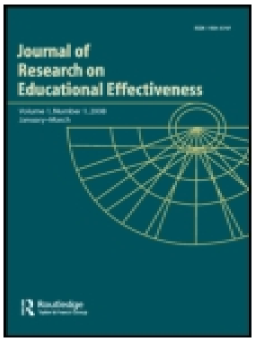

Routledge
Taylor & Francis Group

Journal of Research on Educational Effectiveness

ISSN: 1934-5747 (Print) 1934-5739 (Online) Journal homepage: https://www.tandfonline.com/loi/uree20

A Random Assignment Evaluation of Learning Communities at Kingsborough Community College—Seven Years Later

Michael J. Weiss, Alexander K. Mayer, Dan Cullinan, Alyssa Ratledge, Colleen Sommo & John Diamond

To cite this article: Michael J. Weiss, Alexander K. Mayer, Dan Cullinan, Alyssa Ratledge, Colleen Sommo & John Diamond (2015) A Random Assignment Evaluation of Learning Communities at Kingsborough Community College—Seven Years Later, Journal of Research on Educational Effectiveness, 8:2, 189-217, DOI: 10.1080/19345747.2014.946634

To link to this article: https://doi.org/10.1080/19345747.2014.946634

Journal of Research on Educational Effectiveness, 8: 189–217, 2015

ISSN: 1934-5747 print / 1934-5739 online
DOI: 10.1080/19345747.2014.946634

A Random Assignment Evaluation of Learning Communities at Kingsborough Community College—Seven Years Later

Michael J. Weiss
MDRC, New York, New York, USA

Alexander K. Mayer
MDRC, Oakland, California, USA

Dan Cullinan, Alyssa Ratledge, Colleen Sommo, and John Diamond
MDRC, New York, New York, USA

Abstract: Community colleges play a vital role in higher education, enrolling more than one in every three postsecondary students. While their market share has grown over the past 50 years, students' success rates remain low. Consequently, community college stakeholders are searching with mounting urgency for strategies that increase rates of success. We evaluate the effects of one such strategy, learning communities, from a randomized trial of over 1,500 students at a large urban college in the City University of New York (CUNY) system. We find that the program's positive effects on short-term academic progress (credit accumulation) are maintained seven years after random assignment. We find some limited evidence that the program positively affected graduation rates, particularly for those students without remedial English needs, over this time period. While there is not clear evidence that the program improved economic outcomes, this article concludes by offering sobering reflections on trying to detect the effects of higher education interventions on future earnings.

Keywords: Learning communities, community college, randomized experiment, developmental English, earnings

INTRODUCTION

Over the last five decades, community colleges have played an increasingly prominent role in American postsecondary education and workforce development. In 1963, community colleges enrolled 740,000 students, representing 15% of all fall enrollment in postsecondary institutions. By 2010, enrollment had increased by more than 875%, to 7.2 million students, representing 34% of all fall postsecondary enrollment (National Center for Education Statistics, 2011). Community colleges are generally open-access institutions with minimal, if any, academic admission requirements. Their cost of attendance is significantly less than the cost at 4-year colleges and universities—community college tuition and fees are around

Address correspondence to Michael Weiss, MDRC, YAPSE, 16 East 34th Street, 19th Floor, New York, NY 10016-4326, USA. E-mail: michael.weiss@mdrc.org

one third the cost of public 4-year colleges and universities (College Board Advocacy and Policy Center, 2012). Unlike most 4-year institutions, community colleges frequently offer certificate and degree programs aimed at training workers in disciplines where they can immediately transition from a short college program to a full-time job or career. All of these characteristics contribute to community colleges' appeal to low-income, at-risk populations.

Empirical evidence shows that increased education is positively associated with higher earnings across a wide spectrum of fields and student demographics (Bailey, Kienzl, & Marcotte, 2004; Barrow & Rouse, 2005; Card, 2001; Carneiro, Heckman, & Vytlacil, 2011; Dadgar & Weiss, 2012; Dynarski, 2008; Jacobson & Mokher, 2009; Jepsen, Troske, & Coomes, 2009; Kane & Rouse, 1995). College degree holders (including community college degree holders) earn higher salaries and experience unemployment less frequently than those who do not have a college degree (Dadgar & Weiss, 2012), and those who attended college, regardless of degree receipt, report higher rates of job satisfaction, promotion opportunities, increased work responsibilities, and improved work performance (Hoachlander, Sikora, & Horn, 2003). Studies seeking to quantify the economic returns to postsecondary education have found mixed results regarding whether it is the degree itself that makes a difference or if credits alone can improve earnings. Some studies have found that students who enroll in college but do not receive a degree experience a significant bump in earnings, with one study finding as much as a 5% increase in salary for each year of credits earned (30 credits) and no statistically significant difference between outcomes for students who receive a credential and students who complete the same number of credits without receiving a degree (Kane & Rouse, 1995). Other studies find that degree receipt produces significantly higher wage outcomes compared to credits alone: Although each additional semester or year of credits earned leads to an increase in wages, degree receipt is associated with a larger difference (Dadgar & Weiss, 2012; Jepsen et al., 2009).

Although enrollment in community colleges has increased over the last half century and increased education appears to be a path toward increased economic outcomes, persistence and completion rates leave much room for improvement. Among students beginning at public 2-year colleges, fewer than half earn a credential or transfer to a 4-year institution within 6 years after their initial enrollment (Radford, Berkner, Wheeless, & Shepherd, 2010). Many community college students stop or drop out long before graduation day, losing out on the full complement of higher earnings and employment outcomes associated with higher education.

One factor contributing to these low success rates is the fact that many students arrive on campus only to find that they are required to take non-credit-bearing courses, called "developmental" or "remedial" education courses, prior to enrolling in college-level courses (Duke & Strawn, 2008). One estimate of the prevalence of students required to complete developmental coursework suggests that among students whose first institution attended was a community college, around 60% took a remedial course at a postsecondary institution (Adelman, 2004; Bailey, Jeong, & Cho, 2010). Unsurprisingly, degree or certificate attainment rates among students who need developmental education are even lower than those of the general population (Attewell, Lavin, Domina, & Levey, 2006). Students in need of developmental education are not just failing to earn a credential—the majority never complete the developmental course sequence to which they are referred (Bailey et al., 2010). One popular strategy aimed at addressing the issues of students referred to developmental courses (as well as students in college-level courses) is called "learning communities" (LCs).

LEARNING COMMUNITIES

Although definitions of LCs vary, there is much agreement about their core elements. LCs involve the coenrollment of a cohort of students into two or more courses. Typically, the curricula of these courses are linked or integrated, sometimes around a theme. For example, a learning community centered around and titled "Poverty and Inequality" might link a developmental English course and a college-level sociology course. Students might learn various elements of essay writing and argumentative rhetoric in their English course and use those techniques to write a paper in their sociology course exploring the relationship between income status and race. LCs tend to involve faculty collaboration, which enables teachers to communicate about their shared students and to integrate the curriculum across courses. Some definitions of LCs also include a pedagogical component, usually focusing on active engagement, active pedagogy, and/or collaborative learning. Finally, some LCs include additional student support services, like enhanced counseling or tutoring (Minkler, 2002; Richburg-Hayes, Visher, & Bloom, 2008; Smith, 2001; Smith, MacGregor, Matthews, & Gabelnick, 2004).

In practice, LCs vary from extremely basic (e.g., simply coenrolling students in two or more classes) to multifaceted (e.g., coenrollment, tightly integrated curriculum with joint assignments and assessments, extensive faculty collaboration, and group tutoring). Some form of LCs can be found at an estimated 400 to 500 colleges and universities throughout the United States (Smith, 2001; Smith et al., 2004). In community colleges in particular, a national study reported that around half of 288 surveyed colleges were operating at least one LC (Center for Community College Student Engagement, 2012).

Why Learning Communities Are Expected to Work

LCs at community colleges have the potential to improve student outcomes through several mechanisms. Tinto's classic theoretical framework on dropout from higher education posits that

> it is the person's normative and structural integration into the academic and social systems that lead to new levels of commitment. Other things being equal, the higher the degree of integration of the individual into the college systems, the greater will be his commitment to the specific institution and to the goal of college completion. (Tinto, 1975, p. 96)

At community colleges, this is especially acute: Community colleges are attended largely by commuter students, many of whom contend with the competing demands of work, family obligations, and school (Brock, LeBlanc, & MacGregor, 2005; Minkler, 2002). Consequently, for many students, the only time they are on campus is when they attend class, limiting opportunities for integration to those that can take place in the classroom.

By coenrolling students into two or more courses, LCs are intended to foster stronger connections among students (Smith et al., 2004). Through sustained academic relationships among students and faculty, students are expected to feel more integrated into a community of peers and college life, leading to a greater level of commitment. In addition, integrating course materials may help students understand connections between disciplines and between what they are learning in school and their personal lives, and in doing so engage students more deeply in learning. Curricular integration may be particularly effective when a developmental-level course and college-level course are paired in a learning community, as this allows students to use the basic skills they are learning in their developmental classes in their college-level coursework. Broadly speaking, LCs are theorized to lead to improved academic outcomes by fostering stronger connections among students and between students and faculty, integrating students into campus life, and providing a more engaging academic environment. This is hypothesized to increase students' likelihood of persisting in school and improve their academic attainment.

Evidence of the Effectiveness of Learning Communities

Over the last two decades, a number of researchers have argued that LCs positively affect students. Among the first of the studies published in a peer-reviewed journal was an article by Tinto and Russo (1994) that found that students who participated in a learning community reported increased involvement in academic and social activities, had more positive views of college, and were more likely to persist to the following semester.

Subsequent to Tinto and Russo's findings, multiple researchers, studying a variety of LCs, supported the finding that LCs positively affect student engagement, retention, and academic progress (Engstrom & Tinto, 2008; Goldberg & Finkelstein, 2002; Gordon, Young, & Kalianov, 2001; Stassen, 2003; Taylor, 2003; Tinto, 1997, 1998; Wilmer, 2009; Zhao & Kuh, 2004). These results appeared quite promising; however, the studies all had major limitations with respect to their designs' ability to make causal claims about the effectiveness of LCs.

