The Impact of a Classroom-Based Guidance Program on Student Performance in Community College Math Classes

Kristin F. Butcher
Wellesley College

Mary G. Visher
MDRC

Passing through remedial and required math classes poses a significant barrier to success for many community college students. This study uses random assignment to investigate the impact of a "light-touch" intervention, where an individual visited math classes a few times during the semester, for a few minutes each time, to inform students about available services. Entire class sections, rather than individuals, were randomly assigned to program and control groups, reducing the administrative burden for the college of a randomized-controlled experiment. This study finds that the intervention increased students' use of tutoring services and reduced math class withdrawal rates, but had no effect on overall pass rates. The program did, however, increase the math class pass rates for part-time students, who represented almost 50% of the participants.

Keywords: *community college, remedial education, random assignment evaluation*

COMMUNITY colleges currently enroll about 34% of U.S. undergraduates, a disproportionate share of them from low-income families (Knapp, Kelly-Reid, & Ginder, 2009). At least half of these students will fail to earn a college credential, and success rates are even lower for students deemed in need of remediation (Attewell, Lavin, Domina, & Levey, 2006; Hoachlander, Sikora, & Horn, 2003). For many students, passing out of remedial math, in particular, proves an insurmountable barrier, with many dropping out before they even attempt a college credit-bearing transfer-level course (see Brock, 2010; Visher, Schneider, Wathington, & Collado, 2010; Visher, Wathington, Richburg-Hayes, & Schneider, 2008).

Many colleges have identified moving students through remedial and gatekeeper courses as key goals, and there are a number of studies

examining different methods to improve outcomes along this dimension (Zachry & Schneider, 2008, 2011). This study presents the results of one such effort: the "Beacon Mentoring Program" at South Texas College (STC).

Although this study shows the results of a single-site evaluation, there are particular features of this program that we believe make it worth bringing to the attention of a national audience. First, and most importantly, the Beacon Program represents a "light-touch" intervention that was designed by STC to reach a large number of students in its high-failure math courses in as low cost a manner as possible. Rather than creating a costly new one-on-one traditional mentoring program, the College trained current employees to go directly into the classroom to deliver information about existing services, like the tutoring center, and to urge students to avail

themselves of these services. The "Beacon Volunteers"[1] presented this information to students during a few brief visits over the course of the semester, and these presentations formed the bulk of the Volunteers' interactions with students. A key question in deciding how much weight to place on results of any given intervention is whether other institutions could hope to harness the resources necessary to replicate the program. We argue that the main elements of this intervention—in particular, bringing information into the classroom where it is sure to reach the students—are within reach of virtually all educational institutions.

In addition to the low-cost program design, the randomization methodology itself should be of interest. The randomization scheme used placed a relatively low burden on the institution: instead of randomly assigning individual students to classrooms in the program or control group, more than 2,000 students were allowed to register for 83 sections of math classes as in any normal registration period. Then, a block randomization scheme was used to allocate these 83 sections of developmental and gatekeeper math courses to the program group, which received visits from a Beacon Volunteer, or the control group, which did not. In post-secondary settings, whole-class random assignment is relatively rare but is much less burdensome for the institution than individual random assignment, and the results here demonstrate that whole-class randomization delivered program and control groups that balanced observable characteristics. Whole-class randomization is relatively rare but is much less burdensome for the institution than individual random assignment, and the results here demonstrate that whole-class randomization delivered program and control groups that balanced observable characteristics. If whole-class random assignment is possible, demonstrating to administrators that it can deliver evaluation results with relatively little disruption of standard procedures may reduce resistance to rigorous evaluation. In addition, unlike most randomization studies, where faculty and students self-select to participate in an experiment, raising questions of whether the program would have similar effects if applied to everyone, this study includes every student and every faculty member taking or teaching the targeted levels of math.

Finally, although STC is just one location and may have some differences from the typical U.S. community college,[2] the challenges faced by students in passing developmental and gatekeeper math courses are similar to many other institutions. Thus, lessons learned from this low-cost intervention at STC are likely to be useful to other administrators and policymakers.[3]

This study evaluates the effects of the Beacon Program on use of tutoring services, math class withdrawal and pass rates, and registration for the subsequent semester, for students overall, and part-time versus full-time students. In addition, a change in the rollout of the evaluation allows us to compare the impact of the Beacon Program for students who actively enrolled in an individual random assignment study with the impact for students who did not. This yields insight into whether program effects are different for the types of people who enlist in randomized-controlled experiments, an issue of external validity that frequently arises in policy discussions surrounding randomization studies. Finally, the whole-class randomization methodology allows us to control for teacher effects, and to investigate whether the program effects differ by an often-used measure of teacher quality.

