\beta} - f_2(50)\frac{\partial(50)}{\partial\beta} + \int_{50}^{\alpha}\frac{\partial f_2(x)}{\partial\beta}dx + [f_1(\beta) + 1 - f_4(\beta)]\frac{\partial(\beta)}{\partial\beta} \\ - [f_1(\alpha) + 1 - f_4(\alpha)]\frac{\partial(\alpha)}{\partial\beta} + \int_{\alpha}^{\beta}\frac{\partial[f_1(x) + 1 - f_4(x)]}{\partial\beta}dx \\ + [1 - f_3(150)]\frac{\partial(150)}{\partial\beta} - [1 - f_3(\beta)]\frac{\partial(\beta)}{\partial\beta} + \int_{\beta}^{150}\frac{\partial[1 - f_3(x)]}{\partial\beta}dx + \lambda_2 - \lambda_3 \\ = 0\end{aligned}$$

The assumptions outlined in section 4.3 are that:

1. $\frac{\partial(\beta)}{\partial\alpha} = \frac{\partial(\alpha)}{\partial\beta} = 0$
2. $\frac{\partial f_n(x)}{\partial\alpha} = \frac{\partial f_n(x)}{\partial\beta} = 0$

If we assume no corner solutions, then plugging in these values and simplifying gives

$$f_2(\alpha) - f_1(\alpha) = 1 - f_4(\alpha)$$

and

$$f_1(\beta) = [f_4(\beta) - f_3(\beta)]$$

A.2: Minimizing Weighted Error Rates

Assume that policymakers regard double misplacement (in either direction) as receiving penalty term ω and overplacement (at either level) as receiving penalty term τ.[1] When incorporating τ and ω, the initial Lagrangian becomes

$$\begin{aligned}\mathcal{L} = \int_{50}^{\alpha}(1+\omega)f_1(x)\,dx + \int_{50}^{\alpha}[f_2(x) - f_1(x)]\,dx + \int_{\alpha}^{\beta} f_1(x)\,dx + \int_{\alpha}^{\beta}(1+\tau)[1 - f_4(x)]\,dx \\ + \int_{\beta}^{150}(1+\tau)[f_4(x) - f_3(x)]\,dx + \int_{\beta}^{150}(1+\tau+\omega)[1 - f_4(x)]\,dx \\ + \lambda_1(\alpha - 50) + \lambda_2(\beta - \alpha) + \lambda_3(150 - \beta)\end{aligned}$$

which has first-order conditions

[1] Including an interaction term would require solving for three unknowns using two equations.

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \alpha} = (1+\omega)&\left[f_1(\alpha)\frac{\partial(\alpha)}{\partial \alpha} - f_1(50)\frac{\partial(50)}{\partial \alpha} + \int_{50}^{\alpha}\frac{\partial f_1(x)}{\partial \alpha}dx\right] + [f_2(\alpha) - f_1(\alpha)]\frac{\partial(\alpha)}{\partial \alpha} \\
&- [f_2(50) - f_1(50)]\frac{\partial(50)}{\partial \alpha} + \int_{50}^{\alpha}\frac{\partial[f_2(x) - f_1(x)]}{\partial \alpha}dx + f_1(\beta)\frac{\partial(\beta)}{\partial \alpha} \\
&- f_1(\alpha)\frac{\partial(\alpha)}{\partial \alpha} + \int_{\alpha}^{\beta}\frac{\partial f_1(x)}{\partial \alpha}dx \\
&+ (1+\tau)\left[[1 - f_4(\beta)]\frac{\partial(\beta)}{\partial \alpha} - [1 - f_4(\alpha)]\frac{\partial(\alpha)}{\partial \alpha} + \int_{\alpha}^{\beta}\frac{\partial[1 - f_4(x)]}{\partial \alpha}dx\right] \\
&+ (1+\tau)\Bigg[[f_4(150) - f_3(150)]\frac{\partial(150)}{\partial \alpha} - [f_4(\beta) - f_3(\beta)]\frac{\partial(\beta)}{\partial \alpha} \\
&+ \int_{\alpha}^{\beta}\frac{\partial[f_4(x) - f_3(x)]}{\partial \alpha}dx\Bigg] \\
&+ (1+\tau+\omega)\Bigg[[1 - f_4(150)]\frac{\partial(150)}{\partial \alpha} - [1 - f_4(\beta)]\frac{\partial(\beta)}{\partial \alpha} \\
&+ \int_{\beta}^{150}\frac{\partial[1 - f_4(x)]}{\partial \alpha}dx\Bigg] + \lambda_1 - \lambda_2 = 0
\end{aligned}$$