A key limitation in these studies is that they use nonexperimental designs. Comparisons were made between students enrolled in LCs and those who were not, sometimes attempting to statistically control for a narrow set of available background characteristics. With respect to establishing the causal effect of LCs, the past studies' designs were limited, and none reported conducting sensitivity checks to establish the robustness of their findings to selection bias or other assumptions. Many of the studies' authors acknowledged these limitations. For example, Stassen (2003) noted, "It is possible that students who are most motivated to succeed take advantage of the [learning communities'] opportunities and, as a result, retention and academic performance rates for [learning communities] are better because of individual student selection—not the program components themselves" (p. 587).

To fill this gap in the research, in 2003, the first large-scale random assignment study of LCs was launched, by MDRC, to evaluate the effectiveness of Kingsborough Community College's (KCC) Opening Doors Learning Communities program—the subject of this article.[1] In the study, more than 1,500 students were randomly assigned to either a program group, which had the opportunity to participate in the LCs program, or a control group, which could enroll in the college's usual courses. KCC's LCs model placed freshmen into groups of up to 25 students who took three classes together during their first semester: an English course (either college-level or developmental), an academic course required for the student's major, and a one-credit freshman orientation course. KCC also provided LC students with enhanced counseling and tutoring services, as well as textbook vouchers, leading to a robust learning community model.

[1]The first small-scale random assignment study of LCs appears to be by Goldberg and Finkelstein (2002), but it included only 25 students.

After finding promising early results on full-time enrollment, credits attempted, and credits earned in the evaluation of KCC's LCs (Richburg-Hayes et al., 2008), MDRC, through the National Center for Postsecondary Research, launched a demonstration to evaluate LCs' effects on students (primarily in need of developmental education) at six community colleges. In randomized controlled trials of these programs, MDRC found that the studied LCs (which varied in focus, curricular integration, and student supports), on average, had a small positive effect on students' progress in the subject targeted by the LC and total credit accumulation (Visher & Teres, 2011; Visher, Weiss, Weissman, Rudd, & Wathington, 2012).

The KCC LC evaluation represents one of the first large-scale random assignment evaluations in higher education, and the findings presented in this article represent one of the first large-scale higher education experiments to track students for a long period—7 years at the time of writing—and analyze program effects on academic outcomes such as graduation. The article also includes estimates of the effects of the intervention on employment and earnings outcomes. It concludes by considering the potential for detecting impacts due to higher education interventions on economic outcomes more broadly.

THE KINGSBOROUGH LEARNING COMMUNITIES EVALUATION

This evaluation used a random assignment research design to estimate the effects of KCC's LCs program compared to the college's regular classes and services. This section describes the main research questions addressed in this article, how students became part of the research sample, and how the program was implemented at KCC—a large college in southern Brooklyn, New York. It also presents some characteristics of the sample members and discusses the data sources used in this article. More detailed information can be found in Scrivener et al. (2008).[2]

Research Questions

This evaluation sought to answer the following primary research question:

RQ1: What was the overall average effect of the opportunity to participate in KCC's LCs program on students' *academic outcomes* (including academic progress and completion)?

In addition, we explored the following secondary research questions:

RQ2: What was the overall average effect of the opportunity to participate in KCC's LCs program on students' *labor market outcomes* (including employment and earnings)?

RQ3: Did the effects of the opportunity to participate in KCC's LCs program vary by students' pre-random assignment characteristics?

[2]The majority of information presented in this section is adapted from Scrivener et al. (2008).

Eligibility Criteria

For students to be eligible to participate in the evaluation, they had to meet several eligibility criteria. Students were eligible for the study if they (a) were first-time incoming freshmen who planned to attend college full time during the day, (b) did not test into English as a Second Language (ESL)[3] (i.e., they tested into either developmental English or college-level English), and (c) were between the ages of 17 and 34.[4]

Recruitment and Random Assignment

Potential study participants were identified during the weeks prior to the start of each semester. Scores on the college's reading and writing placement tests determine most students' English placement level. Applicants whose scores placed them in a developmental English course or in freshman English were invited to come to campus to register early for classes.

Students who came in to register received a brief, general description of the LCs program and were told that a random process would be used to determine which study participants would be placed in the program. Students who agreed to participate in the study signed an informed consent form, provided baseline demographic information, and were randomly assigned, by MDRC, either to the program group or to the control group and were given assistance registering for classes.

Students were brought into the research sample in four different groups, or cohorts, before four semesters: fall 2003, spring 2004, fall 2004, and spring 2005. Throughout the study, a total of 1,534 students were randomly assigned at KCC (769 program students, 765 controls).

Data Sources

The analyses presented in this article rely on several data sources, described next. All data sources were provided for both program and control group members.

Baseline Data. As mentioned, just before students were randomly assigned to the study groups, they completed a Baseline Information Form. The questionnaire collected demographic and other background information. Baseline data are used in this article to describe the sample, for subgroup analyses, and as covariates in sensitivity analyses.

CUNY Transcript Data. MDRC was provided with transcript data from the City University of New York (CUNY) Institutional Research Database. These data include information on courses taken, such as course name, credits, and grades, as well as degrees earned from

[3]Students whose scores placed them in ESL were not included in the study, as they were eligible for the college's ESL LCs program.

[4]During the first semester of program operations, KCC's LCs program was open only to students between ages 18 and 34 who reported household income below 250% of the federal poverty level. In subsequent semesters, the income criterion was removed, having been deemed unnecessary because such a large proportion of KCC students are from low- or moderate-income families, and 17-year-olds were admitted to the program with parental consent. Age limits were implemented by funder request.

all CUNY institutions, including KCC. The analyses for this report include data through the fall 2011 semester, which represents 7 years of follow-up for all cohorts of the full study sample. The transcript data are used to estimate the program's effect on academic outcomes.

National Student Clearinghouse Data. The National Student Clearinghouse, a nonprofit organization, collects and distributes enrollment, degree, and certificate data from more than 3,500 colleges that enroll more than 98% of the nation's college students.[5] The Clearinghouse data are used to provide enrollment and degree attainment information for students in the study who attended a postsecondary institution outside the CUNY system. The Clearinghouse data are available for all semesters of the 7-year follow-up period for all cohorts and are used to estimate the program's effects on enrollment and degree attainment.

New York State Department of Labor Unemployment Insurance Data. The New York State Department of Labor provided MDRC with individual-level employment records and aggregate earnings data by research group and subgroups for our sample from 2001 through 2012. These data are used to explore the program's effects on economic outcomes, and capture employment at New York employers during the period.[6]

CUNY Skills Assessment Test Score Data (Placement Test). Students are required to take the CUNY reading, writing, and math skills assessment tests before they begin classes at KCC. MDRC collected test score data for all sample members who took the tests at KCC or any other institution in the CUNY system. In this article, baseline placement test score data are used to make statistical adjustments in the sensitivity analysis. Placement tests were also used by the college to determine students' placement into either college-level English or one of several developmental-level English courses. Placement level is used to define subgroups of sample members for moderator analysis.

Characteristics of the Sample

Table 1 displays a selection of characteristics of the study sample members obtained from the baseline questionnaire, CUNY placement test, and New York State Department of Labor data. The table shows the characteristics for the full sample by research group, that is, the program group and control group. Just over half of the full sample is women. The full sample is racially and ethnically diverse. Sample members were quite young: Only 21% were 21 or older. Slightly less than 10% of the sample has children, and around 6% of the sample was unmarried and had one or more children. More than two thirds of these students had a high school diploma, making this sample more "traditional" than the overall population of American community college students (Adelman, 2005). Around 30% of students delayed their enrollment in college (defined as enrolling in college in a different calendar year than the year in which they graduated high school.) Almost half reported speaking a language other than English at home.

[5]For any semester that a student attended only an institution that does not submit to the Clearinghouse, the student is treated as not enrolled in our analyses.

[6]These records cover about 90% of employment, but they do not capture certain types of jobs, including self-employment, federal government employment, military personnel, informal jobs, and out-of-state jobs.

Table 1. Selected characteristics of sample members at baseline

Characteristic	Program Group	Control Group	Difference	*p* Value
Gender (%)				
Female	57.7	51.4	6.4	.012
Race/ethnicity (%)[a]				.307
Hispanic/Latino	21.2	19.6	1.5	
Black, non-Hispanic	38.1	37.2	0.9	
White, non-Hispanic	24.6	29.3	−4.7	
Asian or Pacific Islander	9.4	7.8	1.5	
Other	6.8	6.0	0.7	
Age (%)				.127
17–18-years-old	44.9	44.1	0.8	
19–20-years-old	35.8	32.6	3.2	
21–34-years-old	19.3	23.3	−4.0	
Has one or more children (%)	8.3	9.1	−0.8	.580
Unmarried and has one or more children (%)	6.5	6.4	0.1	.916
Diplomas/degrees earned (%)[b]				
High school diploma	72.9	68.9	4.1	.079
General Educational Development certificate	25.8	31.4	−5.7	.014
Occupational/technical certificate	2.1	2.0	0.1	.868
Did not earn high school diploma, or did not pass 12th grade (%)	29.1	33.8	−4.7	.048
First person in family to attend college (%)	34.9	31.9	3.1	.208
Delayed postsecondary enrollment (%)[c]	31.2	29.5	1.7	.476
Language other than English spoken regularly in home (%)	48.6	45.2	3.4	.179
Financially independent (%)[d]	13.4	15.5	−2.1	.246
English skills assessment (%)				.819
Passed both English tests at baseline	29.3	28.8	0.5	
Failed one English test at baseline	45.1	46.7	−1.5	
Failed both English tests at baseline	25.6	24.6	1.0	
Placement test scores				
COMPASS M1 Math	37.4	39.2	−1.8	.163
COMPASS M2 Math	28.5	28.0	0.4	.665
ASSET Math	22.3	22.6	−0.4	.569
COMPASS Reading	71.2	72.0	−0.8	.355
ACT Writing	5.9	5.9	0.0	.925
Employment in year prior to random assignment[e]				
Ever employed (%)	52.4	55.6	−3.2	0.214
No. of quarters employed	1.35	1.49	−0.14	0.080

(*Continued on next page*)

Table 1. Selected characteristics of sample members at baseline (*Continued*)

Characteristic	Program Group	Control Group	Difference	*p* Value
Quarterly earnings prior to RA ($)				
Two quarters prior to RA	740	881	−141	0.083
One quarter prior to RA	850	1031	−181	0.048
Sample size (total = 1,534)	769	765		

Source. MDRC calculations using Baseline Information Form data, New York State Department of Labor data, and City University of New York skills assessment test data. *Note.* Distributions may not add to 100% because of rounding. Missing values are not included in individual variable distributions. Either a chi-square test or a two-tailed test was applied to differences between research groups.

