ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 20-302

An Evaluation of Credit Recovery as an Intervention for High School Students Who Fail Courses

Samantha Viano
George Mason University

Gary T. Henry
University of Delaware

Credit recovery (CR) refers to online courses that high school students take after previously failing the course. Many have suggested that CR courses are helping students to graduate from high school without corresponding increases in academic skills. This study analyzes administrative data from the state of North Carolina to evaluate these claims using full data from public and private CR providers. Findings indicate that students who fail courses and enroll in CR have lower test scores of up to two tenths of a standard deviation and are about seven percent more likely to graduate high school on time than students who repeat courses traditionally. Test score differences are particularly large for Biology compared to Math I and English II. Hispanic and economically disadvantaged CR students are more likely to graduate high school than their peers.

VERSION: September 2020

Suggested citation: Viano, Samantha, and Gary T. Henry. (2020). An Evaluation of Credit Recovery as an Intervention for High School Students Who Fail Courses. (EdWorkingPaper: 20-302). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/kcx1-vk12

An Evaluation of Credit Recovery as an Intervention for High School Students Who Fail Courses

Samantha Viano

George Mason University

Gary T. Henry

University of Delaware

Abstract

Credit recovery (CR) refers to online courses that high school students take after previously failing the course. Many have suggested that CR courses are helping students to graduate from high school without corresponding increases in academic skills. This study analyzes administrative data from the state of North Carolina to evaluate these claims using full data from public and private CR providers. Findings indicate that students who fail courses and enroll in CR have lower test scores of up to two tenths of a standard deviation and are about seven percent more likely to graduate high school on time than students who repeat courses traditionally. Test score differences are particularly large for Biology compared to Math I and English II. Hispanic and economically disadvantaged CR students are more likely to graduate high school than their peers.

In 1996, the high school graduation rate dropped to a low of 71 percent after three decades of decline and stagnation since its peak of 79 percent in 1970. Over the last fifteen years, the high school graduation rate has bounced back, rising each year since 2002, reaching a record high in the 2010-11 school year (*U.S. High School Graduation Rate Hits New Record High | U.S. Department of Education*, n.d.). Without parallel trends to explain why this is occurring (Murnane, 2013), one potential explanation is increased access to online credit recovery (CR) courses. CR refers to an opportunity for a student who has previously failed a course to retake it through an online course—sometimes with in-person classroom supports—to earn the lost credit (*Dropout Prevention Intervention Report: Credit Recovery Programs*, 2015). While few studies have empirically examined the effectiveness and potential unintended consequences of CR (notable exceptions include Hart et al., 2019; Rickles et al., 2018), many reports in the media have accused CR of allowing schools to game the system by helping students to graduate without these students gaining the knowledge or skills that are expected from the course. National Public Radio called CR a "questionable quick fix," Education Week wrote about it as "a scandal," and the New York Times included it as part of a system that has produced a "counterfeit high school diploma" (Gardner, 2016; Rich, 2015; Turner, 2015). In this study, we seek to empirically examine these claims that CR has increased graduation rates at the expense of student learning.

Graduation rate increases have not occurred evenly across all students. Larger gains in graduation rates of Black and Hispanic students have closed the gap between the graduation rate of White students and Black and Hispanic students by approximately 20 percent—or three percentage points—between 2003 and 2014 (*National Center for Education Statistics*, n.d.). This pattern is also reflected in the status dropout rate—the percent of 16- to 24-year-olds who are not enrolled in school and do not have a high school credential—by income level. At all income quintiles, the status dropout rate decreased between 2003 and 2014, with the largest decrease for students in the lowest income quartile, for whom status dropout rates decreased approximately 40 percent (*National Center for Education Statistics*, n.d.). The evidence clearly shows the graduation rate is increasing overall and by student subgroup, but it is unclear why it is increasing. If CR is more effective for certain student subgroups than others, this could help explain recent trends in graduation rates.

In this study, we examine the effectiveness of CR as a tool that schools are using in order to graduate more students from high school or prevent them from dropping out of high school, particularly for traditionally underserved student populations. Whether or not CR is associated with increasing graduation rates and reducing dropouts, some critics have opined that these benefits may be hollow, coming at the expense of learning that is expected to occur in high school (see Dynarski, 2018; Smiley, 2017). To examine the possibility of adverse side effects, we examine the effects of CR on end of course exam scores.

Specifically, we address two research questions:

1. Do students who take CR courses differ from students who retake courses traditionally in terms of end of course exam scores, dropping out, or graduating from high school within four years?
2. Do Black, Hispanic, or economically disadvantaged students who take CR courses differ from other students in their respective sub-population who retake courses traditionally in terms of end of course exam scores, dropping out, or graduating from high school within four years?

Policy Context of Current Study

While high school graduation rates in recent years continue to increase, a significant proportion of students continue to not graduate high school on time with traditionally disadvantaged subgroups of students having much lower graduation rates (*Public High School Graduation Rates*, 2020). Both the accountability movement and the corresponding push for the high school diploma to signify college or career readiness may have suppressed graduation rates by introducing new graduation requirements (Holme et al., 2010; Howard et al., 2015; Papay et al., 2010; Plunk et al., 2014; Reardon et al., 2010).

At the same time, the average four year cohort graduation rate continues to reach record highs on an annual basis, with most recent estimates indicating the graduation rate was almost 85 percent in 2016-17 (*Public High School Graduation Rates*, 2020). A key shift in the accountability movement quietly took place around the 2010-11 school year when the No Child Left Behind Act began requiring high schools to report their four year cohort graduation rates (*No Child Left Behind High School Graduation Rate Non-Regulatory Guidance,* 2008). At the same time, the Race to the Top grant competition required applicant states to create a system that would dramatically reform their lowest performing schools where these schools were identified by test score proficiency or graduation rates (*Overview Information; Race to the Top Fund; Notice Inviting Applications for New Awards for Fiscal Year 2010; Notice*, 2010). Ten years after test score proficiency was made a national priority, high school graduation was added to the roster of high-stakes accountability targets across the U.S. With high schools facing sanctions if they had low graduation rates, the key question was how would they go about increasing the graduation rate?

Course Failure as a Root Cause of Reduced Graduation Rates

Students face many institutional and personal barriers to graduating from high school. In particular, studies have consistently found that, even when controlling for demographics, test scores, and other key characteristics, failing courses or low credit accumulation are associated with lower rates of high school graduation (Bowers, 2010; Mac Iver & Messel, 2013; Neild et al., 2008; Silver et al., 2008). The most proximal solutions to address course failure or low credit accumulation are for the students to earn the credits they lost when they failed the course. High schools have historically done this through remedial courses in the summer/after school or repeating the course in full (Cooper et al., 2000; Lauer et al., 2006). Although no available studies assess the effectiveness of these traditional options on likelihood of high school graduation, one evaluation estimated the effects of repeating a course in full after course failure on earning course credit, finding three quarters of students who repeated Algebra I in a California school district passed the course the second time (Fong et al., 2014).

Currently, high schools are left with the knowledge that they are held accountable for graduation rates and students are less likely to graduate from high school if they fail courses with the research-based solutions assessing the means to both prevent course failure and address course failure are severely lacking. Yet, high school graduation rates have been increasing over the last fifteen years. This is the point when CR entered the picture as a possible strategy for increasing graduation rates. Many have hypothesized that students finish CR courses much faster and with less effort than traditional courses (see Smiley, 2017). If CR is more efficient than repeating a course in full, then students who fail courses and enroll in CR would theoretically be more likely to graduate from high school than those repeating the course.

These hypotheses have only begun to be empirically tested. In a study of CR courses offered through the state-run North Carolina Virtual Public School (NCVPS), CR students were less likely to graduate from high school overall but more likely to graduate within four years than students who retook courses in full (Stallings et al., 2016). This study did not include CR course-taking from privately-run CR providers (a popular option in North Carolina, as discussed below) and estimated associations using covariate-adjusted regressions. The second study used a randomized control trial (RCT) design to randomly assigned students in 17 Chicago high schools who had failed Algebra I in ninth grade to enroll in traditional summer school or a CR Algebra I course the following summer. Despite higher passage rates and lower posttest scores of the traditional retake as compared to the CR course, there was no statistically significant difference in the high school graduation rate of the CR versus face-to-face assigned students (Rickles et al., 2018). While this study is causal, the generalizability is limited. CR is often offered during the school year and there are a multitude of different providers. Two additional studies have provided suggestive evidence linking CR or courses that are likely to be CR with persistence to 12th grade, graduating from high school, and enrollment in postsecondary education (Hart et al., 2019; Heinrich & Darling-Aduana, 2019).

Potential Unintended Consequences of Credit Recovery

Journalists and others have suggested that CR courses are low quality and lead to a reduction in academic standards to earn a high school diploma (e.g., Burke et al., 2013; Dynarski, 2018; Gardner, 2016; Smiley, 2017). Critics are concerned that online courses as currently designed and offered are unable to help at-risk students to learn course material they did not learn in a face-to-face course. Many point out that online courses were designed for students taking extension courses or advanced courses, not for students who are struggling academically (see Viano, 2018). Instead of helping students learn material, CR course enrollment could lead to less learning and lower test scores if not designed and administered properly. This is especially a concern if CR courses have especially negative effects on already academically vulnerable student populations like students of color and those who come from low socio-economic backgrounds.