and

$$\frac{\partial \mathcal{L}}{\partial \beta} = (1+\omega)\left[f_1(\alpha)\frac{\partial(\alpha)}{\partial \beta} - f_1(50)\frac{\partial(50)}{\partial \beta} + \int_{50}^{\alpha}\frac{\partial f_1(x)}{\partial \beta}dx\right] + [f_2(\alpha) - f_1(\alpha)]\frac{\partial(\alpha)}{\partial \beta}$$
$$- [f_2(50) - f_1(50)]\frac{\partial(50)}{\partial \beta} + \int_{50}^{\alpha}\frac{\partial[f_2(x) - f_1(x)]}{\partial \beta}dx + f_1(\beta)\frac{\partial(\beta)}{\partial \beta}$$
$$- f_1(\alpha)\frac{\partial(\alpha)}{\partial \beta} + \int_{\alpha}^{\beta}\frac{\partial f_1(x)}{\partial \beta}dx$$
$$+ (1+\tau)\left[[1 - f_4(\beta)]\frac{\partial(\beta)}{\partial \beta} - [1 - f_4(\alpha)]\frac{\partial(\alpha)}{\partial \beta} + \int_{\alpha}^{\beta}\frac{\partial[1 - f_4(x)]}{\partial \beta}dx\right]$$
$$+ (1+\tau)\left[[f_4(150) - f_3(150)]\frac{\partial(150)}{\partial \beta} - [f_4(\beta) - f_3(\beta)]\frac{\partial(\beta)}{\partial \beta}\right.$$
$$\left. + \int_{\alpha}^{\beta}\frac{\partial[f_4(x) - f_3(x)]}{\partial \beta}dx\right]$$
$$+ (1+\tau+\omega)\left[[1 - f_4(150)]\frac{\partial(150)}{\partial \beta} - [1 - f_4(\beta)]\frac{\partial(\beta)}{\partial \beta}\right.$$
$$\left. + \int_{\beta}^{150}\frac{\partial[1 - f_4(x)]}{\partial \beta}dx\right] + \lambda_2 - \lambda_3 = 0$$

Imposing assumptions 1-3 and simplifying partials then gives

$$\frac{\partial \mathcal{L}}{\partial \alpha} = (1+\omega)f_1(\alpha) + [f_2(\alpha) - f_1(\alpha)] - f_1(\alpha) - (1+\tau)[1 - f_4(\alpha)] = 0$$

and

$$\frac{\partial \mathcal{L}}{\partial \beta} = f_1(\beta) + (1+\tau)[1 - f_4(\beta)] - (1+\tau)[f_4(\beta) - f_3(\beta)] - (1+\tau+\omega)[1 - f_4(\beta)] = 0$$

which simplify to

$$\omega f_1(\alpha) + [f_2(\alpha) - f_1(\alpha)] = (1+\tau)[1 - f_4(\alpha)]$$

and

$$f_1(\beta) = (1+\tau)[f_4(\beta) - f_3(\beta)] + \omega[1 - f_4(\beta)]$$

Using probabilities $f_1(\cdot)$ through $f_4(\cdot)$, we can solve for ω and τ, the preferences of a perfectly rational policymaker.

A.3: Determining Optimal Cutoff Values

After taking first-order conditions, we make two assumptions:

1. α and β are not functions of one another.
2. The probability of success in either course level is independent of cutoff values; i.e. peer effects and manipulation of course placement have little effect on whether students are misplaced *at particular grade and placement margins* $\left(\frac{\partial f_n(x)}{\partial \alpha} \approx \frac{\partial f_n(x)}{\partial \beta} \approx 0\right)$.