We find that the Beacon Program successfully guided students to the campus tutoring center: Program group students were about 30% more likely to have visited the tutoring center. The program also reduced the withdrawal rate from math classes for all students. However, only part-time students—who made up almost 50% of the students—increased their math class pass rates. Part-time students in the program group had pass rates that were 10% higher than those in the control group. Finally, we find no evidence to suggest that the students who enlisted in a randomization study are those who drive the statistically significant effects, nor do we find evidence that the program had a differential effect by "teacher quality."

The following section describes STC and places this study in the context of other evaluations. The next section describes the Beacon Program, which is followed by the "Data and Methodology" section. The section "Estimated Program Effects" presents the estimated program impacts for students overall, and for part-time versus full-time students. The section "Extensions:

Differences in Program Impacts by Enrollee/Nonenrollee Status and by Teacher Quality" presents extensions of the evaluation—comparing students who actively enlisted in a random assignment study to those who did not, and examining whether the program had a differential effect by an available measure of teacher quality. The section "Discussion and Avenues for Future Research" concludes.

Background: STC and Other Studies

STC is located in McAllen, Texas, near the Mexican border. It serves a population of about 21,000, mostly (94%) Hispanic, students on its five campuses. The largest of these is the Pecan Campus, which serves about 10,000 students, and was the site of the Beacon Program and evaluation.[4] Although the student population it serves is more Hispanic than at the "average" community college, the challenges facing students and administrators at STC are in many ways characteristic of this educational setting.[5] At STC, as at many community colleges, the majority of incoming students are referred to developmental education, and of those, most are referred to developmental math.[6] Of those who take developmental math classes at STC, only about half pass the course (see Table 1), a rate that is comparable with the national average at community colleges.

As part of their response to these issues, in 2005, STC joined 26 other community colleges in the first round of Achieving the Dream: Community Colleges Count, a national initiative funded by Lumina Foundation for Education. Achieving the Dream is designed to encourage colleges to undertake institutional, data-driven improvement processes to develop priorities for reform, with a special focus on the needs of students in developmental education and other at-risk student populations. The Beacon "Mentoring" Program, described in detail in the next section, was the flagship strategy developed at STC.

Given that community colleges serve such a large proportion of U.S. higher education students, that completion rates tend to be low, and that developmental math seems to be a particular barrier to completion, community colleges have become a focal point for educational interventions and evaluations. Research has examined questions from "what is the impact of being referred to developmental education on students' outcomes,"[7] to what types of specific programs can improve students' outcomes generally, and in developmental education courses more specifically?

Research suggests that students often lack "college knowledge," and often do not know the requirements to graduate or whether the courses they are taking will fulfill those requirements (Goldrick-Rab, 2010). Thus, many of the educational interventions designed to improve student outcomes have counseling, advising, or mentoring components to them. For example, Bettinger and Baker (2011) randomly assigned "coaches" to students at a variety of colleges and universities. These coaches contacted students regularly and helped them to devise plans such that their daily activities supported their long-term goals. The study finds that students who received coaches were more likely to persist in college. Other interventions often pair some form of coaching or guidance with other types of supports. For example, a number of studies have examined the impact of combining advising with financial incentives for improved academic outcomes and have found positive program impacts (Angrist, Lang, & Oreopoulos, 2009; Brock & Richburg-Hayes, 2006; Scrivener & Weiss, 2009). Another approach has been to put students into "Learning Communities," a group of students who stay together for blocks of classes, where one of these classes is a "college success" course aimed at delivering timely advice to students (Scrivener et al., 2008; Visher, Weiss, Weissman, Rudd, & Wathington, 2012; Weiss, Visher, & Wathington, 2010). Rigorous studies of the Learning Community model have found modest positive impacts on progress through developmental math courses, and these effects tended to fade over time (Visher et al., 2012).

Virtually every college and university has some form of advising system in place, suggesting that there is robust agreement that navigating one's way through higher education can be very challenging. Research supports the idea that students often do not know what courses they should be taking, often fail to recognize that they need help, and do not know what help is available or how to get it, when and if, they decide they need it (Deil-Amen & Rosenbaum, 2003). Many interventions, some of which are described above, attempt to tackle

TABLE 1

Differences Between Program and Control Students and Classes.