How Credit Recovery is Implemented in Schools Nationwide and in North Carolina

In the 2015-16 school year, 69 percent of high schools across the country had active CR programs with an average of six percent of high school students enrolled in at least one CR course (Tyner & Munyan-Penney, 2018). Data from the National Survey on High School Strategies Designed to Help At-Risk Students Graduate (HSS), a nationally-representative survey sponsored by the U.S. Department of Education, indicated that 71 percent of schools offered at least one CR course in the 2014-15 school year, 42 percent of schools allowed students to repeat courses traditionally, and a quarter of schools offered both options[1] (*Issue Brief: Credit Recovery*, 2018). In the 2016-17 school year, 82 percent of schools in North Carolina had at least one student enrolled in CR. Among schools that offered CR, the median school had a fifth of eligible students enrolled in CR (range of 0.7 percent to 100 percent; Viano, 2019). We discuss how CR is used in schools nationwide and in North Carolina, specifically, because, while CR is popular nationwide, much of the research available on CR selection is based on information from North Carolina. This study also analyzes data from North Carolina Public Schools (see the Methods section), so in this section we situate the North Carolina setting within the broader CR context.

How Credit Recovery Courses Are Offered

North Carolina schools have a configuration of CR offerings that is similar to other states. North Carolina has a publicly-run online course provider, NCVPS, and allows districts and schools to contract with privately-run providers of CR. As of 2016, 24 states had publicly-run providers of online courses, enrolling over half a million students, and privately-run CR providers are very popular nationwide (Gemin & Pape, 2017). NCVPS creates course content and serves as a state-run vendor of courses (students do not enroll full time in NCVPS) including asynchronous instructional support for the courses. The majority of schools offer CR through third-party vendors that offer these courses to schools nationwide including NovaNET, OdysseyWare, Edmentum/E2020, Plato, Grad Point, Edgenuity, and Apex (*Approved Vendor List*, n.d.; Stallings et al., 2016; authors' analysis). The providers have online course options for all traditional high school courses offered in North Carolina, marketing their ability to provide an online version for hundreds of different courses, with options for synchronous and asynchronous instruction (*Edmentum Courseware*, 2020; Stallings et al., 2016; authors' analysis and interviews with school districts). While private providers are a substantial source for CR courses, most prior studies were limited to public providers (Hart et al., 2019; Stallings et al., 2016) or a single private provider (Heinrich & Darling-Aduana, 2019; Rickles et al., 2018).

Which Students Enroll in Credit Recovery

More information is available on student enrollment/selection into CR. For instance, the 2014-15 HSS indicated schools targeted students for CR most often based on their academic performance (87 percent), attendance problems (73 percent), staff referral (60 percent), and discipline or behavioral issues (48 percent; *Issue Brief: Credit Recovery*, 2018). A study on 24 Massachusetts high schools that were part of the MassGrad initiative found that students' grade level was a common CR selection criterion. In particular, some schools used CR for 12th grade students who only needed one or two credits to graduate while other schools gave preference for 10th graders to enroll in CR over 9th graders. Across all schools, a slightly higher percentage of students enrolled in CR in 12th grade (36 percent) with the other high school grades enrolling very similar percentages of students in CR. Some schools reported being less likely to enroll students with low English proficiency, high need for teacher support, or external commitments (jobs, child care) after seeing these types of students struggle in CR. While schools over time reported beginning to more purposefully identify students most likely to succeed in CR, most schools said they were willing to enroll any student in CR (Levine et al., 2017).

A study of students who failed courses in the 2008-09 through 2011-12 school years compared students enrolled in NCVPS CR to those who repeated courses traditionally or enrolled in non-NCVPS CR/repeated courses traditionally. They found that Hispanic students were slightly less likely to enroll in NCVPS CR. White students and economically disadvantaged students were slightly more likely to enroll in NCVPS CR. There were no differences in enrollment by gender or by special education status (Stallings et al., 2016).

In order to gain a more nuanced understanding of CR assignment in North Carolina, we conducted interviews with school district officials with CR assignment responsibilities from three school districts in North Carolina in fall 2017. Each district had different policies for enrolling students in CR although with some commonalities between them. For instance, all districts allowed students to request assignment to CR courses and updated their CR policies on a regular basis, indicating that CR assignment was evolving over time. In the district with the most formalized CR assignment policies, the district had a specific contract with stated criteria for enrolling in CR. Students were not to enroll in CR in 9th grade and were discouraged from doing so in 10th grade. Students had to have a final grade in the course they failed of 50-59 percent (a policy that another school district also had) or have failed due to absences (i.e., they had a passing grade in the course). If the course they failed was associated with an end of course exam, they had to have earned a proficient score on the exam. The district only enrolled students in CR as part of their normal course schedule, so they preferred to enroll students in CR if they had failed at least two classes since they believed students should be able to make up two course credits in the space of one regular class period. Another district with more informal assignment policies instructed counselors to enroll students in CR if, “for the most part, they’re a decent student.” The third district had no system for enrolling students in CR. They reported that sometimes teachers requested CR assignment (sometimes mid-semester) and students also commonly requested to be in CR.

For this study, we investigated the efficacy of CR, from both public and private providers, as a tool to increase high school graduation rates for all students and for groups of students who were less likely to graduate from high school, such as students of color and economically disadvantaged students. In addition, we explored unintended consequences of CR on academic achievement. Information on selection into CR is especially important in order to separate the effects of selection into CR from the effectiveness of CR. Information from nationwide surveys, reports, and interviews, presented above, paint a complicated picture of CR enrollment. Schools could have used several signal indicators for CR enrollment including academic performance, behavior, attendance, the number of courses they failed, and grade level. Some students were assigned to take CR while others chose to enroll in CR or repeat a previously failed class. We addressed each of these assignment mechanisms in this study as described below with the goal of exploring whether CR was related to students’ course credits, test scores, and high school graduation.

Methods

Data and Sample

The data for this project included longitudinal student-level records from an administrative database, including all students enrolled in public schools in North Carolina. Student course roster files and student grade records were particularly important data files to identify students who have enrolled in CR. The sample in this study was restricted to students who failed a core, required academic course anytime in high school (see *High School Graduation Requirements*, n.d.).[2] Students who fail courses can be identified between the 2012-13 and 2016-17 school years. This sample restriction is due to CR being an intervention limited to students who previously failed a course. Including only students who failed at least one course in the sample will allow students eligible for the treatment in this study (i.e., CR) but did not receive the treatment to be compared with those students who did receive this treatment.

Within this sample, some students have been identified as taking a CR course (i.e., the treatment group), while other students repeated the course in full in a traditional fashion during a class period, after school, or in summer school, which are jointly termed “repeating a course face-to-face” (henceforth referred to as F2F).[3] Also, students who failed a course could have neither taken the course through CR nor F2F, or they could do both CR and F2F at different points in time. In this sample, it was very common for students within the same school to have had the option of taking CR or F2F. About 90 percent of the sample attended a school where students enrolled in both options, less than one percent of the sample enrolled in a school that only had CR, seven percent of the sample enrolled in schools where students only enrolled in F2F, and three percent of the sample were in schools where students didn’t enroll in either option.

Measures

Credit Recovery and Repeating a Course Face–to-Face

Enrollment in CR and/or F2F was determined using course transcript files. Students were observed to be F2F when they failed a course and repeated that same course within the same year or in subsequent school years. Courses were classified as CR when the student repeats the failed course online, as expressed by a course code indicator and through course titles. All students who enroll in CR, regardless of the provider (i.e., NCVPS or through a private provider), were included in this study although data limitations did not allow us to distinguish between CR providers.

Outcome Variables

The three dependent variables for this study were end of course exam (EOC) scores, dropping out of high school in North Carolina, and graduating from high school within four years (i.e., on-time). EOCs were offered in three subjects during this time period: Math I, English II, and Biology. Students took the EOC when they have completed the associated course. Students who were failing the course were required to take the EOC and local districts were not allowed to exempt students from taking the EOC (*READY End-of-Course Assessments: Biology, English II, NC and Math I*, 2014). Students retook the EOC every time they enrolled in an EOC-associated course (including a CR course) unless they previously earned a proficient score on that EOC. Students who earned a proficient score were allowed to retake the EOC after enrolling in CR or F2F.[4] We defined the original score as a pre-treatment variable and subsequent EOC scores as post-treatment observations. We matched EOC scores to student course records by the date the EOC was taken. If a student took both CR and F2F with in the same year, we were unable to differentiate if the EOC was taken at the end of the CR course or after F2F. However, if a student was only in CR or F2F, we are certain the EOC was taken at the end of the CR course or the F2F course that year.[5] We standardize the EOC scores by subject and year.

We included both high school graduation and dropping out because many students did not drop out but also did not finish high school within four years (about 10 percent of our sample). While graduating in four years has strong salience for accountability and reporting, students regularly exercised the option to continue high school beyond their fourth year. High school graduation was determined through the graduation file, which indicated whether a student graduated from high school in a particular year. Students were considered high school dropouts if they were no longer enrolled in a North Carolina public high school four to five years (four years for 2013-14 cohort and five years for the 2012-13 cohort) after entering ninth grade and there is no record of their having graduated from high school.[6]

Covariates

The school-level covariates were measured by year. They included student body enrollment percentages for the following student subgroups: percent classified as Limited English Proficient, percent identified as gifted, percent who qualify for special education services, percent eligible for free or reduced-price lunch, percent Black, and percent Hispanic. We also included enrollment as a school-level covariate.

All student-level behavioral and academic covariates were measured pre-CR enrollment to avoid covariates that potentially mediated the effects of CR on the outcomes, mostly during middle school. Students' prior academic performance was measured through end-of-grade test scores (standardized by test and year then averaged), enrollment in accelerated or remedial courses, grade point average, and whether or not they failed a course in eighth grade. Students were enrolled in accelerated courses if they were enrolled in a course that was designed to be offered to students at a higher grade level when in eighth grade. Whether students were enrolled in remedial courses was determined in the same fashion as accelerated courses, but for enrollment in courses designed for lower grade students. Student attendance information was included from middle school and operationalized as the percentage of school days that a student was absent in eighth grade and an indicator if the student was chronically absent that year (i.e., was absent 15 or more days). Other covariates include gender, whether the student was designated as Limited English Proficient (LEP; as of eighth grade), whether the student was gifted, whether the student was enrolled in special education services (SPED), whether the student was overage in 8^{th} grade, and their approximate age in 8^{th} grade. We included measures of student race/ethnicity—in particular, whether or not a student is Hispanic, Black, and whether the student is in a family that is economically disadvantaged for the moderation analysis and as covariates.