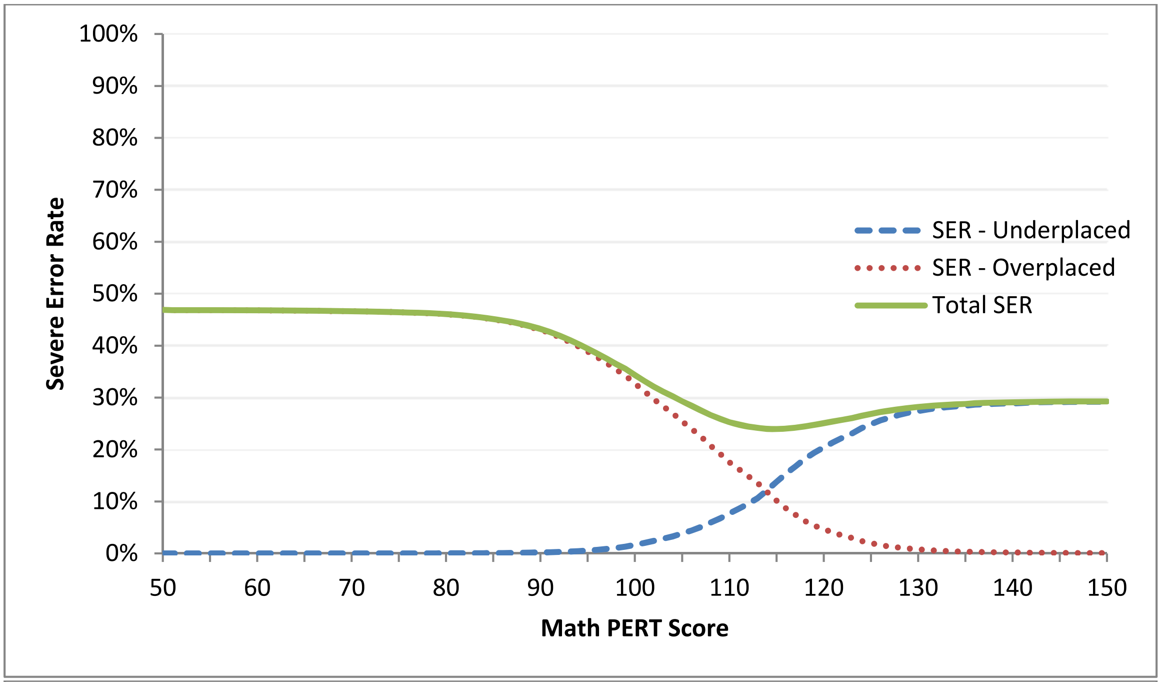
The first assumption is intuitive—we cannot determine optimal cutoff values if either is constrained unnecessarily. The second assumption is a narrow one—some peer effects do exist, and students can improve their PERT scores through retesting—so we merely claim that peer effects are unlikely to push many students within a certain score range into a given grade range.

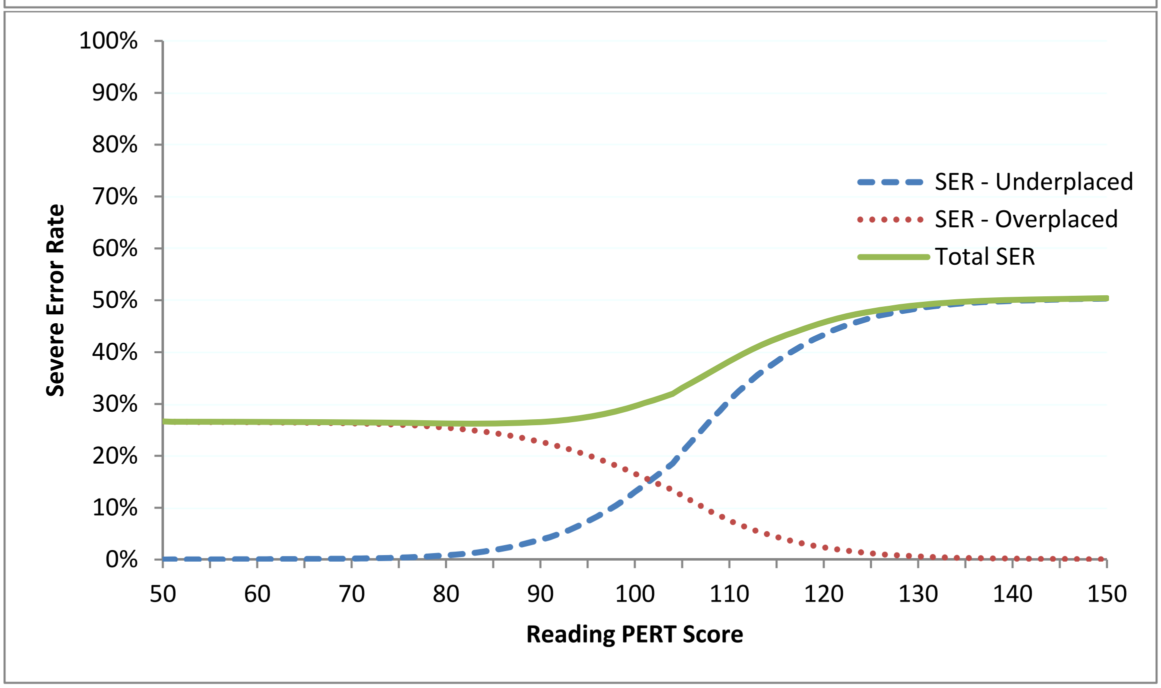
If there are no corner solutions ($50 \neq \alpha \neq \beta \neq 150$), then marginal single underplacement should equal marginal single overplacement at each cutoff. Double misplacement is irrelevant: because moving one cutoff cannot change double misplacement to accurate placement, these students will be misplaced either way, and the SER treats all misplacement as equally harmful.