[a]Respondents who indicated that they are Hispanic and who also chose a race are included only in the Hispanic/Latino category. [b]Distributions do not add to 100% because categories are not mutually exclusive. [c]Following Horn et al. (2004), students were considered to have delayed postsecondary enrollment if they graduated from high school and enrolled at college in different calendar years. An exception was made for 27 students who graduated from high school in August through November and enrolled in college in January through March of the following year. These students were not treated as having delayed postsecondary enrollment. [d]Students were considered financially independent if any of the following applied: was 24 or older as of random assignment, was married, had children (excluding spouse). [e]These values include data from the four quarters prior to the quarter in which students were randomly assigned.

Prior to enrollment at KCC, students took two tests (one English, one reading) that were used to determine their placement into either college-level English or one of several developmental English courses. Students who passed both tests were placed in college-level English, students who passed only one test were placed into the highest level of developmental English, and students who failed both tests were placed into a lower level of developmental English. A plurality of students passed one test only, placing them into the highest level of developmental English (46%). The rest of the sample is roughly evenly split between students passing neither test and students passing both tests. Average placement test scores are also reported for the math and English placement tests. Finally, during the year prior to the evaluation, on average, study participants were employed in slightly more than half of the quarters, and around 15% of students were financially independent.[7]

Some statistically significant differences between research groups were found at baseline. For example, there are more female participants and slightly lower rates of pre-random-assignment employment in the program group. Although some of these baseline differences are on factors that may correlate with the outcomes, there is no reason to believe that random assignment was compromised or that our effect estimator is biased (MDRC controlled the random assignment process). However, sensitivity analyses are conducted, and described in the appendix, to explore the effect on impacts and standard errors when

[7]The research sample, which consists of students who met program eligibility criteria and agreed to participate, may not be representative of the broader KCC student body or the eligible population at KCC. For example, according to IPEDS 2003 fall cohort data, male participants are slightly overrepresented in the research sample (45.4% of the research sample vs. 40.7% of the student body). There is a smaller range of ages in the research sample than the student body because of the eligibility criteria. The research sample also has a higher proportion of Black and Hispanic students than the student body (58.1% of the research sample vs. 48.7% of the student body). In addition, the research sample may look different from the student body on unobservable characteristics.

regression adjustment is made based on a number of covariates that prior research has found to be correlated with academic success.[8]

Program Implementation

As previously discussed, LC staff recruited and randomly assigned 1,534 students to the program or control group over four semesters (fall 2003 through spring 2005). Over the four semesters, KCC ran 40 LCs for the study: 31 with developmental English and nine with college-level English.

The key structural feature of the program (coenrollment in three courses) operated as intended, with only minor glitches, throughout the study period. A large majority of students in the program group enrolled in linked classes (85.3%), tutors and case managers with relatively small caseloads were assigned to each learning community, and textbook vouchers were distributed as planned.[9] Although all of the LCs had the same basic structure, there was variation in content, class size,[10] and the degree to which faculty worked together and integrated their courses. A faculty survey administered to one cohort of learning communities instructors also indicated that in preparation for teaching in an LC, faculty developed a new syllabus or adapted their regular syllabus and that they gave at least some joint assignments. Thus, the study provides a strong test of the structural features of the learning community, but it may not fully detect the effects of tightly integrating course curricula.

In addition to the question of implementation, it is important to know how the program differed from the college's usual services. The contrast is clearest with respect to course assignments and scheduling. LC students took three linked courses that were scheduled in a block, and all of them took an English course and the freshman orientation class. Control group students took whatever courses were available to them, at whatever times those courses met, and were not required to take English or the freshman orientation.

A typical full-time course load at KCC involves 12 credits (12 hr of class per week). Because the lower level developmental English courses meet for 8 hr each week, the content courses are typically three credits, and the freshman orientation class is one credit, students at the lower English levels usually took no additional unlinked courses. In contrast, students in higher level English courses, which meet for fewer hours per week, usually took at least one non-LC course.

RESULTS

Program Effects

The program's effects on two outcome types are examined—academic outcomes and employment outcomes. For details on the estimation strategy, see the appendix. With respect to academic outcomes, confirmatory (or primary) outcomes include students' (a) *progress*

[8]These were "risk characteristics associated with students' likelihood of leaving postsecondary education without attaining a credential" described by Horn, Berger, and Carroll (2004). Further information on our proxies of these variables can be found in the sensitivity analysis section.

[9]For a detailed discussion of the implementation research conducted for this program, see Scrivener et al. (2008).

[10]Because there were challenges in managing registration and predicting how many students would test into each level of English, class size varied from six to 25 students, with an average of around 17.

Table 2. Key outcomes, 7 years after random assignment

Outcome	Program Group	Control Group	Estimated Effect	SE	*p* Value
Academic outcomes					
Total credits earned[a]	58.8	54.7	4.0	2.4	.092
Earned a degree (%)	39.5	36.2	3.3	2.8	.236
Economic outcomes					
Quarters employed in Year 7 (%)	56.4	54.2	2.3	2.5	.355
Earnings in Year 7[b] ($)	15,820	14,652	1,168	1,033	.258
Sample size (total = 1,534)	769	765			

Source. MDRC calculations from City University of New York (CUNY) Institutional Research Database, National Student Clearinghouse data, and New York State Department of Labor. *Note.* Rounding may cause slight discrepancies in sums and differences. Estimates are adjusted by research cohort. For academic outcomes and quarters employed in Year 7, cluster-robust standard errors are used when calculating standard errors. Students are clustered by learning community link. Degree and enrollment measures include outcomes from any college. Credits refer to credits earned at any CUNY college.

[a]Total credits include both college-level and developmental credits. Values of zero credits earned have been imputed for nine students for whom CUNY data were unavailable. [b]Sixty-three additional students are included in other analyses but are not included here because earnings data were not available. Social security numbers for these students were not available at random assignment.

toward a degree as measured by total credits earned at any CUNY college and (b) *completion of a degree* as determined by whether students earned any type of degree at any institution covered by the National Student Clearinghouse. With respect to employment and earnings, outcomes are (a) employment and (b) earnings in New York State.[11]

Table 2 summarizes the main findings 7 years after students entered the study (detailed results are presented in the following sections). With respect to academic progress, KCC's single-semester LC program had a positive effect on credit accumulation, increasing credits earned by an estimated 4.0 credits (7.5%) over 7 years ($p = .09$).

Although not statistically significant, the learning community program is estimated to have increased graduation rates by 3.3 percentage points (from 36.2% to 39.5%) after 7 years. The positive estimate, along with positive estimates in each of the preceding 4 years (shown in Table 3), is encouraging for a short-duration intervention, but it is tempered because the lowest *p* value associated with the estimated effects on graduation is .104, in Year 6.[12]

With respect to employment and earnings, during the 7th year after students were randomly assigned there was no discernible effect on economic outcomes. It is important

[11]In addition to the four outcomes described here, many other outcomes of interest were explored, some of which are described in the text next. To reduce the multiple hypothesis testing problem (Schochet, 2008), two outcomes were prespecified in each outcome domain (academics and employment).

[12]Note that in a previous MDRC report, Sommo, Mayer, Rudd, and Cullinan (2012) presented 6-year graduation impact estimates that had a *p* value just below .10. Results presented here show a 6-year graduation impact estimate that has a *p* value of .104, slightly above 10. From a research perspective, the difference is inconsequential. The *p* value changed because of new information and improved understanding regarding the 6-year degree status of two sample members, including updated data from the Clearinghouse on one of the students.

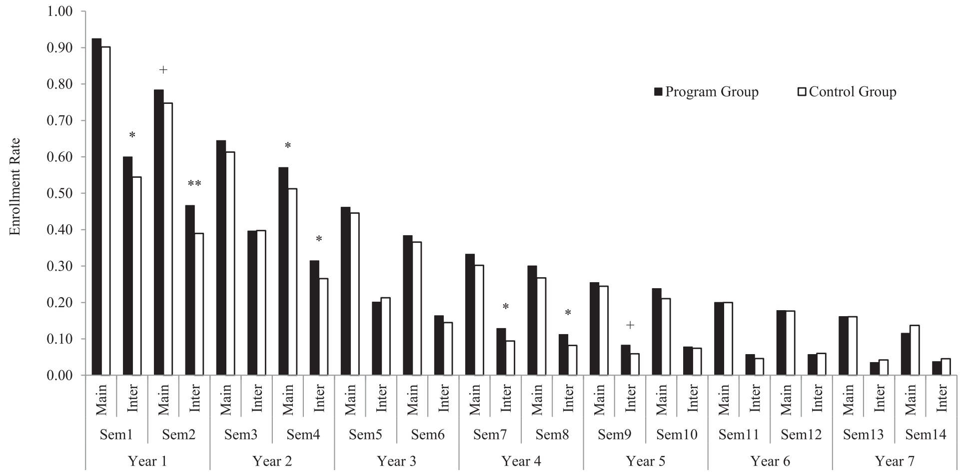

Figure 1. City University of New York (CUNY) enrollment in main sessions and intersessions. *Source.* MDRC calculations from CUNY Institutional Research Database. *Note.* Estimates are adjusted by research cohort. Cluster-robust standard errors are used when calculating *p* values; students are clustered by learning community link. Statistical significance levels are indicated as $^{+}10\%$. $^{*}5\%$. $^{**}1\%$.

to note that this study was not designed to be powered to detect effects on employment and earnings; still, as one of the largest experiments in higher education history it provides a unique opportunity to reflect on the possibility of connecting higher education interventions to effects on employment and earnings. In the concluding section of this article insights into the possibility of making such connections are offered, grounded in what has been learned from this real-world experiment. We now turn to details regarding the impact findings.