We included several covariates measured in high school in select analyses that were unlikely to be affected by CR enrollment (i.e., unlikely to mediate the relationship between CR assignment and the outcomes), and potentially proxy for engagement or effort. We included an indicator of the year in high school that they failed their first course (e.g., their second year in high school). We used four indicator variables, one for each of the four core subjects that they failed, which equal one to indicate they failed a course within the specific core subject. As a measure of student behavior, we included an indicator of whether the student received a behavioral infraction in the year they first failed a course. To proxy for engagement and ability, we calculated the numeric distance from earning a passing grade the first time the student failed each course and average this distance across these courses. For courses associated with an EOC, we observed the numeric grade students earned in the course and their EOC score. We included both the distance from passing for EOC courses and their EOC scores for all initial EOC enrollments (i.e., including EOC courses they passed initially and failed) as individual covariates and as an interaction. The interaction is, in essence, the EOC by how close the student was to passing.

Missing Data

These measures contained significant amounts of missing data due to the nature of this dataset and traditional limitations of administrative data. We were missing the middle school measures for students who were not attending a public school in North Carolina in the year prior to ninth grade or did not take the end of grade assessment that year. This resulted in missing between 11 and 46 percent of data on the middle school covariates. Discipline data was not available until 2015. Because we measure discipline in the year of first failure, we did not have discipline data on students who failed courses prior to 2015, 55 to 62 percent of the sample. Another pattern of missing data was for the measures of distance from passing and EOC classes/scores. About half of the data in the grades files was numeric, but the other half indicated a student's letter grade, so we did not calculate distance from passing for the latter group of students. Students might also have been missing EOC scores, although this was less common (7-20 percent of the sample). We accounted for this missing data using indicators for missing and transforming the associated variables to equal 0 when the missing indicator equals 1 for the graduation/dropout analysis only (note that we did not transform the EOC scores when this was the outcome, this missing indicator was only included in the EOC scores as a covariate in the graduation analysis). We also ran specification checks with case-wise deletion for missing data, described in the robustness checks section.

Empirical Framework

The goal of this study was to explore the relationship between enrolling in CR courses test scores, dropping out, and graduating from high school for students who failed core, required courses compared to other students who failed courses followed by F2F. The ideal empirical framework for this study would have been to randomly assign students to CR and F2F to assess the effectiveness of CR on these outcomes. Since this study used secondary data, the empirical framework was designed to remove or reduce the influence of potential confounders to the extent possible.

The empirical framework sought to address several levels of endogeneity: endogenous differences across time and between schools as well as within school and year selection bias. Students might have had differential access to CR courses due to the school they attend and the years they were enrolled in that school. The school itself might have varied CR enrollment in response to endogenous trends in the school like an escalating dropout rate. Different schools took different approaches to CR enrollment due to preferences of the leadership, school budgets, and other school-level trends that might have been correlated with the outcomes. Finally, the validity of any causal estimate is also challenged by likely student-level selection bias—in other words, the systematic selection of students who failed core, required classes into CR courses rather than F2F based on variables correlated with the outcomes. We took several approaches to address these different types of endogeneity that differed due to the two types of outcome variables, course-level outcomes and student-level outcomes.

End of Course Exam Outcome Analysis

For the EOC score analysis, the sample included all student-by-course observations for students who failed an EOC subject in 2012-13 through 2017-18. In North Carolina, EOC scores were included as part of the student’s final grade, but their actual EOC score did not determine whether the student failed the course or not (i.e., a student can pass an EOC exam and still fail the course and vice versa). Therefore, observations were included in the analysis regardless of their score on the original or subsequent EOC exams. The unit of analysis was the student-by-course-by-test date, such that each record represented one of the times when the student took an EOC exam. We included all EOC exam scores for students as long as they had originally failed the course associated with that EOC exam (e.g., if a student failed Biology, we include all of their Biology EOC scores, if this same student passed English II the first time they enrolled then we do not include their English II EOC scores). Descriptive statistics for this sample are listed in appendix Tables A1 and A2.

The analysis assessing the relationship between CR and EOC scores is estimated using the following equation:

$$(1)\ Y_{icgsy} = \beta_0 + \beta_1 CR_{icgsy} + \beta_2 F2F_{icgsy} + \beta_3 CR\&F2F_{icgsy} + \boldsymbol{\rho X_{iy}} + \boldsymbol{\tau V_{sy}} + \sigma_c + \theta_g + \gamma_y + \varphi_{is} + \varepsilon_{icsy}$$

Where Y_{icsy} represented student *i's* standardized EOC score (depending on the specification) in course *c*, in grade *g*, school *s*, and year *y*. We saturated the model with a variety of fixed effects to account for time-invariant differences across years, schools, and students. The φ_{is}term was a student-by-school fixed effect. We also included a course fixed effect (σ_c), grade fixed effect (θ_g), year fixed effect (γ_y), and two vectors of time-varying student characteristics ($\boldsymbol{X_{iy}}$) and school characteristics ($\boldsymbol{V_{sy}}$). Standard errors were clustered at the school level since treatment was distributed at this level (Abadie et al., 2017; Primo et al., 2007).

The coefficient β_1 represented the adjusted difference between enrolling in CR and the subsequent EOC exam score in comparison to the original EOC scores for that course. The coefficient β_2 represented the effect of students who enrolled a F2F course, and β_3 represented the effect of enrolling in both CR and F2F in the same year[7] in comparison to the original EOC scores by that student for that course. Because of the student-by-school fixed effect, the coefficients compared the standardized EOC score the first time the student took the exam to the score when the student originally failed the course and removed all stable student and school characteristics from the estimates including students' innate ability. Due to this specification, the model estimated within student learning differences between the original EOC score when the student failed the course and EOC scores subsequent to the original failure for the same course. We were specifically interested in comparing the coefficients β_1 and β_2 which estimated the differences between the EOC score after the intervention and the original score for that course averaged across all of the students who originally failed the course. While students were expected to earn higher EOC scores on subsequent enrollments, the difference in the coefficients for CR students and F2F students provided the estimate of the effect of CR compared to the traditional alternative treatment, F2F (significance testing using Wald tests).

Our approach addressed many of the endogeneity concerns described above since we estimated within-student estimates for the period or spell when the student was in the same school, eliminating stable student and school variation. The main endogeneity concern that remained was time-varying student and school characteristics that affected both assignment and the outcomes and were not addressed by covariates.

We cannot include indicators for race/ethnicity or economic disadvantage since these to do not vary meaningfully within student over this time period.[8] To examine our second research question, we estimated model (1) separately for subsamples of students and compare coefficients (e.g., testing whether $CR_A – CR_B = 0$ comparing models fit for subsamples A and B) and difference-in-differences (e.g., testing whether $(CR_A – F2F_A) – (CR_B – F2F_B) = 0$ comparing models fit for subsamples A and B) from these different models using Chow tests.

High School Graduation and Dropout Outcome Analysis. We were unable to analyze the high school graduation/dropout outcomes using the same empirical strategy as the EOC scores since high school graduation and permanent dropout were one-time outcomes, and student fixed effects requires at least two measurements of the outcome of interest. The analytical sample for examining high school graduation was organized as a panel data set with two cohorts of students: first-time ninth graders in the 2012-13 and 2013-14 school years. These were the only two cohorts that had enough information to include full course-taking information in high school up to an on-time graduation date four years after entering high school. The unit of analysis is at the student-school level such that each student has one unique observation for every school they attended.

In order to address endogenous differences between cohorts and between schools, the models included cohort-by-school fixed effects such that only students who failed courses within the same cohort within the same school were compared to each other. Cohort-by-school fixed effects accounted for between-cohort and school factors such as the availability of CR courses, the likelihood of assignment-to-CR based on the school in which a student was enrolled, and the availability of CR courses based on the year that could have correlated CR enrollment with the outcomes of interest.

In this sample, we defined the treatment as enrolling in at least one CR course at any time in high school. We focus our comparison on students who only enrolled in F2F although we include indicators for students who do CR & F2F and neither option in our models. To see a breakdown of the sample into these groups by number of CR and F2F courses students enrolled in, see appendix Table A3. We addressed within-cohort and school selection bias by implementing weighted propensity scores, following a technique known as marginal mean weighting through stratification (MMWS) adjustment to estimate propensity scores across the four treatment/comparison conditions (CR, F2F, CR&F2F, and Neither). Rosenbaum and Rubin (1983) demonstrated through a proof that these probabilities of assignment can produce a study sample in which assignment to treatment is independent of the outcome in that sample, based on an assumption known as strong ignorability. According to Stuart (2010), the likelihood that the assumption held true using a matching procedure is more reasonable when the covariates that predict both assignment and the outcomes of interest were identified and included in the estimation of propensity scores.

The goal was to weight the sample such that, conditional on the covariates, the students in the treatment and comparison conditions were as similar as they would be if treatment assignment were randomized. Following the advice of Stuart, the covariates that this study included were highly correlated with assignment to CR and the outcomes of interest. The matching procedure reduced bias in assignment to treatment to the extent that the correlation between CR assignment and the outcomes of interest were controlled by the included covariates. We compile a robust set of covariates associated with taking a CR course and with dropout/graduation for the matching procedure. Basic student variables that prior data have shown to be related to high school graduation, such as student race/ethnicity, gender, and an indicator of economic disadvantage, were included. Also included were key indicators of the students' learning capabilities that are highly predictive of high school graduation and were likely taken into account when assigning students to CR courses: whether the student had a disability, whether the student was gifted, whether the student was Limited English Proficient (LEP), their approximate age, and whether they were overage in eighth grade.