Setting $\alpha = \beta$ would eliminate placement into upper DE—students below the combined cutoff would be placed into lower DE, while those above would be placed into for-credit courses. In this case, total marginal overplacement equals total marginal underplacement—since upper DE would be eliminated, the only relevant question would be whether students were underplaced in lower DE or overplaced in for-credit courses.

Appendix B: Visualizing the SER

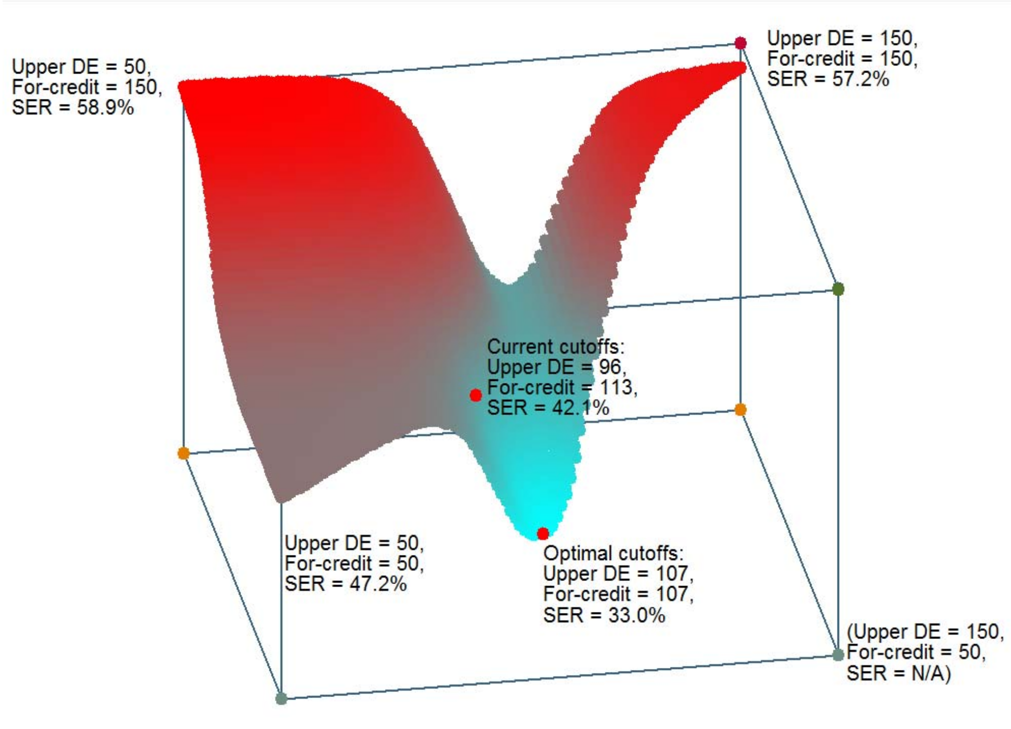
Figure B.1. SER by cutoff value, single cutoff model in math (top) and reading (bottom).