Academic Outcomes

Persistence. Although the confirmatory academic outcomes are credit accumulation and degree receipt, we first examine the programs' estimated effect on persistence (as measured by enrollment rates), a hypothesized mediator of the program's effect on academic success. One aspect of LCs' theory of change is that the coenrollment of students will foster stronger connections among them and between students and faculty members, thus increasing students' likelihood of persisting, and in doing so, improve academic outcomes like credit accumulation. Figure 1 plots enrollment rates *at any CUNY college*, over the course of 7 years.[13] Each year comprises two semesters (fall and spring). Each semester comprises two sessions that we refer to as the main session (12 weeks) and the intersession (6 weeks). Notably, the bulk of the LC experience took place during the main session of the first semester (the program semester).

The first striking feature of Figure 1 is that during six of 28 sessions the program's estimated effects on enrollment rates had a *p* value below .05, as indicated by star(s) above the program and control group bars. For example, during the main session of Semester 4,

[13]National Student Clearinghouse data enables us to examine colleges beyond CUNY. However, we focus on CUNY colleges in this section because it allows us to break apart enrollment during the main sessions (fall and spring) and the intersessions (winter and summer).

57.1% of program group members enrolled compared to 51.2% of control group members. The 5.9 percentage point difference represents the estimated effect on enrollment.

Equally noteworthy is the program's effects on enrollment rates during *intersessions*, the 6-week winter and summer sessions that are included in students' tuition at KCC but are attended by a lower proportion of students than the 12-week fall and spring *main* sessions. During five of the first eight semesters, the program's estimated effect on intersession enrollment had a p value below .05, with estimated effects averaging just above 3 percentage points. Of importance, some components of the learning community program, such as enhanced counseling and the use of textbook vouchers, were available for the first full semester, including the intersession. It is possible that these services, coupled with advice and encouragement to attend intersessions from the program case managers, enticed more program group students to give intersessions a try and they continued to use them in subsequent postprogram semesters. It is also possible that some students who enrolled in learning communities felt more connected to campus and thus enrolled at a higher rate compared to their control group counterparts. This study cannot determine the specific mechanisms that led to positive impacts.

When considering Figure 1 it is important to note that in later semesters examining program effects on enrollment rates alone can be misleading. Some students graduated and were no longer enrolled. In addition, to highlight the intersession findings, Figure 1 considers enrollment *at CUNY colleges only* using CUNY's database; however, in later semesters more students enrolled at non-CUNY colleges.[14]

Because enrollment rates at the end of the 7-year follow-up provide an indication of what is to come beyond 7 years, it is noteworthy that during the final study semester, 22% of program group students were enrolled *at any college* compared to 24% of control group students—or under one fourth of the study sample (not shown in the figure). This finding is returned to later when considering LC's long term effects on degree completion.

Credit Accumulation. Table 3 provides information on cumulative total credits[15] attempted and earned at the end of Years 1 through 7 *at CUNY colleges*. The LCs program operated during the first semester of Year 1. During that year, students randomly assigned to the program attempted 1.4 more credits and earned 2.1 more credits than their control group counterparts. A 95% confidence interval for the program's estimated effect on credit accumulation through one year ranges from 0.8 to 3.3 ($2.1 \pm 1.96 \times 0.6$). The estimated effect on credits earned in Year 1 represents an 11% increase over the control group average of 18.2 credits. The effect on credits earned was evenly split between developmental credits and college-level credits (not shown in table).

[14]Using National Student Clearinghouse data enrollment *at any college* covered by the Clearinghouse database was also examined by semester (the Clearinghouse data does not allow a clean breakdown by sessions). During the first 2 years after random assignment, enrollment rates outside of CUNY were low (less than 5%), so enrollment in CUNY colleges (i.e., what is shown in Figure 1) is similar to enrollment at any college. During the 3rd through 7th years, enrollment rates at any college were between 5 and 11 percentage points higher than enrollment at CUNY colleges alone. This is important to note when considering the magnitude of enrollment rates over time, because Figure 1 underrepresents enrollment at any college. Of importance, program-control group differences in enrollment rate generally were not affected by enrollment outside of CUNY (where they could be estimated), so the estimated program effects shown in Figure 1 are likely about the same as if they could be examined at any college.

[15]Total credits include both college-level credits (which are generally degree applicable) and developmental credits (which do not count toward a degree).

Table 3. Academic outcomes, Years 1–7

Outcome	Program Group	Control Group	Estimated Effect	SE	*p* Value
Cumulative credits attempted (at CUNY)					
Year 1	28.3	26.9	1.4	0.6	.024
Year 2	47.4	44.6	2.7	1.2	.024
Year 3	58.8	55.8	3.0	1.7	.071
Year 4	67.1	63.1	4.1	2.0	.045
Year 5	73.4	68.7	4.7	2.3	.041
Year 6	77.9	73.0	4.8	2.5	.054
Year 7	80.9	76.4	4.6	2.7	.087
Cumulative credits earned (at CUNY)					
Year 1	20.3	18.2	2.1	0.6	.002
Year 2	33.9	31.1	2.8	1.2	.020
Year 3	42.0	39.4	2.6	1.6	.098
Year 4	48.1	44.8	3.4	1.9	.076
Year 5	52.8	49.0	3.7	2.1	.074
Year 6	56.3	52.3	4.0	2.3	.078
Year 7	58.8	54.7	4.0	2.4	.092
Earned a degree (at any college) (%)					
Year 1	0.0	0.0	0.0		
Year 2	6.1	5.4	0.8	1.5	.615
Year 3	20.2	17.0	3.2	2.5	.202
Year 4	26.5	23.8	2.7	2.6	.302
Year 5	31.9	28.4	3.5	2.8	.210
Year 6	35.9	31.5	4.4	2.7	.104
Year 7	39.5	36.2	3.3	2.8	.236
Highest degree earned by Year 7 (%)[a]					
Bachelor's degree or higher	16.5	14.8	1.7	2.1	.404
Associate's degree	22.2	20.5	1.7	2.2	.434
Sample size (total = 1,534)	769	765			

Source. MDRC calculations from City University of New York (CUNY) Institutional Research Database and National Student Clearinghouse data. *Note.* Rounding may cause slight discrepancies in sums and differences. Estimates are adjusted by research cohort. Cluster-robust standard errors are used when calculating standard errors. Students are clustered by learning community link. Cumulative credits include both college-level and developmental credits. Values of zero credits attempted/earned have been imputed for nine students for whom CUNY data were unavailable.

[a]Percentage who earned bachelor's degree or higher and percentage who earned associate's degree do not add up to total because the degree type of some degree earners was unknown.

After the 1st year, the magnitude of the estimated effect on total credits earned increased, leveling at 4.0 credits 7 years after random assignment.[16] At this time, program group members earned an average of 58.8 credits and control group members earned an average of 54.7 credits, the difference represents a 7% increase in credits earned. To put the magnitude of this effect in perspective, most classes are worth three or four credits and most associate degrees require a minimum of 60 college-level credits to graduate. Therefore, on average, the estimated effect represents completing an addition 1 to 1.33 more courses or about one fifteenth of the credits required to earn an associate's degree.

[16]$p = .092$. The p value increases due to increased variance in the outcome over time.

After 7 years, college-level credits account for a little more than three fourths of the estimated effect on total credits earned (not shown in the table).[17] Tying the intersession effects on enrollment rates to the credit accumulation effects—one fourth of the estimated effect on total credits earned (i.e., one credit) occurred during the winter and summer intersessions.

Degree Completion. The third panel in Table 3 provides information on degree attainment *at any college* at the end of Years 1 through 7. Given the program's positive effect on credit accumulation through 7 years, it is plausible that the LCs program could affect degree attainment. Before delving into the program's effects, we first consider overall patterns of degree attainment.

Following national trends, degree attainment rates are low for both program and control group members. Two years after entering the study, around 6% of all sample members earned a degree at any college. After 7 years, degree attainment rates rose to nearly 40%. Notably, in the final term of study, only 10% of the full sample was enrolled *and* had not yet earned any degree.[18] Thus, for the degree completion rates to change dramatically, many students not enrolled in Year 7 would need to reenroll.

Although not statistically significant, the program's estimated effect on degree attainment is positive in Years 2 through 7 after random assignment. This positive trend is encouraging, especially in light of the estimated impacts on credit accumulation. Even in Year 6, however, when the estimated effect is largest (4.4 percentage points) a 95% confidence interval includes zero. In Years 3 through 7 the estimated effect hovered around the 3.5 percentage point range, an effect that, if real, some might consider a practically significant effect for a single-semester intervention. However, despite being a large randomized experiment, this study was not powered to detect effects of this magnitude.

Economic Outcomes

The previous section established the KCC LCs program's positive effect on credit accumulation. Some evidence suggests that the program may have affected graduation rates as well. Here, we explore whether there were also effects on labor market outcomes, because past research has connected both credit accumulation and degree attainment to economic returns. The focus is on two main labor market outcomes: employment and earnings.