We included a wide set of variables from eighth grade to attempt to account for the various pieces of information that schools likely used when assigning students to CR. We used several measures of academic success including eighth grade test scores, GPA, failing a course, enrollment in advanced courses, and enrollment in remedial courses. Students with lower academic ability might be discouraged from taking CR courses because these students needed more guidance and scaffolding than an online platform could provide. Students who enrolled in accelerated courses were more likely to be college-bound and F2F enrollment results in their original failing grade being expunged from their transcript (but would not be expunged for CR enrollment). The eighth grade GPA was expected to predict both of the outcomes, since this was an indicator of both motivation and academic ability. Failing a course in eighth grade may also be measures of students' motivation and academic ability. Student attendance and chronic absenteeism in eighth grade were the final covariates included in the matching procedure. Attendance might also be an important assignment mechanism, since students who have poor prior attendance might be seen as being a better fit for a flexible, online platform than a structured, traditional course.

We included several covariates measured in high school as long it was very unlikely these covariates were endogenous to assignment to CR. As indicated in the interviews with school districts, when a student failed a course (i.e., earlier or later in high school) had salience for CR assignment, so we included year of first failure. We also included an indicator for being involved in a disciplinary incident. Since discipline could be an assignment mechanism for CR, we measure discipline in the year of first failure in high school with an indicator of whether the student received a behavioral infraction.

As a robustness check, we included the other covariates measured in high school (see the Measures section) in the matching procedure. While we cannot guarantee they are not mediated by CR assignment, we construct these variables to lessen this concern. We included distance from passing for first time failure, distance from passing for first time enrollments in all EOC courses, average first-time EOC scores, and an interaction between the latter two variables as measures of ability, effort, and engagement.

To empirically examine the extent to which our strong identification assumption was likely to hold, we assessed whether these covariates were correlated with treatment assignment and the outcomes in two ways. First, as shown in appendix table B1, we examined within cohort-school correlations between the covariates, treatment assignment, and the two outcomes (see Steiner et al., 2010). For CR, 23/40 covariates had significant correlation coefficients with CR assignment within school-cohort. The comparison groups had similarly high proportion of the variables with significant correlations (29/40 for F2F, 23/40 for the Neither condition, 19/40 for CR & F2F). All of the covariates have significant correlations with graduating high school with only three being insignificant for dropping out. In addition, we compared a naïve linear probability model predicting CR assignment with no covariates to a linear probability model predicting CR assignment with the covariates (both models include school-cohort fixed effects). The R^2 in the naïve model with graduation as the outcome is 0.058, and the R^2 in the model with covariates is 0.141, a noticeable improvement in the R^2 between models, even allowing for the attenuation of the R^2 due to the binary dependent variable.

Our matching approach used the marginal mean weighting through stratification (MMWS) adjustment. MMWS combines propensity score and stratification methods of matching and allowed us to calculate propensity scores for multiple treatment/comparison groups (Hong, 2010, 2012; Linden, 2017). The MMWS procedure began by estimating propensity scores through a multinomial logistic regression model then stratified the sample based on the propensity score. We then calculated a marginal mean weight based on the propensity score and stratification procedure that was designed to reweight the sample to represent the proportion of the sample we would expect to be assigned to each condition if treatment was assigned randomly using the mmws command in Stata (Linden, 2014). We then ran an OLS model with school-by-cohort fixed effects and full covariates weighted by this marginal mean weight.

We assessed balance on the sample reweighted by the MMWS procedure using two metrics based on the advice of Stuart (2010): standardized difference in means and the ratio of standard deviations (SD) between the treated and comparison groups. According to Stuart, the difference in standardized means should all be less than 0.25. The ratio of SD should be between 0.5 and 2. In Table 1, we show the within-group means and SD for each covariate with the first line including information on CR students and the second line being students in the F2F group (full balance checks comparing balance on all four treatment/comparison groups are listed in appendix Table B2) for each covariate pre-matching and post-matching. As shown in Table 1, the CR and F2F groups were already very similar to one another prior to matching with only two covariates having standardized differences larger than 0.25 (First Failed in 4th Year and Missing Discipline Data). These differences were much smaller post-matching with an average standardized difference pre-matching of 0.067 and an average standardized difference post-matching of 0.018. The SD ratios post-matching are also very close the one with the smallest ratio being 0.939 and the largest 1.287.[9]

We also assess balance using a variety of covariates that we did not include in the matching procedure. Covariates included in MMWS were correlated with treatment assignment and the outcomes, but the rich administrative dataset we have access to has many other covariates that we can use to assess balance post-matching. We create a variety of variables using data on students' middle schools, their middle school classmates (i.e., in individual classes using rosters), and middle school teachers. We find in this exercise that the groups were already very similar pre-matching (average standardized difference of 0.063) and are even more similar when weighted (average standardized difference of 0.055; see the appendix Table B3 for the results of this balance check). These weights produced balanced samples based on the metrics suggested by Stuart (2010) on covariates included in matching and covariates not included in matching.

The final models were linear probability models that included cohort-by-school fixed effects with all of the covariates used in matching with probability weights from the MMWS procedure. The model takes the following form,

$$(2)\ Y_{csi} = \beta_0 + \beta_1 CR_{csi} + \beta_2 CR\&F2F_{csi} + \beta_3 Neither_{csi} + \boldsymbol{\rho X_i} + \delta_{cs} + u_{csi}.$$

The outcome, Y_{csi} represented whether student *i* in school *s* and cohort *c* graduated from high school or dropped out of high school, depending on the specification. The model included a vector of individual student characteristics, $\boldsymbol{X_i}$ (see appendix Table B6 for a full list of these covariates broken out by treatment/comparison group), as well as a cohort-by-school fixed effect, δ_{cs}, and an error term, u_{csi}. All models were then fitted with interactions between the treatment/comparison variables and the variable indicating that a student was Black, Hispanic, or economically disadvantaged. Given that we had coefficients for CR, CR&F2F, and Neither the comparison was to students who do only F2F. Standard errors were clustered at the school level

Limitations

While our goal was to reduce endogeneity to estimate a plausibly unbiased treatment effect of CR on the outcomes, several different sources of endogeneity may not be fully addressed. In the EOC score student-by-school fixed effect analysis, we accounted for non-time-varying student and school characteristics but only accounted for time-varying student and school characteristics by adjusting for the time-varying covariates we include in the model. If something about the student or school changed substantially over time (in ways not accounted for by the year and grade fixed effects) but related to both assignment to CR and the course-level outcomes, then these changes could lead to bias.

For the analysis of graduation and dropout, we are aware of additional concerns for potential endogeneity. For instance, if students were assigned to CR courses in schools in a highly-personalized manner, such that two students who looked identical according to their data (e.g., same middle school attendance, test scores, and grades) were assigned to CR based on specific information about those students that was uncorrelated with the available covariates and also predicted their high school graduation propensity, then the treatment effect estimates would be biased. The extent to which this was true is unknown and may vary across and within schools. We reviewed above some of the reasons students were assigned to CR. We accounted for most of these explanations using available covariates. Likely, students were assigned to CR in some mixture of random and highly purposeful ways, with some schools perhaps using one approach or the other, while others mix the two. However, purposeful assignment only biased the estimates to the extent that this purposeful assignment was not correlated with any included covariate like gender, test scores, or attendance. In addition, a continuing debate among causal inference researchers asks whether matching or propensity scores represent a substantial improvement in bias reduction over ordinary least squares regression, particularly when matching without a pre-test measure (Angrist & Pischke, 2009; Cook et al., 2008). Considering pre-test measures of one-time events like graduation do not exist, it is possible that the matching process does not lessen bias beyond what we would observe using covariate-adjustment alone.

We were also limited in our ability to evaluate the overall effect of CR due to our focus on core courses only. Students often fail elective courses and CR versions of these courses were available. We limit our analysis to core courses because of their clearer connection to graduation requirements and due to limitations in our ability to identify the hundreds of different elective courses offered in NC. This limited our inferences to CR coursetaking for math, science, English, and social studies core courses.

Results

We display summary statistics including average values on the outcomes and observations sizes by treatment/comparison status in both analytical samples in Table 2. When comparing initial to post-enrollment EOC scores, CR students' standardized EOC scores stayed the same at -0.79 from their first time taking the course to after they enrolled in CR. F2F students had a 0.09 standardized unit improvement in EOC scores from their initial enrollment (-0.73) to post-F2F enrollment (-0.64). While the EOC analysis is at the course enrollment level, the high school graduation analysis is at the student level. As is shown in Table 2, CR students enrolled in an average of 1.71 CR courses during high school, and F2F students enrolled in an average of 1.91 F2F courses (for a more detailed breakdown, see appendix Table B5). For high school graduation and dropout, a higher proportion of CR students graduated from high school than F2F students (0.74 versus 0.63, respectively), with these results mirrored for dropping out of high school.

Results from student-by-school fixed effects models with EOC scores as the outcomes are listed on Table 3 (see appendix Table D1 for the coefficients on the CR&F2F variable). The first panel shows the results for all EOCs, and the first column has the coefficient for the full sample. Taking a CR course was estimated to increase EOC scores by 0.048 standard deviation units compared to their other EOC exams, and F2F students were estimated to have a 0.165 standard deviation increase in their EOC scores. The difference between these coefficients is over a tenth of a standard deviation and is statistically significant (0.117, Wald test $p<0.001$). We show the results for the student-by-school fixed effects models restricted to subgroups of students by economically disadvantaged status and race/ethnicity in columns 2-6. Again, each of the five student subgroups had significantly lower test scores for CR compared to F2F of a similar magnitude as the overall average. We found no significant differences between the CR coefficients across models or difference-in-differences in the effect of CR compared to F2F for these subgroups using Chow tests.