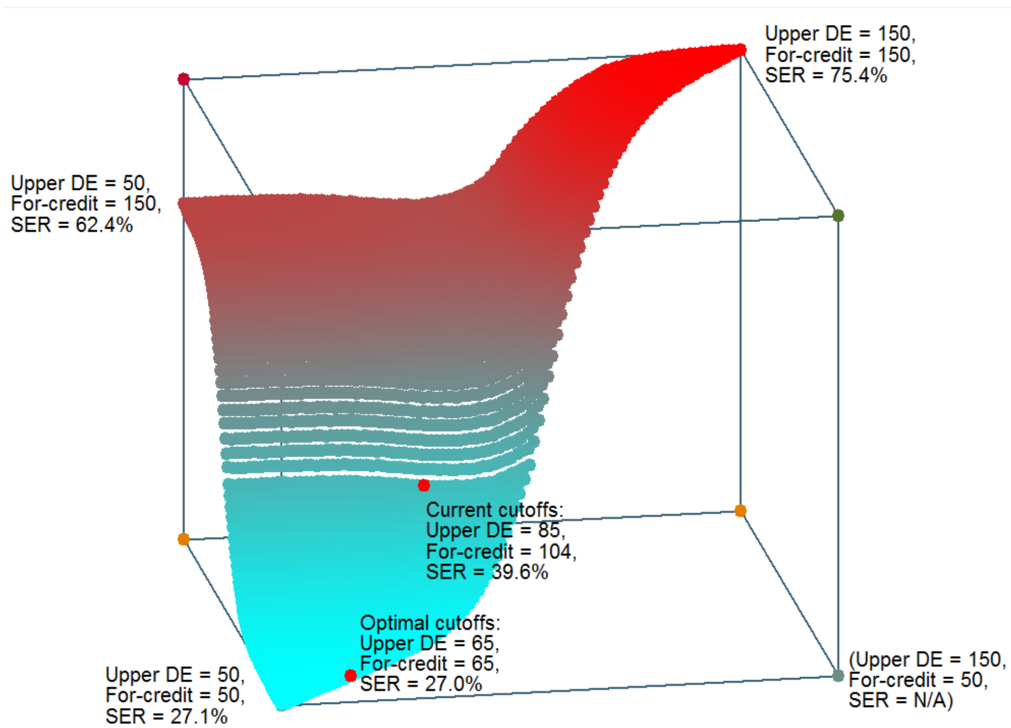




Note: The x axis shows the PERT scores usable as the cutoff for for-credit placement in a one-cutoff model of severe error. The y axis shows the SER. The dashed line shows the component of the SER at a given cutoff due to underplacement. The dotted line shows the component of the SER at a given cutoff due to overplacement. The solid line shows the total SER, the sum of these two components, which we hope to minimize.

Figure B.2. SER by cutoff value, double cutoff model in math (above) and English (below).





Note: The x and z axes (measuring length and depth respectively) represent possible values of the cutoff scores to place into upper DE and for-credit courses respectively. Figures are shaded on a gradient with red representing the highest SER and light blue representing the lowest SER, with two points in red illustrating current cutoffs and SER-minimizing cutoffs. Points at maximum and minimum PERT values have been labeled, along with the current and SER-minimizing cutoffs.

Appendix C: Implications of Policymaker Preferences

Table C1. Implications of policymaker preferences for various forms of misplacement, in math and English.