Employment. Table 4 provides yearly data on the percentage of quarters employed in New York State during the 2 years prior to random assignment and the 7 years after random assignment. Prior to random assignment, there were small differences in employment rates—the control group was slightly more likely to be employed during that time.[19] During the 7 years following random assignment, employment rates remained similar

[17]The estimated effect on developmental credits all occurred during the 1st year of study and was maintained (but did not grow) during subsequent years.

[18]As noted earlier, more than 20% of the full sample were still enrolled, but nearly half of those enrolled had already completed a first degree.

[19]In this data a positive correlation was found between prerandom assignment employment and future employment. Thus, this difference would appear to favor the control group. Sensitivity analyses were conducted controlling for prerandom assignment employment status—the results were substantively the same.

Table 4. Economic outcomes, years 1-7

Outcome	Program Group	Control Group	Estimated Effect	SE	*p* Value
Percentage of quarters employed (%)					
Pre-RA Year 2	21.4	24.7	−3.3	1.8	.062
Pre-RA Year 1	33.7	37.4	−3.7	1.9	.058
Post-RA Year 1	43.5	44.5	−1.0	2.1	.621
Post-RA Year 2	51.6	51.2	0.4	2.3	.852
Post-RA Year 3	54.7	55.8	−1.1	2.3	.640
Post-RA Year 4	56.3	56.2	0.1	2.3	.975
Post-RA Year 5	56.3	53.8	2.5	2.4	.301
Post-RA Year 6	56.2	54.2	2.1	2.5	.407
Post-RA Year 7	56.4	54.2	2.3	2.5	.355
Average earnings following RA ($)					
Year 1	4,060	4,611	−551	346	.112
Year 2	6,006	6,166	−160	436	.713
Year 3	8,417	8,126	291	556	.601
Year 4	10,295	10,204	91	666	.892
Year 5	12,103	11,970	134	789	.866
Year 6	13,656	12,655	1,002	893	.262
Year 7	15,820	14,652	1,168	1,037	.260
Sample size (total = 1,471)	739	732			

Source. MDRC calculations from New York State Department of Labor data. *Note.* Rounding may cause slight discrepancies in sums and differences. For percentage of quarters employed, estimates are adjusted by research cohort, and cluster-robust standard errors are used when calculating standard errors. Students are clustered by learning community link. For average earnings, outcomes were not adjusted for covariates, and cluster-robust standard errors were not used owing to data restrictions. Sixty-three additional students are included in analyses of academic outcomes but are not included here because earnings and employment data were not available. Social security numbers for these students were not available at random assignment (RA). For the percentage of quarters employed, Post-RA Year 1 and Pre-RA Year 1 do not include the quarter during which students were randomly assigned. Employment data from this quarter are not shown above. Yearly pre-RA earnings data are not available.

between program and control group members. During the final year of study, around 55% of the full sample was employed in New York. Overall, there is no evidence that the LCs program had a discernible effect on employment rates during the study follow-up period.

Earnings. The second panel in Table 4 provides information on average yearly earnings in New York State during the 7 years after random assignment. Although not statistically significant, the program's estimated effect on average earnings is more than $1,000 during Years 6 and 7. The estimated effect represents an increase of around 8% over the control group's base of $12,700 in Year 6 and $14,700 in Year 7. These effects are not statistically significant, meaning that if the true effect of the program was zero, there is a fairly high probability (over .26) that effect estimates as large as or larger than these could occur by chance. Despite its large sample size, this experiment is not sufficiently powered to detect an effect of $1,000. More details on issues of statistical power are provided in the Discussion section.

Subgroup Analyses

In addition to estimating the overall average effect of the opportunity to participate in KCC's LCs program, it is also of interest to know whether the program helps all eligible participants or only particular types of individuals. To limit multiple hypothesis testing, the choice of subgroups focuses on three student characteristics identified by previous researchers examining this same sample at an earlier period (Scrivener et al., 2008; Sommo et al., 2012). Specifically, we examine whether program effects varied based on baseline English placement level, gender, and race.[20]

English placement level is examined because there is generally a relationship between students' placement level and their likelihood of succeeding academically (i.e., it is an indicator of "risk"). Moreover, students' placement level determined the LCs available to them, such that students in different levels may have experienced different implementation of the program. Gender and race are explored because they are often viewed as policy relevant subgroups because men tend to underperform in community college and certain racial subgroups have been historically disadvantaged. As recommended by Bloom and Michalopoulos (2011), an explicit test of whether the estimated impacts vary significantly among groups was conducted (e.g., was the estimated effect for men statistically distinguishable from the estimated effect for women).

Table 5 shows the program's estimated effects on credit accumulation, based on students' initial English placement test level. The first panel shows the program's effects on the least at-risk group—those who began the study "college ready" in English having passed both placement tests. The second panel focuses on students who failed one baseline English test[21] and the third panel centers on student who failed both baseline English tests.[22]

Table 5 shows that the greatest estimated increases in credit accumulation occurred for students who passed both English tests at baseline and those who failed both English tests. Students in the program group who passed both English tests earned an average of 7.2 credits more than their counterparts in the control group. Students in the program group who failed both tests earned an average of 8.1 credits more than their counterparts in the control group. In contrast, students in the program group who failed one test earned essentially the same number of credits compared to their counterparts in the control group. The pattern of these results, moreover, was evident as early as the first semester following the program semester, and continued throughout the follow-up period.

Table 6 shows the program's estimated effects on degree attainment, based on students' initial English Placement Level. The program appears to have had a positive effect on the least at-risk students—students who passed both English tests—with effect estimates on graduation ranging from 9.4 percentage points to 12.2 percentage points in Years 3 through 7. In contrast, the estimated effects for students who failed one or two English tests at baseline (the bottom two panels) are much smaller.

Evidence that the program's effects on the three groups differed more than is expected by chance, is mixed. For the three groups, the *p* value for the test of variation in estimated impacts on degrees (the last column in Table 6) is below .10 in 3 out of 6 years, suggesting

[20]The information presented in this section is adapted from Sommo et al. (2012).

[21]The group of students who failed one test consisted almost entirely of those who had passed the reading test but failed the writing test. Only 0.3% of the total sample passed the writing test and failed the reading test.

[22]Similar tables for other outcome measures are available upon request. They are not included here to save space.

Table 5. Cumulative credits earned at any CUNY college, by English skills assessment at baseline, years 1-7

Outcome	Program Group	Control Group	Estimated Effect	SE	*p* Value	Subgroup Difference *p* Value[a]
Passed both English tests at baseline						
Cumulative credits earned						
Year 1	22.6	20.1	2.5	1.0	.018	.399
Year 2	37.7	33.8	4.0	2.0	.049	.427
Year 3	47.1	42.4	4.7	2.5	.062	.352
Year 4	55.0	48.5	6.6	3.0	.031	.190
Year 5	60.6	53.5	7.1	3.4	.040	.171
Year 6	64.2	56.9	7.3	3.7	.051	.160
Year 7	67.0	59.7	7.2	3.9	.066	.202
Sample size (total = 445)	225	220				
Failed one English test at baseline						
Cumulative credits earned						
Year 1	19.5	18.3	1.2	0.9	.180	
Year 2	32.7	31.4	1.3	1.7	.450	
Year 3	40.0	39.7	0.4	2.2	.859	
Year 4	45.3	45.3	0.0	2.5	.991	
Year 5	49.7	49.8	−0.1	2.8	.961	
Year 6	53.1	53.4	−0.4	3.1	.908	
Year 7	55.5	55.8	−0.3	3.3	.932	
Sample size (total = 704)	347	357				
Failed both English tests at baseline						
Cumulative credits earned						
Year 1	19.1	15.8	3.3	1.3	.015	
Year 2	31.7	27.1	4.6	2.4	.057	
Year 3	39.6	35.2	4.5	3.2	.167	
Year 4	45.2	39.5	5.8	3.8	.126	
Year 5	49.3	42.3	7.0	4.1	.091	
Year 6	52.7	44.7	8.0	4.3	.065	
Year 7	55.0	46.9	8.1	4.5	.075	
Sample size (total = 385)	197	188				

Source. MDRC calculations from City University of New York (CUNY) Institutional Research Database data. *Note.* Rounding may cause slight discrepancies in sums and differences. Estimates are adjusted by research cohort. Cluster-robust standard errors are used when calculating standard errors. Students are clustered by learning community link. The H-statistic was used to calculate the Subgroup Difference *p* Value, as described in Greenberg, Meyer, and Wiseman (1994). Cumulative credits include both college-level and developmental credits. Values of zero credits attempted/earned have been imputed for nine students for whom CUNY data were unavailable.

[a]The Subgroup Difference *p* Value is the *p* value for a test of variation in program impacts across subgroups.The null hypothesis is that the program's effects are homogeneous across groups.

the program had different impacts across the groups. For credits earned (Table 5); however, the difference in estimated impacts had a minimum *p* value of .16. Nonetheless, the pattern is consistent and the magnitude of the differences is meaningful.