We next examined the results by EOC subject. In panel B for the Math I EOC, the effect of CR on the Math I EOC scores is smaller and not statistically significant, and the F2F coefficient retains statistical significance. The difference between the CR and F2F coefficients in the Math I models is just below a tenth of a standard deviation (-0.083, Wald test $p=0.021$). The only subgroup with significant differences between CR and F2F is Hispanic students with over a tenth of a standard deviation difference in scores, but none of the difference-in-differences across models were statistically significant according to Chow tests.

As is shown in panel C, the coefficient on CR for the English II EOC is very close to zero, but the coefficient on F2F is close to the magnitude from the full sample, showing a 0.147 difference in learning gains between CR and F2F (Wald test $p=0.001$). The difference between

CR and F2F is significant for the sample of economically disadvantaged (ED) students (-0.155, Wald test p=0.010) and Black students (-0.168, Wald test p=0.005), but not statistically significant for the other student subgroups. The difference-in-differences between CR and F2F for Black students compared to Non-Black, Non-Hispanic students is not significant (Chow test p=0.165) and these difference-in-differences are not significant when comparing ED to non-ED students (Chow test p=0.460). While it might appear that effects vary across student subgroup for English II, the Chow tests did not find significant difference-in-differences between these models.

While the coefficient on CR is similar in Panel D for the Biology EOC compared to the CR coefficient for the Math I EOC, the coefficient on F2F is larger for the Biology EOC. The difference between the CR and F2F coefficients for the Biology EOC is now almost two tenths of a standard deviation (-0.191, Wald test p=0.004). Just as in the case of English II, the difference between CR and F2F is only statistically significant for ED students (-0.209, Wald test p=0.024) and Black students (-0.216, Wald test p=0.040), but these difference-in-differences are not statistically significant compared to non-ED (Chow test p=0.810) and non-Black, non-Hispanic (Chow test p=0.423) students. While F2F course enrollments are consistently associated with larger learning gains, the difference in learning gains is smallest for the Math I EOC and largest for the Biology EOC. We did not find evidence that these negative effects are more or less pronounced for certain subgroups of students.

The results for the school-by-cohort fixed effects analysis with weighting are shown on Table 4 (for coefficients on CR&F2F and Neither, see appendix Table E1).[10] Referring to columns (1) and (5), we find enrolling in CR appears to be beneficial for high school graduation and drop out. Compared to students who only remediate courses F2F, students in CR are predicted to 7.2 percent more likely to graduate from high school within four years and 6.1 percent less likely to drop out of high school.

The results of the moderation analysis are shown in Table 4 for the covariates of interest only. Columns 2 and 6 show the results for interactions for Black students, columns 3 and 7 show the results for the interactions for Hispanic students, and columns 4 and 8 show the results for the interactions for ED students. To ease interpretation of these interaction terms in the models estimating the association between CR and graduating from high school, refer to Figure 1 (the equivalent figure for dropping out of high school is almost identical). This figure has the predicted probabilities of graduating from high school from the models including interaction terms between treatment/comparison status and race/ethnicity or ED status. As is shown in Figure 1, CR students, regardless of the subgroup, are predicted to be more likely to graduate from high school. These differences are particularly pronounced for Hispanic students and ED students which had significant interaction terms in the models (see Table 4). Hispanic F2F students have a predicted probability of graduating of 0.59 and Hispanic CR students have a predicted probability to graduating of 0.69 (Wald test $p<0.001$). ED F2F students have a predicted probability of graduating of 0.71 compared to 0.79 for ED CR students (Wald test $p<0.001$). Black CR students also are shown to be more likely to graduate high school (0.77) than Black F2F students (0.70; Wald test $p<0.001$). Note that ED CR students have the same predicted graduation probability as non-ED F2F students (0.79; Wald test $p=0.79$). F2F Hispanic students have the lowest predicted probability of graduating (0.59), but CR Hispanic students have an slightly higher, but not significantly, predicted probability of graduating (0.69) compared to White CR students (0.66; Wald test $p=0.73$).

Robustness Checks

Several robustness checks were performed to see if the results were sensitive to the specification of the model and the matching strategy, all of which are available in the appendix as supplemental tables. First, we assessed if the results could be an artifact of the different samples of students across the two analyses or the different method and covariates used to estimate each model (also see appendix Table A5 for a descriptive comparison of the samples of CR students across analyses). In appendix Table C1, we show the results of estimating a school-by-cohort fixed effect model with the same covariates used in the high school graduation analysis for all outcomes. We found the coefficients on CR and F2F were all similar in magnitude, direction, and statistical significance as the estimates in Tables 3 and 4.

As an additional check on the sensitivity of the estimates to different modeling specifications, we conducted an identical matching process with the EOC sample as we did with the high school graduation sample. We then estimated the original models we used for those outcomes but weighted with these propensity score first with student-by-school fixed effects and then with school-by-cohort fixed effects. As is shown in appendix Tables D1 and D2, the results were very similar to the original estimates.

We assessed the sensitivity of the estimates to a couple of basic modeling choices. First, we estimated models with binary outcomes using logistic regression. As is shown on appendix Table E2, all results for graduating from high school and dropout maintain the same statistical significance levels and direction. Second, we compared estimates of the results for the high school graduation and dropout models that allow for case-wise deletion compared to the main estimation results that used a missing indicator for variables with systematically missing data, see appendix Table E3. The results with case-wise deletion were smaller in magnitude but in the same direction and statistical significance level as the main results.

We included a couple of robustness checks on the EOC analysis to see if the results were sensitive to the sample of students. First, we restricted the sample to only students who had at least two EOC scores in each course, eliminating students who did not take an EOC exam after a CR or F2F course. Second, students who scored proficient on their first EOC were not required to take it again when they retook the class through CR or F2F, so we restricted the sample to only include students who did not score proficient on their initial EOC exam. As is shown in appendix Tables D4 and D5, these different samples produced results that are very similar to the results using the full sample.

The final robustness check assessed if the EOC results were due to the specific nature of the EOC exams. Since the 2012-13 school year, all students enrolled in 11^{th} grade are required to take the ACT. For students with ACT scores, about 90 percent of the sample took the ACT in their third year of high school (as expected) with almost all of the remainder taking the ACT in their fourth year of high school. However, since this study's sample is comprised of students who fail courses, members of the sample have a high likelihood of dropping out during high school, and only 63 percent of the sample took the ACT. As might be expected based on the results thus far, CR students were more likely to persist to 11^{th} grade and therefore take the ACT. For this reason, we did not include ACT scores in our main set of results, but we use ACT scores as an outcome as a robustness check on the EOC test score results. We excluded students from the analysis if they did not take CR or F2F until the year after they took the ACT. We use the graduation analysis sample and methodology to conduct this robustness check as ACT scores are typically a one-time event in this sample. As is shown on appendix Table D6, CR students are predicted to have ACT scores that are 0.22 standard deviation units lower than F2F students, very similar results of the EOC analysis.

We now discuss several additional robustness checks for the high school graduation/dropout analysis. As was mentioned in the Empirical Framework section, we had access to additional variables measured in high school that we did not include in the main specifications because they could be potential mediators between CR assignment and the outcomes. These variables include average initial EOC scores, distance from passing, distance from passing for EOC courses, whether they failed a course in each of the four subjects, and an interaction of the EOC scores and distance from passing for EOC courses. As a robustness check, we reran the matching procedure including these variables with the original set of covariates. As is shown in appendix Table E4, the CR and F2F groups originally had large standardized differences on several of these additional covariates, specifically the rates of failing courses in each subject area, but post-matching all differences and standard deviation ratios were within the acceptable bounds. As shown on appendix Table E5, the results were almost identical to what we found when estimating these results without these additional covariates.

It's possible that the positive association between CR and high school graduation was a result of some students having more access to CR courses and not enrolling in CR courses themselves. To test this potential threat, we restrict the sample to only students who had access to CR courses since their first ninth grade year of high school. As is shown on appendix Table E6, the results were almost identical to the main estimates.

To test the robustness of our results to an alternative matching approach, we followed the advice of Iacus, King, and colleagues and combined exact matching, coarsened exact matching, and matching based on a distance measure (Iacus et al., 2019, 2019; King & Nielsen, 2019). We exact match on receipt of a disciplinary infraction, cohort, and school. We then took a coarsened exact matching approach by matching students on the quintile of their average eighth grade test scores (dropping those who didn't match with students in their comparison group). The students were then matched on the full set of covariates using a Mahalanobis distance measures instead of propensity score measures (King & Nielsen, 2019) using all available covariates described above. Students who only were F2F were excluded from the sample if they are not one of the top five matches, according to the Mahalanobis distance, for any member of the treatment group. We then estimated models using this matched sample with all of the covariates from the original model. We conducted this process twice once with case-wise deletion and once with missing indicators (including the missing indicators in the matching procedure). As is shown on Table E7, the results were similar to our preferred matching approach. The exception is the coefficient on CR for graduating high school in the case-wise deletion version which loses statistical significance and changes direction, but this results could be a remnant of restricting the sample to only students with full data who also met the strenuous matching requirements.

One threat to causality was the potential endogeneity stemming from students who fail multiple courses when assignment to CR was related to the number of courses they fail as well as the outcomes. We restrict the sample to only students who have failed one class; results in appendix Table E8. After matching within the sample of students who only failed one class, the results are slightly smaller but retain statistical significance and the same direction.

The final robustness check explored one source of plausibly exogenous variation in CR enrollment: access to CR based on whether the school offered CR to that cohort. In 52 schools (eight percent of schools), one cohort had access to CR courses (i.e., at least one student enrolled in CR) while the other cohort in the same school did not appear to have access to CR. We estimated a model with school fixed effects, cohort fixed effects, and student-level covariates where the treatment was defined as having CR as an option for that cohort (at the cohort-school level) restricted to only schools that had CR for one cohort and not for the other. The benefit of this approach was that we were not as concerned about student-level selection mechanisms when it appeared that students either did or did not have access to CR as a rule instead of on a discretionary basis. The results were an intent to treat analysis since we estimated the association of being in a cohort with access to CR and the outcomes, not actual receipt of CR. Results are shown in appendix Table E9. Overall, the results were relatively similar although not statistically significant. The coefficient on being in a cohort with access to CR for the graduation rate outcome was similar to (but smaller in magnitude) the main set of results. The coefficient on the dropout outcome was close to 0 and not significant. These results provided some evidence that access to CR could be correlated with graduation probability.