		Math								English							
		Cutoff		Placement		Misplacement				Cutoff		Placement		Misplacement			
τ	ω	UDE	For-credit	Lower DE	Upper DE	Double Under	Single Under	Single Over	Double Over	UDE	For-credit	Lower DE	Upper DE	Double Under	Single Under	Single Over	Double Over
-1	0.0	50	50	0.0%	0.0%	0.0%	0.0%	17.4%	29.8%	50	50	0.0%	0.0%	0.0%	0.0%	14.6%	12.4%
-0.5	0.0	101	101	22.0%	0.0%	1.6%	1.6%	16.7%	16.2%	50	50	0.0%	0.0%	0.0%	0.0%	14.6%	12.4%
0	0.0	107	107	37.9%	0.0%	5.0%	5.5%	14.2%	8.3%	65	65	0.6%	0.0%	0.1%	0.1%	14.7%	12.2%
0.5	0.0	111	114	51.4%	10.9%	8.5%	12.9%	11.1%	2.7%	83	83	5.0%	0.0%	1.3%	0.8%	14.7%	10.6%
1	0.0	113	123	57.8%	30.6%	10.5%	24.1%	5.9%	0.2%	92	92	14.8%	0.0%	5.0%	2.7%	13.7%	8.3%
-1	0.5	67	105	0.2%	32.1%	0.0%	3.8%	34.5%	10.3%	50	65	0.0%	0.6%	0.0%	0.1%	14.9%	12.2%
-0.5	0.5	102	108	24.5%	16.9%	2.4%	5.9%	20.0%	7.2%	65	65	0.6%	0.0%	0.1%	0.1%	14.7%	12.2%
0	0.5	106	112	35.0%	19.6%	4.3%	9.8%	16.3%	3.9%	76	76	2.0%	0.0%	0.4%	0.3%	14.7%	11.6%
0.5	0.5	109	117	44.7%	29.6%	6.6%	17.3%	11.3%	1.2%	83	83	\	0.0%	1.3%	0.8%	14.7%	10.6%
1	0.5	111	124	51.4%	39.1%	8.5%	24.9%	6.7%	0.1%	90	90	11.9%	0.0%	3.8%	2.1%	14.1%	8.9%
-1	1.0	50	109	0.0%	44.7%	0.0%	6.6%	36.5%	6.3%	50	82	0.0%	4.4%	0.0%	1.1%	16.3%	10.8%
-0.5	1.0	99	112	17.1%	37.5%	1.3%	9.4%	24.5%	3.9%	50	82	0.0%	4.4%	0.0%	1.1%	16.3%	10.8%
0	1.0	104	115	29.4%	37.3%	3.2%	13.8%	17.5%	2.1%	76	83	2.0%	3.0%	0.4%	1.2%	15.6%	10.6%
0.5	1.0	107	119	37.9%	42.6%	5.0%	19.7%	12.3%	0.7%	82	90	4.4%	7.5%	1.1%	3.4%	15.9%	8.9%
1	1.0	109	124	44.7%	45.8%	6.6%	24.7%	8.4%	0.1%	86	94	7.3%	11.0%	2.1%	5.6%	15.7%	7.6%
-1	1.5	50	112	0.0%	54.6%	0.0%	9.4%	36.8%	3.9%	50	86	0.0%	7.3%	0.0%	2.1%	17.0%	10.0%
-0.5	1.5	98	114	15.3%	47.0%	1.0%	12.0%	25.0%	2.7%	50	90	0.0%	11.9%	0.0%	3.8%	17.6%	8.9%
0	1.5	103	117	27.0%	47.2%	2.8%	16.8%	17.6%	1.2%	65	92	0.6%	14.2%	0.1%	5.0%	17.6%	8.3%
0.5	1.5	105	121	32.3%	52.6%	3.8%	21.5%	13.6%	0.4%	76	97	2.0%	22.8%	0.4%	9.3%	17.6%	6.5%
1	1.5	108	127	41.5%	52.9%	5.8%	26.6%	8.4%	0.0%	82	103	4.4%	36.6%	1.1%	16.8%	16.7%	4.2%
-1	2.0	50	113	0.0%	57.8%	0.0%	10.5%	36.7%	3.3%	50	92	0.0%	14.8%	0.0%	5.0%	17.9%	8.3%
-0.5	2.0	97	115	13.4%	53.2%	0.8%	13.6%	25.7%	2.1%	50	94	0.0%	18.4%	0.0%	6.5%	18.1%	7.6%
0	2.0	102	118	24.5%	53.2%	2.4%	18.2%	18.1%	0.9%	50	97	0.0%	24.8%	0.0%	9.4%	18.5%	6.5%
0.5	2.0	104	122	29.4%	57.3%	3.2%	22.3%	14.3%	0.3%	65	102	0.6%	37.7%	0.1%	15.9%	18.2%	4.5%
1	2.0	107	127	37.9%	56.5%	5.0%	26.5%	9.5%	0.0%	78	107	2.6%	53.3%	0.6%	24.8%	16.5%	2.6%

Appendix D: Sensitivity Analysis of Placement Accuracy Using a Grade of "D" as Passing instead of "C"

Table 1. Minimizing the math SER under a double cutoff model, using a grade of "D" as passing instead of "C".