Together, these analyses suggest positive academic impacts for students who either passed both tests or failed both tests and weaker evidence that the program improved

Table 6. Degree earned at any college, by English skills assessment at baseline, years 1–7

Outcome	Program Group	Control Group	Estimated Effect	SE	p Value	Subgroup Difference p Value[a]
Passed both English tests at baseline						
Earned a degree (%)						
Year 1	0.0	0.0	0.0			
Year 2	13.5	8.9	4.6	3.2	.152	.292
Year 3	32.1	21.3	10.8	4.8	.026	.110
Year 4	40.1	28.6	11.5	4.6	.014	.069
Year 5	46.7	34.9	11.8	4.4	.008	.052
Year 6	50.3	38.1	12.2	4.4	.006	.068
Year 7	53.0	43.6	9.4	4.9	.057	.303
Sample size (total = 445)	225	220				
Failed one English test at baseline						
Earned a degree (%)						
Year 1	0.0	0.0	0.0			
Year 2	4.1	5.0	−1.0	1.5	.534	
Year 3	16.1	17.1	−0.9	2.9	.742	
Year 4	23.3	23.8	−0.5	3.1	.879	
Year 5	27.4	28.9	−1.5	3.3	.649	
Year 6	32.3	32.8	−0.5	3.2	.882	
Year 7	37.2	37.0	0.2	3.3	.944	
Sample size (total = 704)	347	357				
Failed both English tests at baseline						
Earned a degree (%)						
Year 1	0.0	0.0	0.0			
Year 2	1.5	1.6	−0.1	1.1	.900	
Year 3	13.9	11.5	2.4	3.6	.505	
Year 4	17.0	17.9	−0.9	4.1	.826	
Year 5	23.0	19.5	3.5	4.5	.436	
Year 6	26.0	21.1	4.9	5.0	.323	
Year 7	28.4	26.1	2.4	5.0	.637	
Sample size (total = 385)	197	188				

Source. MDRC calculations from City University of New York (CUNY) Institutional Research Database and National Student Clearinghouse data. *Note.* Rounding may cause slight discrepancies in sums and differences. Estimates are adjusted by research cohort. Cluster-robust standard errors are used when calculating standard errors. Students are clustered by learning community link. The H-statistic was used to calculate the Subgroup Difference P-value, as described in Greenberg, Meyer, and Wiseman (1994).

[a]The Subgroup Difference p Value is the p value for a test of variation in program impacts across subgroups. The null hypothesis is that the program's effects are homogenous across groups.

outcomes for students who passed only one test. One possible explanation for the pattern is that the variation is being driven by differences in implementation of the LCs model. A faculty survey conducted for one cohort in the study suggests that there was less collaboration among some faculty teaching learning communities for students who failed one test during that semester. A small student survey conducted for one cohort in the study also suggests potentially different experiences for these students; although a much larger student survey conducted 1 year after random assignment does not.[23]

Another possible explanation is that the students who failed only one test had different characteristics compared with the other groups of students, in addition to their baseline English skills. Baseline demographic data were examined for each group of students to assess whether this might be true (results not tabled). Most observed baseline characteristics were either similar between all three groups of students or varied monotonically, rendering them unlikely sources of the pattern of impacts observed. However, on characteristics of gender, race, and ethnicity, and rate of high school diploma receipt, there is some evidence that students who had passed one placement test differed from those who passed or failed both.

Ultimately, the data are too limited to provide conclusive evidence. This evaluation cannot determine whether the pattern of results observed in Table 5 and Table 6 stems from differences in program implementation for the three subgroups, differences in students' characteristics or academic preparation, chance, or other factors. Consequently, these finding suggests that future evaluations of LCs (or related programs) should examine variation in program effects by initial skill levels to see if this type of findings holds in other settings.

With respect to gender and race, we do not find any evidence that the LCs program is more effective for one gender (not shown in tables). There is some evidence that LCs were more effective for members of a catchall "other" race category, comprised of students who, when asked their race/ethnicity prior to being randomly assigned, responded "other" ($n = 72$), checked more than one race ($n = 20$), or responded American Indian ($n = 2$). However, uncertainty around the effect estimates for this group is very high owing to its small size.

Overall, this study finds some evidence that LCs effects may vary across different types of individuals. However, the findings are not consistent enough within this study, nor confirmed across other studies (Visher et al., 2012), to draw definitive conclusions from these analyses.

Limitations

Before summarizing the findings on the effects of KCC's LCs program, a few important limitations to this research are described.

External Validity. This article focuses on one LC program implemented at one community college. The randomized design offers a major improvement over past studies with respect to *internal validity* (for a definition, see Campbell & Stanley, 1963). The program served a diverse group of students and operated 40 LCs sections. Nonetheless, inference regarding the effects of LCs operating at other colleges on other populations of students requires major speculation.

Teacher Effects. In this study, random assignment was conducted at the student level, creating two similar groups of students for future comparisons. Instructors, however, were not randomly assigned. Thus, the selection of instructors may have influenced the estimated effects of LCs. The presented effect estimates represent the combined effect of LCs and the types of instructors delivering those services, or teacher effects, whereas we are most

[23] See Scrivener et al. (2008) for more information on these data sources.

interested in the effects of LCs alone (Weiss, 2010). To some extent this concern is mitigated because many learning community instructors also taught stand-alone versions of these classes, sometimes to control group students. However, it is unclear how frequently control group students took classes with instructors who also taught the program group. It is also noteworthy that this concern exists in all past evaluations of LCs, so the elimination of student selection bias through the randomized design is a major improvement over past attempts to estimate the effects of LCs.

Summary

KCC's LCs program boosted students' academic progress—students offered this one-semester program earned significantly more credits than their control group counterparts—a result that was maintained 7 years after random assignment. There is some limited evidence that the program increased students' chances of earning a degree, especially for those students that began the study prepared for college-level English. There is not strong evidence of a statistically discernible effect on employment or earnings. These results may not be surprising given the magnitude of the effects on academic outcomes and the limited statistical power to detect effects on employment and earnings.

DISCUSSION

Community colleges provide access to higher education for millions of students, yet a large proportion of these students do not earn a college degree. Few strategies that have been tested using random assignment have been shown to substantially increase students' academic progress or graduation rates. The evaluation of the LC program described in this article marks the first large-scale randomized trial of a LC program, and the first large-scale randomized trial in a community college. It is unique not only for the long duration of the follow-up period and the ability to track both academic outcomes and employment outcomes, but also because the program itself appears to have produced a long-term effect on students' academic progress.

Other Randomized Trials of Learning Communities Programs

These positive impacts, however, may not be representative of the effects of LC more generally. The final report on the National Center for Postsecondary Research's Learning Communities Demonstration describes results from randomized trials of six LC programs operated at six colleges[24] and suggests modest short-term impacts, on average, with KCC's estimated impacts standing out as the largest.

Several factors distinguish the LCs at KCC. They were particularly comprehensive (including additional student support services), and some of the services extended into the subsequent intersession period between the standard fall and spring semesters. In addition, the research sample had important distinguishing characteristics. For example, the evaluation explicitly recruited students intending to enroll in college full time and

[24]The six programs include the subsample of developmental education students in the evaluation described in this article.

Table 7. Necessary returns to degree in year 7, assuming earnings effects only through degree receipt

Sample Size and MDE	Hypothetical Degree Effect (Percentage Points)	Necessary Returns to Degree ($)	Necessary Returns to Degree (%)
$N = 1{,}500$	3.3	87,091	592%
MDE = $2,874	5.0	57,480	391%
	10.0	28,740	196%
	15.0	19,160	130%
	20.0	14,370	98%
$N = 10{,}000$	3.3	33,727	229%
MDE = $1,113	5.0	22,260	151%
	10.0	11,130	76%
	15.0	7,420	50%
	20.0	5,565	38%

Note. Minimum Detectable Effects assume power is 80%, the significance level is 5%, and the standard deviation of the earnings outcome is $19,880.

included both developmental and college-ready English students. The program also had unusually strong support from the college leadership. Therefore, although the KCC results are encouraging, it is not clear how easily the positive impacts could be replicated at other institutions.

Reflections on Detecting the Effects of Educational Intervention on Earnings

The employment and earnings data raises further questions. Policymakers and academics alike believe in the importance of postsecondary education for success in the labor market. This study, however, suggests that identifying the causal effect of an educational intervention on earnings may be more difficult than generally acknowledged. This study, and other studies in education, may be underpowered to identify anything but very large effects on earnings. To illustrate, consider the minimum detectable effect (MDE)—the smallest true effect that a study is likely to detect.[25] In this study, the MDE is around $2,900, representing a 20% increase over the control group's average earnings in Year 7 ($14,652). This means that if the true average effect on earnings was $2,900, then there was an 80% chance that the study would find a positive effect at the .05 significance level. To put the size of the MDE in context, consider that Dadgar and Weiss (2012) estimated the average economic returns to earning an associate degree at around 6%.

To offer insight into designing evaluations of college programs aimed at improving economic outcomes, Table 7 presents scenarios that pose the following question: "If a college reform programs' only path to improved earnings is through degree-attainment, what would the average economic returns to a degree have to be in order for a study to be adequately powered?"

The first row of the first panel in Table 7 parallels our study, where n is around 1,500, the MDE is around $2,874, and the estimated degree effect is 3.3 percentage points. If effects on earnings derive solely from effects on earning a degree, then the first row shows that the required returns to a degree would need to be $87,091 in the 7th year, which

[25]For simplicity of exposition, clustering is ignored here. Clustering only exacerbates the issues described.

constitutes a 592% increase over the average earnings in the control group. In fact, the last row of Table 7 shows that even in an evaluation with 10,000 students where an intervention increases degree receipt by 20 percentage points, the return to a degree would have to be $5,565, or 38%, for the study to be adequately powered; this remains well above current estimates of the average returns to an associate's degree.

This may imply that for a higher education intervention to yield detectable effects on earnings, the intervention may need to provide services that more directly link to the labor market (e.g., enhanced career and employment counseling) and/or target degree completion in fields that offer higher than average returns to a degree (e.g., nursing), and also that evaluations probably require very large samples.

Conclusion

The study reported on here shows that a short-term intervention in community colleges can have long-term educational effects, and potentially affect degree completion. Although randomized trials of LCs have produced varied results, some findings presented here reinforce findings from other research. Notably, the positive program effects during intersessions are consistent with findings from several other evaluations. For example, random assignment evaluations of unrelated programs have also found significant impacts on intersession enrollment and intersession credit accumulation (Patel & Rudd, 2012; Richburg-Hayes et al., 2009; Scrivener, Weiss, & Sommo, 2012). Taken together, these findings suggest that intersessions may be well suited to educational interventions aimed at helping students progress toward a degree more quickly. Program developers may want to design new interventions with intersession enrollment in mind, as students seem responsive to interventions aimed to increase enrollment during these times.

More broadly, practitioners should interpret the results from this study in the context of the more varied results from other randomized trials of LCs. Given these mixed findings, LC practitioners should be encouraged, but cautious. They would also benefit from rigorous research on the impact of LCs in their own colleges as well as detailed qualitative research on the experiences of students in their LC classrooms.