Conclusion

Students who fail courses in high school are commonly faced with two options to make up for that failed credit: F2F (over the summer, after school, or during the school year) or retake the course through CR. While many have hypothesized that CR courses have contributed to the rising high school graduation rate, this study shows empirically that CR is associated with higher probability of high school graduation for students who fail core courses. Using carefully selected methods that account for most likely sources of bias, this study finds that CR lead to a higher likelihood of graduating and a lower likelihood of dropping out of high school compared to students who fail courses and were F2F. These results are especially pronounced for Hispanic students and ED students. We share evidence that CR could help to create more equitable outcomes for Hispanic and ED students for graduating from high school. Hispanic CR students have equivalent predicted probabilities of graduating as their non-Black, non-Hispanic CR student peers, and ED CR students have very similar predicted graduation probabilities compared to non-ED F2F students. Since graduation rates for Hispanic and ED students are increasing at a faster rate than White or non-ED students, CR could be part of why this is occurring.

However, CR appears to reduce student learning compared to F2F as measured by EOCs, especially the EOCs in English and Biology, and, in robustness checks, the ACT. The negative effect sizes for these assessments could be described as medium to large effects (range is -0.22 to -0.08), showing that these negative associations are practically significant (Kraft, 2020). These negative associations for test scores appear to be equally shared by students of different race/ethnicities and socioeconomic status. In some ways, CR might be allowing traditionally marginalized students to have a higher probability of graduating while not exacerbating inequality in other ways, at least compared to their traditionally advantaged peers who enroll in CR.

These results confirm much of what others found when evaluating the effects of CR on students' outcomes. For instance, a prior study in NC found NCVPS CR students had higher on-time gradation probabilities of seven percent (our effect size is also seven percent) compared to students who did F2F and non-NCVPS CR (Stallings et al., 2016). Our results also mirror findings from Florida which found CR students had a seven percent higher probability of remaining in high school through the spring of 12th grade (their proxy for graduating; Hart et al., 2019). A study of a large, urban school district also found students possibly enrolled in CR (i.e., they had failed a course in the year before taking a virtual course) were 13 percent more likely to graduate high school than F2F students, a finding in the same direction but larger than the findings in this study. However, this study did not find the students potentially enrolled in CR had significantly lower test scores than F2F students (Heinrich & Darling-Aduana, 2019). While the RCT on CR in Chicago initially found CR students had lower posttest scores in Algebra I than F2F students (-0.19, twice as large as what we found for Math I EOCs but similar in size to what we found for Biology), they found no effect on high school graduation (Heppen et al., 2016; Rickles et al., 2018). We confirm much of what has been found in observational studies of CR, but our findings differ from the RCT. Our study differs from the RCT in that students typically do not only take one CR or F2F course (as is shown in appendix Table A3, 44 percent of the sample takes more than one of either CR or F2F) and take CR in courses other than Algebra I (equivalent to Math I in NC). This does leave open the possibility though that CR can have differing effects based on the context and course such that future studies can continue to investigate how to promote the positive aspects of CR while lessening the negative.

While the empirical design of this study seeks to approximate random assignment, we are unable to rule out the possibility that these results are simply due to students with a higher likelihood of graduating from high school on time and lower ability (as measured through test scores) enrolling in CR. While we include data on many of the indicators school are likely using to assign students to CR, it is difficult without more information to ascertain other potential assignment mechanisms that have not yet been written about in the literature or were discussed in the interviews with North Carolina school districts. Fortunately, this study was able to account for the vast majority of the known CR assignment mechanisms.

This study is part of building a literature base on CR, whether it helps students to graduate, but possibly at the expense of their learning. From these results, we can hypothesize that CR courses might tend to be overly simplistic, lack rigor, and/or lack the same content as traditional courses. This study finds lends empirical support to much of what has been said by critics of CR courses, but should be considered an exploratory study to prompt further study.

Future research should continue to test these hypotheses both in other states and with more years of data from NC. While we now have evidence on the effectiveness of CR in a few different settings (Hart et al., 2019; Heinrich & Darling-Aduana, 2019; Rickles et al., 2018), which should help to justify a randomized experiment to give schools and policymakers a clearer understanding on the effectiveness of CR in helping students graduate and possible tradeoffs in student learning.

However, the relative merits of CR have not been explored in this study. Graduating from high school might have greater benefits, both individually and societally, than small to medium changes in test scores. In addition, students with more problematic behaviors in the classroom might benefit from an online environment with potential positive spillovers to their peers in traditional classrooms. At the same time, the potential lower levels of learning in CR could lead to additional challenges in more advanced coursework in high school and in higher education. This pattern of increasing graduation rates potentially at the expense of student learning might potentially erode the economic signal provided by high school graduation. These kinds of questions about what the high school credential does and should signify are currently playing out in schools, as they decide whether to enact policies that might increase their graduation rates to the detriment of scores on high-stakes assessments. This study gives policymakers and school leaders more empirical information with which to make these kinds of decisions.

Endnotes

[1] Note that these categories were not mutually exclusive.

[2] Regardless of year or track, students were required to complete four years of English, at least three math courses, Biology with at least one other science course, and a minimum of two social studies courses. Go to https://www.dpi.nc.gov/districts-schools/high-school-graduation-requirements for more information.

[3] We used full course enrollment files to determine if a student enrolled in CR or repeating a course for credit. The course enrollment files included course codes, online enrollment indicators, CR indicators, full course names, and a separate file contains final grades (by semester).

[4] In this sample, 11 percent of students who failed an EOC subject earned a proficient score the first time they had taken the exam. Seven percent of students who already scored proficient on an EOC retook the EOC

[5] We were unable to differentiate whether the EOC was associated with CR or F2F for 1.7 percent of the sample or 3,344 tests. Seven percent of the sample or 13,485 tests were associated with CR courses, and 33.1 percent of the sample or 63,469 tests were for F2F courses. The remainder of the sample represents initial course enrollments. See Table 2.

[6] We exclude students from the graduation and dropout analysis who officially transferred out of North Carolina Public Schools, were expelled, left the country, or had a serious illness.

[7] Ideally we would be able to differentiate between whether a student took the EOC exam immediately after their CR course or immediately after their F2F course. Because the transcript files do not necessarily give specific or accurate dates of course enrollment within year (particularly for online courses), we are unable to differentiate EOC exam scores between CR and F2F courses when a student does both within the same year.

[8] Economic disadvantage status is included as a time-varying covariate in the student-by-school fixed effect models. If we were to examine the coefficient on this covariate, it would indicate the association between changing economic disadvantage status across time while the student was in high school. Since we are interested in the broader implications of socioeconomic disadvantage, we treat this covariate as if it were fixed for our subgroup analysis.

[9] As is shown in the appendix Table B2, the CR group had some large standardized differences pre-matching with the CR&F2F and None (i.e., neither CR nor F2F) groups on several variables, but all of these differences shrunk below the 0.25 threshold with SD ratios close to one post-matching.

[10] The coefficient on the variables indicate that being Black or Hispanic is associated with a higher odds of graduating from high school (just Black students) and lower probability of dropping out of high school (both subgroups). While this result is unusual given the literature on the chronic under-attainment of students of color, within the sample of students who fail courses, the data indicate that these students have higher graduation rates and lower dropout rates.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). *When should you adjust standard errors for clustering?* National Bureau of Economic Research.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion* (Vol. 1). Princeton University Press.

Approved Vendor List. (n.d.). North Carolina Virtual Public School.

Bowers, A. J. (2010). Grades and Graduation: A Longitudinal Risk Perspective to Identify Student Dropouts. *The Journal of Educational Research*, *103*(3), 191–207. https://doi.org/10.1080/00220670903382970

Burke, K., Chapman, B., & Monahan, R. (2013, September 24). Critics Blast Credit Recovery as City Data Reveals Frequent Use by Public High School Students. *New York Daily News*.

Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management*, *27*(4), 724–750. https://doi.org/10.1002/pam.20375

Cooper, H., Charlton, K., Valentine, J. C., Muhlenbruck, L., & Borman, G. D. (2000). Making the most of summer school: A meta-analytic and narrative review. *Monographs of the Society for Research in Child Development*, i–127.

Dropout Prevention Intervention Report: Credit Recovery Programs. (2015). U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse.

Dynarski, S. (2018, January 21). Online Courses Are Harming the Students Who Need the Most Help. *The New York Times*. https://www.nytimes.com/2018/01/19/business/online-courses-are-harming-the-students-who-need-the-most-help.html

Edmentum Courseware. (2020, January 7). https://www.edmentum.com/products/courseware

Fong, A. B., Jaquet, K., & Finkelstein, N. (2014). *Who Repeats Algebra I, and How Does Initial Performance Relate to Improvement When the Course Is Repeated?* (REL 2015-059). Regional Educational Laboratory West. http://eric.ed.gov/?id=ED548534

Gardner, W. (2016, August 10). *Summer School Still the Best Way to Catch Up*. Walt Gardner's Reality Check - Education Week. http://blogs.edweek.org/edweek/walt_gardners_reality_check/2016/08/summer_school_is_still_best_way_to_catch_up.html?cmp=eml-enl-eu-news3

Gemin, B., & Pape, L. (2017). *Keeping Pace with K-12 Online Learning 2016*. Evergreen Education Group. https://static1.squarespace.com/static/59381b9a17bffc68bf625df4/t/593efc779f745684e6ccf4d8/1497300100709/EEG_KP2016-web.pdf

Hart, C. M. D., Berger, D., Jacob, B., Loeb, S., & Hill, M. (2019). Online Learning, Offline Outcomes: Online Course Taking and High School Student Performance. *AERA Open*, *5*(1), 1–17. https://doi.org/10.1177/2332858419832852

Heinrich, C. J., & Darling-Aduana, J. (2019). *Does Online Course-taking Increase High School Completion and Open Pathways to Postsecondary Education Opportunities?* (EdWorkingPapers). Annenberg Institute for School Reform at Brown University.