	(1)	(2)	(3)	(4)	(5)	(6)
Metric	PERT	FCAT	Subject GPA	Overall GPA	All HS data	All HS + PERT
	Holding DE rates fixed					
Upper DE cutoff	96	306	13	12	12	12
For-credit cutoff	113	337	59	59	58	58
Lower DE placement	11.7%	12.2%	11.5%	11.6%	12.0%	12.0%
Upper DE placement	46.1%	46.2%	46.0%	46.3%	46.0%	46.0%
Double underplacement	0.7%	1.4%	1.0%	0.8%	0.7%	0.5%
Single underplacement	10.5%	12.1%	11.0%	10.6%	9.8%	9.0%
Single overplacement	20.8%	19.5%	18.3%	18.1%	18.3%	18.6%
Double overplacement	2.6%	4.1%	4.2%	3.3%	2.9%	2.3%
Total SER	34.6%	37.1%	34.4 Table %	32.9%	31.7%	30.3%
	Optimal Cutoffs					
Upper DE cutoff	102	311	26	23	25	28
For-credit cutoff	102	311	26	23	25	28
Lower DE placement	24.6%	18.0%	25.6%	22.9%	25.0%	28.0%
Upper DE placement	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Double underplacement	2.4%	2.4%	2.9%	2.2%	2.1%	2.2%
Single underplacement	2.5%	1.7%	4.6%	3.6%	3.8%	3.7%
Single overplacement	8.1%	6.7%	5.0%	5.3%	5.3%	6.3%
Double overplacement	10.9%	15.4%	12.8%	12.7%	11.6%	9.4%
Total SER	23.9%	26.3%	25.3%	23.8%	22.9%	21.6%

Note: "Upper DE cutoff" and "For-credit cutoff" show scores or percentile values at which students are first placed in upper DE and for-credit courses respectively. "Lower DE placement" and "Upper DE placement" show the percent of students placed in lower and upper DE respectively. "Double underplacement" shows the percent of all students predicted to get a B or better two course levels above their placement. "Single underplacement" shows the percent of all students predicted to get a B or better one course level above their placement. "Single overplacement" shows the percent of all students placed in upper DE or for-credit courses and predicted to fail. "Double overplacement" shows the percent of all students placed in for-credit courses but predicted to fail upper DE. Panels holding DE rates fixed use the scores or percentile cutoffs that keeps both DE placement rates as close to the current rates as possible. Panels using optimal cutoffs use scores or percentile cutoff values under each placement metric that minimize the SER.

Table 2. Minimizing the English SER under a double cutoff model using a grade of "D" as passing instead of "

	(1)	(2)	(3)	(4)	(5)	(6)
Metric	PERT	FCAT	Subject GPA	Overall GPA	All HS Data	All HS + PERT
	Holding DE rates fixed					
Upper DE cutoff	85	265	7	7	6	6
For-credit cutoff	104	312	46	44	43	43
Lower DE placement	6.4%	6.3%	7.0%	6.7%	6.0%	6.0%
Upper DE placement	37.4%	37.5%	37.2%	37.2%	37.0%	37.0%
Double underplacement	1.9%	1.9%	1.5%	1.3%	0.9%	0.9%
Single underplacement	18.1%	17.4%	16.0%	14.7%	13.7%	13.6%
Single overplacement	12.0%	12.0%	10.3%	10.7%	10.9%	11.0%
Double overplacement	3.2%	3.2%	2.5%	1.8%	1.7%	1.6%
Total SER	35.2%	34.5%	30.4%	28.5%	27.2%	27.0%
	Optimal Cutoffs					
Upper DE cutoff	50	167	1	3	6	6
For-credit cutoff	50	167	1	3	6	6
Lower DE placement	0.0%	0.1%	0.7%	3.0%	6.0%	6.0%
Upper DE placement	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Double underplacement	0.0%	0.0%	0.1%	0.5%	0.9%	0.9%
Single underplacement	0.0%	0.0%	0.3%	0.5%	1.1%	1.0%
Single overplacement	9.7%	9.8%	9.6%	9.5%	9.2%	9.3%
Double overplacement	10.0%	9.9%	9.7%	8.7%	7.7%	7.6%
Total SER	19.7%	26.5%	19.6%	19.2%	19.0%	18.8%

Note: "Upper DE cutoff" and "For-credit cutoff" show scores or percentile values at which students are first placed in upper DE and for-credit courses respectively. "Lower DE placement" and "Upper DE placement" show the percent of students placed in lower and upper DE respectively. "Double underplacement" shows the percent of all students predicted to get a B or better two course levels above their placement. "Single underplacement" shows the percent of all students predicted to get a B or better one course level above their placement. "Single overplacement" shows the percent of all students placed in upper DE or for-credit courses and predicted to fail. "Double overplacement" shows the percent of all students placed in for-credit courses but predicted to fail upper DE. Panels holding DE rates fixed use the scores or percentile cutoffs that keeps both DE placement rates as close to the current rates as possible. Panels using optimal cutoffs use scores or percentile cutoff values under each placement metric that minimize the SER.