Randomized trials are still relatively rare in community colleges, and few of those that are conducted result in impacts still evident after 1 or 2 years—and even fewer have led to such a long period of follow-up or the analysis of employment-related outcomes. Identifying strong programs—or periods in students' academic careers—that hold the potential to affect long-term educational outcomes is a pressing need in community college research.

REFERENCES

Adelman, C. (2004). *Principal indicators of student academic histories in postsecondary education, 1972–2000* (pp. 1–161). Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Adelman, C. (2005). *Moving into town—and moving on: The community college in the lives of traditional-age students*. Washington, DC: U.S. Department of Education.

Angrist, J., Imbens, G., & Rubin, D. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*, 444–455.

Attewell, P. A., Lavin, D. E., Domina, T., & Levey, T. (2006). New evidence on college remediation. *Journal of Higher Education*, *77*, 886–924.

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

Bailey, T., Kienzl, G., & Marcotte, D. E. (2004, August). *The return to a sub-baccalaureate education: The effects of schooling, credentials and program of study on economic outcomes*. New York, NY: CCRC.

Barrow, L., & Rouse, C. E. (2005, February). *Do returns to schooling differ by race and ethnicity?* (Working Paper No. 2005-02). Chicago, IL: Federal Reserve Bank of Chicago.

Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review*, *8*, 225–246. doi:10.1177/0193841×8400800205

Bloom, H. S., & Michalopoulos, C. (2011). When is the story in the subgroups? Strategies for interpreting and reporting intervention effects on subgroups. *Prevention Science*, *14*, 179–188.

Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis*, *29*, 30–59.

Brock, T., LeBlanc, A., & MacGregor, C. (2005). *Promoting student success in community college and beyond: The Opening Doors Demonstration*. New York, NY: MDRC.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171–246). Chicago, IL: Rand McNally.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, *69*, 1127–1160.

Carneiro, P., Heckman, J. J., & Vytlacil, E. J. (2011). Estimating marginal returns to education. *American Economic Review*, *101*, 2754–2781.

Center for Community College Student Engagement. (2012). *A matter of degrees: Promising practices for community college student success (a first look)*. Austin: The University of Texas at Austin, Community College Leadership Program.

College Board Advocacy and Policy Center. (2012). *Trends in college pricing 2012*. New York, NY: College Board.

Crits-Christoph, P., Tu, X., & Gallop, R. (2003). Therapists as fixed versus random effects-some statistical and conceptual issues: A comment on Siemer and Joormann (2003). *Psychological Methods*, *8*, 518–523.

Dadgar, M., & Weiss, M. J. (2012). *Labor market returns to sub-baccalaureate credentials: How much does a community college degree or certificate pay?* (CCRC Working Paper No. 45). New York, NY: Community College Research Center.

Duke, A.-E., & Strawn, J. (2008). *Overcoming obstacles, optimizing opportunities: State policies to increase postsecondary attainment for low-skilled adults*. Boston, MA: Jobs for the Future.

Dynarski, S. (2008). Building the stock of college-educated labor. *Journal of Human Resources*, *43*, 576–610.

Engstrom, C. M., & Tinto, V. (2008). Learning better together: The impact of learning communities on the persistence of low-income students. *Opportunity Matters*, *1*, 1–17.

Goldberg, B., & Finkelstein, M. (2002). Effects of a first-semester learning community on nontraditional technical students. *Innovative Higher Education*, *26*, 235–249.

Gordon, T. W., Young, J. C., & Kalianov, C. J. (2001). Connecting the freshman year experience through learning communities: Practical implications for academic and student affairs units. *College Student Affairs Journal*, *20*, 37–45.

Greenberg, D., Meyer, R., & Wiseman, M. (1994). Multisite employment and training program evaluation: A tale of three studies. *Industrial and Labor Relations Review*, *47*, 679–691.

Hoachlander, G., Sikora, A. C., & Horn, L. (2003). *Community college students: Goals, academic preparation, and outcomes. postsecondary education descriptive analysis reports*. Available from http://nces.ed.gov/pubs2003/2003164.pdf

Horn, L., Berger, R., & Carroll, C. D. (2004). *College persistence on the rise? Changes in 5-year degree completion and postsecondary persistence rates between 1995 and 2000*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Jacobson, L., & Mokher, C. (2009). *Pathways to boosting the earnings of low-income students by increasing their educational attainment*. Washington, DC: The Hudson Institute Center for Employment Policy.

Jepsen, C., Troske, K., & Coomes, P. (2009). *The labor-market returns to community college degrees, diplomas, and certificates* (Discussion Paper Series DP 2009–08). Lexington: University of Kentucky, Center for Poverty Research.

Kane, T. J., & Rouse, C. E. (1995). Labor-market returns to two and four-year college. *American Economic Review*, *85*, 600–614.

Minkler, J. (2002). ERIC review: Learning communities at the community college. *Community College Review*, *30*(3), 46–63.

National Center for Education Statistics. (2011). *Total fall enrollment in degree-granting institutions, by control and level of institution: 1963 through 2010*. Retrieved from http://nces.ed.gov/programs/digest/d11/tables/dt11_199.asp

Pals, S. L., Murray, D. M., Alfano, C. M., Shadish, W. R., Hannan, P. J., & Baker, W. L. (2008). Individually randomized group treatment trials: A critical appraisal of frequently used design and analytic approaches. *American Journal of Public Health*, *98*, 1418–1424. doi:10.2105/ajph.2007.127027

Patel, R., & Rudd, T. (2012). *Can scholarships alone help students succeed? Findings from the Performance-Based Scholarship Demonstration in New York*. New York, NY: MDRC.

Radford, A. W., Berkner, L., Wheeless, S. C., & Shepherd, B. (2010). *Persistence and attainment of 2003-04 beginning postsecondary students: After 6 years. First look* (NCES 2011-151). Washington, DC: NCES. Available from http://nces.ed.gov/

Richburg-Hayes, L., Brock, T., LeBlanc, A., Paxson, C., Rouse, C. E., & Barrow, L. (2009). *Rewarding persistence: Effects of a performance-based scholarship program for low-income parents*. New York, NY: MDRC.

Richburg-Hayes, L., Visher, M. G., & Bloom, D. (2008). Do learning communities effect academic outcomes? Evidence from an experiment in a community college. *Journal of Research on Educational Effectiveness*, *1*, 33–65.

Roberts, C., & Roberts, S. A. (2005). Design and analysis of clinical trials with clustering effects due to treatment. *Clinical Trials*, *2*, 152–162. doi:10.1191/1740774505cn076oa

Schelchter, M. D., & Forsythe, A. B. (1985). Post-hoc selection of covariates in randomized experiments. *Communications in Statistics—Theory and Methods*, *14*, 679–699.

Schochet, P. Z. (2008). *Technical methods report: Guidelines for multiple testing in impact evaluations* (NCEE 2008-4018). Washington, DC: National Center for Education Evaluation and Regional Assistance. Available from http://ies.ed.gov/ncee/pubs/.

Scrivener, S., Bloom, D., LeBlanc, A., Paxson, C., Rouse, C. E., & Sommo, C. (2008). *A good start: Two-year effects of a freshmen learning community program at Kingsborough Community College*. New York, NY: MDRC.

Scrivener, S., Weiss, M. J., & Sommo, C. (2012). *What can a multifaceted program do for community college students?* New York, NY: MDRC.

Serlin, R. C., Wampold, B. E., & Levin, J. R. (2003). Should providers of treatment be regarded as a random factor? If it ain't broke, don't "fix" it: A comment on Siemer and Joormann (2003). *Psychological Methods*, *8*, 524–534.

Siemer, M., & Joorman, J. (2003a). Assumptions and consequences of treating providers in therapy studies as fixed versus random effects: Reply to Crits-Christoph, Tu, and Gallop (2003) and Serlin, Wampold, and Levin (2003). *Psychological Methods*, *8*, 535–544.

Siemer, M., & Joorman, J. (2003b). Power and measures of effect size in analysis of variance with fixed versus random nested factors. *Psychological Methods*, *8*, 497–517.

Smith, B. L. (2001). The challenge of learning communities as a growing national movement. *Peer Review*, *3*(4), 4–8.

Smith, B. L., MacGregor, J., Matthews, R., & Gabelnick, F. (2004). *Learning communities: Reforming undergraduate education*. San Francisco, CA: Jossey-Bass.

Sommo, C., Mayer, A. K., Rudd, T., & Cullinan, D. (2012). *Commencement day: Six-year effects of a freshman learning community program at Kingsborough Community College*. New York, NY: MDRC.

Stassen, M. L. A. (2003). Student outcomes: The impact of varying living-learning community models. *Research in Higher Education*, *44*, 581–613.

Taylor, K., with Moore, W. S., MacGregor, J., & Lindblad, J. (2003). *Learning community research and assessment: What we know now* (National Learning Communities Project Monograph Series). Olympia: The Washington Center.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Education Research*, *45*, 89–125.

Tinto, V. (1997). Classrooms as communities: Exploring the educational character of student persistence. *Journal of Higher Education*, *68*, 599–623.

Tinto, V. (1998). *Learning communities and the reconstruction of remedial education in higher education*. Paper presented at the Conference on Replacing Remediation in Higher Education at Stanford University, Stanford, CA. Sponsored by the Ford Foundation and United States Department of Education.

Tinto, V., & Russo, P. (1994). Coordinated studies programs: Their effect on student involvement at a community college. *Community College Review*, *22*(2), 16–25.

Visher, M. G., & Teres, J. (2011). *Breaking new ground: An impact study of career-focused learning communities at Kingsborough Community College*. New York, NY: MDRC.

Visher, M. G., Weiss, M. J., Weissman, E., Rudd, T., & Wathington, H. D. (2012). *The effects of learning communities for students in developmental education: A synthesis of findings from six community colleges*. New York, NY: National Center for Postsecondary Research. Teachers College, Columbia University.