Heppen, J. B., Sorensen, N., Allensworth, E., Walters, K., Rickles, J., Taylor, S. S., & Michelman, V. (2016). The Struggle to Pass Algebra: Online vs. Face-to-Face Credit Recovery for At-Risk Urban Students. *Journal of Research on Educational Effectiveness,* 00–00. https://doi.org/10.1080/19345747.2016.1168500

High School Graduation Requirements. (n.d.). http://www.ncpublicschools.org/docs/curriculum/home/graduationrequirements.pdf

Holme, J. J., Richards, M. P., Jimerson, J. B., & Cohen, R. W. (2010). Assessing the Effects of High School Exit Examinations. *Review of Educational Research*, *80*(4), 476–526. https://doi.org/10.3102/0034654310383147

Hong, G. (2010). Marginal Mean Weighting Through Stratification: Adjustment for Selection Bias in Multilevel Data. *Journal of Educational and Behavioral Statistics*, *35*(5), 499–531. https://doi.org/10.3102/1076998609359785

Hong, G. (2012). Marginal mean weighting through stratification: A generalized method for evaluating multivalued and multiple treatments with nonexperimental data. *Psychological Methods*, *17*(1), 44. https://doi.org/10.1037/a0024918

Howard, K., Romero, M., Scott, A., & Saddler, D. (2015). *Success after Failure: Academic Effects and Psychological Implications of Early Universal Algebra Policies*. http://digitalcommons.chapman.edu/education_articles/46/

Iacus, S. M., King, G., & Porro, G. (2019). A Theory of Statistical Inference for Matching Methods in Causal Research. *Political Analysis*, *27*(1), 46–68. https://doi.org/10.1017/pan.2018.29

Issue Brief: Credit Recovery. (2018). U.S. Department of Education Office of Planning, Evaluation and Policy Development Policy and Program Studies Service. https://www2.ed.gov/rschstat/eval/high-school/credit-recovery.pdf

King, G., & Nielsen, R. (2019). Why Propensity Scores Should Not Be Used for Matching. *Political Analysis*, 20.

Kraft, M. A. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, *76*(2), 275–313.

Levine, E., Johnson, J., Malave, J., & Santaniello, S. (2017). *Online courses for credit recovery in high schools: Effectiveness and Promising Practices*. Nellie Mae Education Foundation, UMass Donahue Institute.

Linden, A. (2014). *MMWS: Stata module for implementing mean marginal weighting through stratification*. http://ideas.repec.org/c/boc/bocode/s457886.html

Linden, A. (2017). Improving causal inference with a doubly robust estimator that combines propensity score stratification and weighting. *Journal of Evaluation in Clinical Practice*, *23*(4), 697–702.

Mac Iver, M. A., & Messel, M. (2013). The ABCs of Keeping On Track to Graduation: Research Findings from Baltimore. *Journal of Education for Students Placed at Risk (JESPAR)*, *18*(1), 50–67. https://doi.org/10.1080/10824669.2013.745207

Murnane, R. J. (2013). U.S. High School Graduation Rates: Patterns and Explanations. *Journal of Economic Literature*, *51*(2), 370–422. https://doi.org/10.1257/jel.51.2.370

National Center for Education Statistics. (n.d.). http://nces.ed.gov/

Neild, R. C., Stoner-Eby, S., & Furstenberg, F. (2008). Connecting entrance and departure the transition to ninth grade and high school dropout. *Education and Urban Society*, *40*(5), 543–569.

No Child Left Behind High School Graduation Rate Non-Regulatory Guidance. (2008). U.S. Department of Education.

Overview Information; Race to the Top Fund; Notice Inviting Applications for New Awards for Fiscal Year 2010; Notice. (2010). U.S. Department of Education.

Papay, J. P., Murnane, R. J., & Willett, J. B. (2010). The Consequences of High School Exit Examinations for Low-Performing Urban Students: Evidence From Massachusetts. *Educational Evaluation and Policy Analysis*, *32*(1), 5–23.

Plunk, A. D., Tate, W. F., Bierut, L. J., & Grucza, R. A. (2014). Intended and unintended effects of state-mandated high school science and mathematics course graduation requirements on educational attainment. *Educational Researcher*, 0013189X14540207.

Primo, D. M., Jacobsmeier, M. L., & Milyo, J. (2007). Estimating the impact of state policies and institutions with mixed-level data. *State Politics & Policy Quarterly*, *7*(4), 446–459.

Public High School Graduation Rates. (2020, May). The Condition of Education. https://nces.ed.gov/programs/coe/indicator_coi.asp

READY End-of-Course Assessments: Biology, English II, NC and Math I. (2014). Public Schools of North Carolina, State Board of Education, Department of Public Instruction.

Reardon, S. F., Arshan, N., Atteberry, A., & Kurlaender, M. (2010). Effects of Failing a High School Exit Exam on Course Taking, Achievement, Persistence, and Graduation. *Educational Evaluation and Policy Analysis*, *32*(4), 498–520. https://doi.org/10.3102/0162373710382655

Rich, M. (2015, December 26). As Graduation Rates Rise, Experts Fear Diplomas Come Up Short. *The New York Times*.

Rickles, J., Heppen, J. B., Allensworth, E., Sorensen, N., & Walters, K. (2018). Online Credit Recovery and the Path to On-Time High School Graduation. *Educational Researcher*, 0013189X18788054. https://doi.org/10.3102/0013189X18788054

Silver, D., Saunders, M., & Zarate, E. (2008). *What Factors Predict High School Graduation in the Los Angeles Unified School District* (No. 14; California Dropout Research Project). UCLA/IDEA and UC/ACCORD.

Smiley, S. (2017, May 23). Fast. Isolating. Superficial. *Slate*. http://www.slate.com/articles/news_and_politics/schooled/2017/05/what_class_is_like_for_high_schoolers_taking_their_courses_online.html

Stallings, D. T., Weiss, S. P., Maser, R. H., Stanhope, D., Starcke, M., & Li, D. (2016). *Academic outcomes for North Carolina Virtual Public School credit recovery students* (REL 2017-177). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. http://ies.ed.gov/ncee/edlabs

Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods, 15*(3), 250. https://doi.org/10.1037/a0018719

Stuart, E. A. (2010). Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science, 25*(1), 1–21. https://doi.org/10.1214/09-STS313

Turner, C. (2015, June 10). Raising Graduation Rates with Questionable Quick Fixes. *NPR*.

Tyner, A., & Munyan-Penney, N. (2018). *Gotta Give 'Em Credit: State and District Variation in Credit Recovery Participation Rates* (p. 250). Thomas B. Fordham Institute. https://edexcellence.net/publications/gotta-give-em-credit

U.S. High School Graduation Rate Hits New Record High | U.S. Department of Education. (n.d.). Retrieved February 16, 2016, from http://www.ed.gov/news/press-releases/us-high-school-graduation-rate-hits-new-record-high-0

Viano, S. (2018). At-Risk High School Students Recovering Course Credits Online: What We Know and Need to Know. *American Journal of Distance Education*, *32*(1), 16–26. https://doi.org/10.1080/08923647.2018.1412554

Viano, S. (2019, April 7). *A Choice Between Second Chances: An Analysis of How Students Address Course Failure*. American Education Research Association, Toronto, CA.

Table 1
Balance check pre and post-matching, comparing CR and F2F only.

		Pre-Match				Post-Match			
Variable		Mean	SD	Std. Diff with CR	SD Ratio	Mean	SD	Std. Diff with CR	SD Ratio
Student is Black	CR	0.384	0.486			0.395	0.489		
	F2F	0.366	0.482	0.036	1.009	0.394	0.489	0.004	1.001
Student is Hispanic	CR	0.153	0.36			0.164	0.37		
	F2F	0.158	0.365	0.015	0.986	0.163	0.37	0.001	1.001
Student is Other Race	CR	0.078	0.268			0.082	0.274		
	F2F	0.086	0.281	0.031	0.954	0.083	0.276	0.004	0.994
Student is Female	CR	0.43	0.495			0.409	0.492		
	F2F	0.454	0.498	0.05	0.994	0.414	0.493	0.009	0.998
Student was Gifted (8th Grade)	CR	0.047	0.213			0.052	0.223		
	F2F	0.089	0.285	0.166	0.746	0.055	0.228	0.012	0.977
Student was SPED (8th Grade)	CR	0.174	0.379			0.18	0.384		
	F2F	0.163	0.369	0.029	1.026	0.185	0.388	0.013	0.99
Student was Previously LEP (8th Grade)	CR	0.046	0.21			0.045	0.206		
	F2F	0.042	0.201	0.019	1.043	0.041	0.198	0.019	1.044
Student was LEP (8th Grade)	CR	0.06	0.238			0.067	0.249		
	F2F	0.055	0.229	0.02	1.039	0.066	0.247	0.005	1.008
Student is Economically Disadvantaged	CR	0.649	0.477			0.67	0.47		
	F2F	0.625	0.484	0.052	0.985	0.67	0.47	0	1
Percent Days Absent (8th Grade)	CR	5.9	6.219			6.283	6.801		
	F2F	6.659	7.989	0.106	0.778	6.609	7.498	0.046	0.907
Chronically Absent (8th Grade)	CR	0.21	0.407			0.225	0.418		
	F2F	0.234	0.423	0.057	0.963	0.24	0.427	0.036	0.977
Approximate Age (8th Grade)	CR	13.163	4.285			13.161	4.318		
	F2F	13.06	4.474	0.024	0.958	13.12	4.415	0.01	0.978
Overage (8th Grade)	CR	0.26	0.439			0.264	0.441		
	F2F	0.25	0.433	0.024	1.014	0.275	0.446	0.025	0.987
Average 8th Test Score (8th Grade)	CR	-0.345	0.718			-0.372	0.731		
	F2F	-0.199	0.797	0.191	0.902	-0.371	0.782	0.001	0.935
Failed a Class (8th Grade)	CR	0.281	0.449			0.315	0.464		
	F2F	0.259	0.438	0.049	1.026	0.315	0.464	0	1