Weiss, M. J. (2010). The implications of teacher selection and the teacher effect in individually randomized group treatment trials. *Journal of Research on Educational Effectiveness*, *3*, 381–405.

Weiss, M. J., Lockwood, J. R., & McCaffrey, D. F. (2014, April). *Estimating the standard error of the impact estimator in individually randomized trials with clustering* (MDRC Working Papers on Research Methodology). New York, NY: MDRC.

Wilmer, E. (2009). The influence of learning communities on the interaction levels of developmental English students. *The Journal of the Virginia Community Colleges*, pp. 55–67.

Zhao, C.-M., & Kuh, G. D. (2004). Adding value: Learning communities and student engagement. *Research in Higher Education*, *45*, 115–138.

APPENDIX

Estimation and Sensitivity Analyses

The analyses presented in this article are intent-to-treat (ITT), comparing the outcomes of all program group students to all control group students, regardless of compliance with treatment assignment. The estimand is the effect of being offered the opportunity to participate in KCC's LCs. This is not necessarily the same as the effect of experiencing KCC's LCs because at least 15% of the program group did not fully participate in an LC.[26]

[26]Seven percent of program group members did not enroll at all during the program semester. An additional 8% enrolled but were not coenrolled in LC courses as of the college's add/drop deadline. Nine percent of control group members did not enroll at all during the program semester. Less than 1% of control group members "crossed over," meaning they enrolled in an LC. In previous analyses on the 2-year effects of KCC's LC program on the same sample presented on here, Richburg-Hayes et al. (2008) used instrumental variables to estimate the local average treatment effect and observed that the instrumental variables estimates were similar in magnitude and direction to the [ITT] estimates.

The ITT estimand reflects the effect of offering LCs in a real-world, school-based setting, where compliance is imperfect.[27]

The Impact Model

In this evaluation, individual students were randomly assigned to the opportunity to enroll in LCs. Equation 1 describes the model used to estimate the average effect of this offer on student outcomes:

$$Y_i = \alpha + \beta T_i + \boldsymbol{Coh_i}\boldsymbol{\delta} + \boldsymbol{X_i}\boldsymbol{\gamma} + \varepsilon_i \tag{1}$$

Here, Y_i is the outcome of interest (e.g., credits earned) for student i; T_i is a treatment indicator for student i; $\boldsymbol{Coh_i}$ is a vector of dummy indicators representing a students' random assignment cohort; and $\boldsymbol{X_i}$ is a vector of baseline characteristics for student i (included in sensitivity analyses only).[28]

The coefficient of interest, β, represents the causal effect of the program offer. The study sample includes four subsequent cohorts of students (fall 2003, spring 2004, fall 2004, spring 2005); cohort dummies are included in the impact model to reflect the different time points that students entered the study. The vector of client characteristics ($\boldsymbol{X_i}$) is not included in the main impact analyses but is included in sensitivity analyses described in the next section of the appendix.[29] Large-scale randomization increases the likelihood that the characteristics (both observable and unobservable) of treatment and control group members are similar at the outset of the study and ensures that there are not systematic differences between research groups. Thus, student-level covariates are not necessary to obtain an unbiased impact estimate.

Sensitivity analyses were conducted including selected baseline characteristics that prior research has found to correlate with academic outcomes (Horn et al., 2004). The indicators used as covariates in the sensitivity analysis were delayed postsecondary

[27]One alternative would be to use instrumental variables to estimate local average treatment effect, as in Richburg-Hayes et al. (2008). Such analyses require the assumption that the treatment has no effect on students assigned to the program who did not experience the program, referred to as "no-shows" or "never-takers" (Angrist, Imbens, & Rubin, 1996; Bloom, 1984). In this study, this assumption may not hold because program participation is defined several weeks into the semester at a point when the program may already have affected students' outcomes. Moreover, our best program participation measure is defined as coenrolling in LC classes; however, the intervention includes services that may have been received by students who did not coenroll (e.g., textbook voucher, counseling). In other words, the "exclusion restriction" may not hold, so the analyses rely on assumptions that may be violated.

[28]In cases where the value of one or more baseline characteristic could not be conclusively determined due to missing student data, values were imputed to the pooled sample mean. A separate set of "missing" dummy indicators were created when a baseline variable was missing. In addition to the baseline characteristics themselves, these missing dummy indicators were also included as covariates in the sensitivity analyses.

[29]The decision of whether the *main* analyses would include X_i was made prior to examining the results of the impact model to reduce any potential researcher bias.

enrollment,[30] financial independence,[31] employed in any of the four quarters prior to random assignment, had one or more children, had children and was unmarried, and either had not completed Grade 12 or had not received a high school diploma. English and Math placement test scores were also used as covariates in the sensitivity analysis. Researchers often include covariates that are correlated with the outcome of interest in their impact model to improve the precision of the effect estimate (Bloom, Richburg-Hayes, & Black, 2007). The greater the correlation between covariates and the target outcomes, the greater the precision gains. Researchers also sometimes include covariates in their impact model to account for observed baseline imbalances; however, there is disagreement about how to do this and even whether it is appropriate in a randomized experiment (Schelchter & Forsythe, 1985).

Sensitivity analyses were conducted for two reasons: first, to understand the precision gains resulting from including covariates in higher education experiments, where less (relative to k-12) historical guidance on their realized benefits exists; second, as noted in the text, despite randomization, there is evidence of baseline differences in the characteristics of program and control group members. Although selecting covariates strictly on the basis of observed pretreatment assignment differences can bias the standard error of the effect estimator (Schelchter & Forsythe, 1985), we are interested in the robustness of our results to a model that includes covariates selected based on prior evidence of a relationship with higher education outcomes, some of which were observed to differ across research groups at the outset of the study, others of which were not.

In the main analyses the error term, ε_i, is modeled using cluster robust standard errors to account for the potential lack of independence caused by shared experiences (e.g., teacher effects) in LCs clusters (for more details and explanation behind the model, see Visher et al., 2012).[32] Additional analyses were conducted using ordinary least squares as a sensitivity check.[33]

[30]Following Horn et al. (2004), students were considered to have delayed postsecondary enrollment if they graduated from high school and enrolled at college in different calendar years. An exception was made for 27 students who graduated from high school in August through November and enrolled in college in January through March of the following year. These students were not treated as having delayed postsecondary enrollment.

[31]Students were considered financially independent if any of the following applied: was 24 or older as of random assignment, was married, or had children (excluding spouse).

[32]Recent research on this topic suggests that cluster robust standard errors can be upward biased in individually randomized group treatment trials such as this one. This occurs because students were nonrandomly sorted into LCs after random assignment resulting in the *appearance* of dependency of observations within clusters, which need not be accounted for in the analyses since it is artificial (Weiss, Lockwood, & McCaffrey, 2014).

[33]It is reasonable to use OLS with this study design *if* the desired inference is with respect to the effect of the specific LCs observed in this study (taught by the specific instructors observed in the study) rather than to a hypothetical superpopulation of LCs from which these particular LCs could have been randomly drawn (Siemer & Joorman, 2003b). Technically, Siemer and Joorman (2003b) recommended a constrained fixed effect model that will often produce standard errors that are smaller in magnitude than OLS; however, recent work on this topic by Weiss et al. (2014) suggests that the constrained fixed effects approach can produce standard errors that are downward biased (owing to the nonrandom a sorting of students into clusters), and thus OLS may be a reasonable alternative for those interested in the fixed effects inference.

Sensitivity Analyses

Two sets of sensitivity analyses were conducted to assess the robustness of the main findings. First, as previously described, impact analyses were conducted with and without student characteristics (measured prior to random assignment) as covariates. These analyses are intended to assess the robustness of findings to the inclusion of covariates, given greater than expected baseline imbalances. These analyses also may inform future research about the potential precision gains achieved by including available covariates in higher education experiments.

Occasionally the inclusion of covariates shifted an impact estimate's p value across the .05 or .10 thresholds that are often used to determine statistical significance. With respect to credit accumulation, for example, after including covariates the estimated effect on cumulative credits earned in Year 7 drops from 4.0 to 3.4 credits, resulting in an increase in p value from .092 to .117. In the opposite direction, the estimated effect on 6-year graduation rates became statistically significant after including covariates—the p value dropped from .104 to .044 due to a 23% decrease in the standard error of the impact estimate (the impact estimate remained about the same).

Including the set of "risk factor" covariates yields important precision gains with respect to credit accumulation and even larger gains in regard to degree completion.[34] In future large-scale studies, researchers are encouraged to consider prespecifying that their main impact analyses include this set (or a similar set) of covariates when estimating program effects.

In the second sensitivity analyses, the robustness of the academic outcome inference findings are examined with respect to the approach used to estimate the standard error of the effect estimate—either using cluster robust standard errors or using ordinary least squares. Education researchers have paid limited attention to this issue in individually randomized experiments; the topic has been debated in the field of psychotherapy without consensus (Crits-Christoph, Tu, & Gallop, 2003; Pals et al., 2008; Roberts & Roberts, 2005; Serlin, Wampold, & Levin, 2003; Siemer & Joorman, 2003a, 2003b). This model choice affects only the standard error of the effect estimate, and in this particular evaluation the decision affects only the 6-year graduation results, which moved from a p value just above .10 to one that is below .10.[35]

Overall, the sensitivity analyses do not dramatically change the interpretation of the overall findings on the effectiveness of the learning community program.

[34]The standard error of the impact estimate decreased by an average of 10.4% for credit accumulation and 18.0% for degree completion.

[35]With respect to credit accumulation, the two approaches yield nearly identical standard errors and p values, with differences in p values all below .01. With respect to graduation, the p values using ordinary least squares are, on average, .07 smaller than when using cluster robust standard errors. Substantively, this does not change the story. With respect to employment, p values are generally slightly smaller using ordinary least squares, with the largest decrease being .025.