Took Remedial Course (8th Grade)	CR	0.027	0.162			0.027	0.163		
	F2F	0.03	0.171	0.019	0.947	0.028	0.166	0.007	0.981
Took Advanced Course (8th Grade)	CR	0.155	0.362			0.179	0.383		
	F2F	0.247	0.431	0.23	0.84	0.177	0.382	0.003	1.003
GPA (8th Grade)	CR	2.225	1.025			2.161	1.022		
	F2F	2.295	1.126	0.064	0.91	2.136	1.088	0.023	0.939
First Failed in 2nd Year	CR	0.285	0.451			0.254	0.436		
	F2F	0.197	0.398	0.206	1.135	0.25	0.433	0.01	1.006
First Failed in 3rd Year	CR	0.262	0.44			0.175	0.38		
	F2F	0.202	0.401	0.142	1.095	0.172	0.377	0.008	1.007
First Failed in 4th Year	CR	0.119	0.324			0.084	0.277		
	F2F	0.169	0.375	0.142	0.865	0.097	0.296	0.046	0.936
First Failed in 5th Year	CR	0.008	0.09			0.005	0.071		
	F2F	0.005	0.07	0.041	1.291	0.004	0.063	0.015	1.121
Received a Disciplinary Infraction	CR	0.178	0.382			0.126	0.332		
	F2F	0.121	0.326	0.16	1.173	0.122	0.328	0.011	1.012
Missing 8th Grade Covariates	CR	0.094	0.292			0.095	0.294		
	F2F	0.103	0.304	0.03	0.96	0.1	0.3	0.015	0.98
Missing 8th Grade Test Scores	CR	0.12	0.325			0.125	0.33		
	F2F	0.139	0.346	0.058	0.938	0.133	0.339	0.023	0.974
Missing 8th Grade Transcript	CR	0.095	0.293			0.097	0.296		
	F2F	0.101	0.302	0.022	0.971	0.101	0.302	0.015	0.98
Missing 8th Grade Grades	CR	0.103	0.304			0.105	0.306		
	F2F	0.111	0.314	0.026	0.968	0.11	0.313	0.017	0.978
Missing Overage in 8th Grade	CR	0.173	0.379			0.186	0.389		
	F2F	0.205	0.403	0.08	0.939	0.197	0.398	0.029	0.977
Missing Discipline Data	CR	0.475	0.499			0.612	0.487		
	F2F	0.532	0.499	0.114	1.001	0.607	0.489	0.012	0.997

Note. Standardized differences and standard deviation (SD) ratios are between the CR mean/SD and the other treatment statuses mean/SD; Post matching estimates are weighted using the propensity score; Differences outside of the acceptable range of values are bolded (greater than 0.25 for standardized differences and outside of the 0.5 to 2 range for SD ratios); Means will be lower in this table than overall descriptive for variables that we transformed to zero when data were missing.

Table 2

Summary statistics on the outcomes and observation sizes by sample.

	CR		F2F		CR & F2F		Initial		Neither	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Panel A: EOC Sample										
Initial EOC Score (std)	-0.79	(0.72)	-0.73	(0.66)	-0.86	(0.64)	-	(-)	-	(-)
Post-Enroll EOC Score	-0.79	(0.75)	-0.64	(0.74)	-0.83	(0.71)	-0.79	(0.72)	-	(-)
Unique Students	13,389		52,759		3,245		91,196		-	
Observations	14,652		66,203		3,366		112,065		-	
Panel B: High School Graduation Sample										
Average # Enrollments	1.71	(1.20)	1.91	(1.42)	CR 2.3	(1.71)	-	(-)	0	(-)
					F2F 2.4	(1.80)				
Prop. Graduated in 4 Years	0.74	(0.44)	0.63	(0.48)	0.59	(0.49)	-	(-)	0.59	(0.49)
Prop. Dropout	0.17	(0.38)	0.28	(0.45)	0.23	(0.42)	-	(-)	0.35	(0.48)
Unique Students	12,566		33,408		22,230				26,791	
Observations	13,226		37,937		24,036		-		29,511	

Note. The EOC sample is at the student-test-test date level. The high school graduation sample is at the student-school level. EOC scores are standardized by year and course. The "Initial" refers to the initial course enrollment for the EOC samples (i.e., the first time the student enrolled in the course and failed). The last column, "Neither," refers to the sample of students who did not enroll in CR or F2F in high school. Initial EOC scores are not defined for the Initial category since this category is made up entirely of initial scores, and the average of initial scores is then shown in the EOC score row. No initial category exists for the high school graduation sample since the sample is constructed at the student-school level and not at the course enrollment level. The Neither category is undefined for the EOC sample because those who never repeat courses are fully captured in the Initial category.

Table 3
Results from student-by-school fixed effect models with standardized end of course exam scores as the outcomes.

	(1) All Students	(2) Non-ED	(3) ED	(4) Non-Black, Non-Hispanic	(5) Black	(6) Hispanic
Panel A: All EOCs						
CR	0.048**	0.057	0.043	0.074*	0.031	0.041
	(0.018)	(0.033)	(0.023)	(0.033)	(0.024)	(0.035)
F2F	0.165***	0.172***	0.160***	0.189***	0.143***	0.167***
	(0.013)	(0.022)	(0.015)	(0.022)	(0.017)	(0.020)
Difference CR – F2F	-0.117***	-0.115***	-0.117***	-0.115***	-0.112***	-0.126***
Observations	196286	57351	138935	79690	81209	35421
R^2	0.81	0.86	0.81	0.82	0.80	0.82
Panel B: Math I EOC						
CR	0.019	0.029	0.017	0.013	0.020	0.035
	(0.042)	(0.062)	(0.053)	(0.070)	(0.055)	(0.068)
F2F	0.102***	0.112*	0.099**	0.092*	0.089*	0.152**
	(0.024)	(0.054)	(0.032)	(0.041)	(0.037)	(0.046)
Difference CR – F2F	-0.083*	-0.083	-0.082	-0.079	-0.069	-0.117*
Observations	83321	23133	60188	34815	33960	14566
R^2	0.89	0.91	0.90	0.88	0.89	0.89
Panel C: English II EOC						
CR	-0.010	-0.025	-0.020	0.027	-0.037	-0.024
	(0.057)	(0.144)	(0.076)	(0.087)	(0.069)	(0.129)
F2F	0.137***	0.097	0.135**	0.141	0.131**	0.110
	(0.041)	(0.107)	(0.052)	(0.075)	(0.048)	(0.100)
Difference CR – F2F	-0.147**	-0.122	-0.155**	-0.114	-0.168**	-0.134
Observations	57504	17174	40330	23322	23483	10705
R^2	0.94	0.96	0.94	0.94	0.93	0.94
Panel D: Biology EOC						
CR	0.026	0.079	-0.006	0.120	-0.026	-0.060
	(0.080)	(0.149)	(0.105)	(0.158)	(0.102)	(0.147)
F2F	0.217***	0.278*	0.203**	0.282*	0.190**	0.146
	(0.061)	(0.139)	(0.069)	(0.113)	(0.073)	(0.098)
Difference CR – F2F	-0.191**	-0.199	-0.209*	-0.162	-0.216*	-0.206
Observations	55461	17044	38417	21553	23766	10150
R^2	0.92	0.93	0.93	0.91	0.91	0.93

Note. Standard errors in parentheses; Standard errors clustered at the school level; Wald tests assed whether the differences between CR and F2F were statistically significant; Covariates omitted for brevity. ED refers to economically disadvantaged. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4

Results from OLS model with school-by-cohort fixed effects and propensity score weighting with graduating in four years and dropping out from high school as the outcomes.

	(1) Graduate	(2) Graduate	(3) Graduate	(4) Graduate	(5) Dropout	(6) Dropout	(7) Dropout	(8) Dropout
CR Student	0.072*** (0.0081)	0.069*** (0.0089)	0.066*** (0.0080)	0.055*** (0.010)	-0.061*** (0.0073)	-0.060*** (0.0082)	-0.056*** (0.0073)	-0.042*** (0.0095)
Black	0.11*** (0.0051)	0.089*** (0.0063)			-0.12*** (0.0049)	-0.10*** (0.0064)		
Black CR Student		0.0082 (0.012)				-0.0032 (0.011)		
Hispanic	-0.0024 (0.0069)		-0.011 (0.0084)		-0.021*** (0.0062)		-0.015 (0.0077)	
Hispanic CR Student			0.038* (0.016)				-0.031* (0.014)	
ED	-0.060*** (0.0042)			-0.076*** (0.0056)	0.045*** (0.0035)			0.058*** (0.0051)
ED CR Student				0.026* (0.011)				-0.029** (0.010)
Observations	104,164	104,164	104,164	104,164	104,164	104,164	104,164	104,164
R^2	0.239	0.240	0.239	0.239	0.238	0.239	0.238	0.238

Note. Standard errors in parentheses; Standard errors clustered at the school level; Models weighted with propensity scores created through a marginal mean weighting through stratification procedure; Covariates omitted for brevity.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figures

Figure 1

Predicted probability of graduating from high school based on student race, economic disadvantage status, and CR versus F2F enrollment.