	Control group	Program group	*p*-value for difference
Characteristics			
Math 0080 (lowest level remedial math course)	0.157	0.161	0.963
Math 0085 (remedial math)	0.416	0.429	0.904
Math 1414 (lowest level college math course)	0.431	0.415	0.883
Female	0.576	0.576	0.995
Young (≤24 years)	0.711	0.703	0.794
Fraction H.S. GPA[a]	0.756	0.765	0.341
Top 25% on placement test[b]	0.237	0.215	0.639
Bottom 25% on placement test[b]	0.219	0.228	0.773
Missing placement test score[b]	0.088	0.080	0.580
Part-time student	0.481	0.466	0.636
Evening class	0.179	0.166	0.880
Filled out "BIF" survey[c]	0.242	0.249	0.811
Class size[d]	26.14	26.02	0.890
Outcomes			
Ever visit tutoring center	0.189	0.263	0.002
Withdraw	0.176	0.161	0.581
Pass	0.525	0.558	0.469
Register for next semester	0.590	0.583	0.749
Observations	1,098	1,067	

Note. Data are for a randomized-controlled experiment where program students were in classes that were randomly assigned to have a class Volunteer. Data include administrative records and a student baseline survey (see text for details).
[a]Fraction H.S. GPA is individuals' high school GPA as a fraction of the highest GPA available at their high school, due to the fact that 4.0 is not the highest GPA at many schools. This information is only available for 1,464 students.
[b]Top 25%, bottom 25%, and missing placement test score are all derived from information on (in-sample) math placement test scores. Not all the students have placement test scores, and among those who do, not all took the same test. Missing placement test is a dummy equal to 1 if there is no placement test data, and 0 otherwise. Top 25% and bottom 25% are defined within this sample and are equal to 1 if the students' placement test score fell within those percentiles for a given test.
[c]Filled Out "BIF" survey is a dummy variable equal to 1 if the student filled out a Baseline Information Form, and indicates the student enrolled in an individualized random assignment study. Not all students were requested to fill out such a form; see text for details.
[d]Class size is calculated for the 83 different classes in the experiment.

this informational barrier through one-on-one advising, often in combination with other supports, and there is evidence, described above, that these interventions can improve student outcomes. Below, we turn to a description of the Beacon Program at STC.

The Beacon Program at STC

The Beacon Program was designed to reach a large number of students in high-failure math classes at low cost to STC. The College recruited 41 "Volunteers" from among existing employees and gave them a brief training course on academic advising and counseling.[8] The logic model demonstrates how the Beacon Program was intended to work (see Figure 1). The Beacon Volunteers began the semester by arranging with their paired instructor times they could meet with their assigned class 3 or 4 times during the semester for 5 to 10 minutes at the beginning of class. During those minutes, the Volunteers handed out printed information about campus services and important dates to remember such as deadlines for applying for financial aid and registering for the next semester.[9] They also passed out their email addresses and office phone numbers, encouraged students to think of them as their "go-to" person for information on campus, and urged students to contact them outside of class so that they could refer them to appropriate services.[10] Volunteers
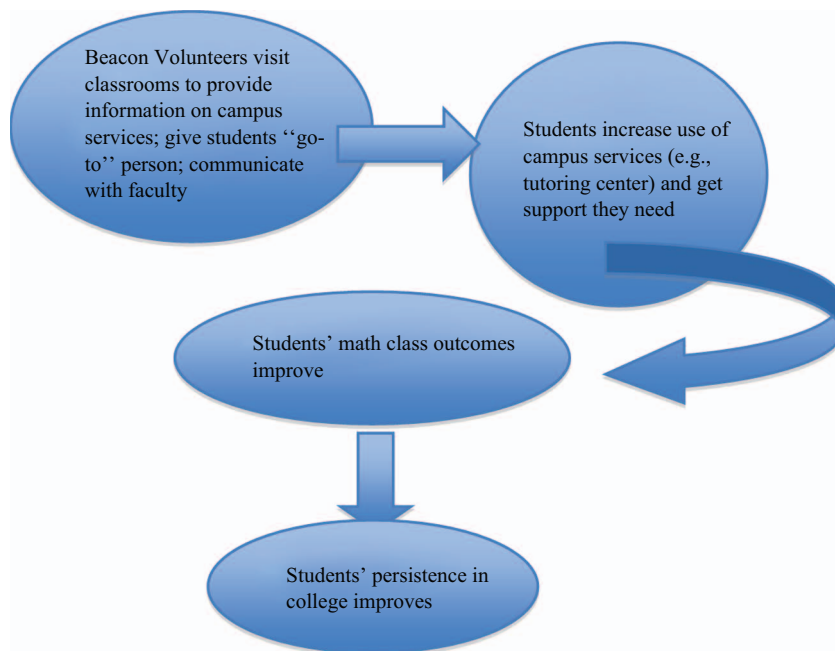
FIGURE 1. *Logic model for Beacon Program.*

and math instructors were also expected to communicate about students who seemed at risk of failure or dropping out. For example, instructors might email Volunteers to ask them to follow up with students who had missed several days of class. This combination of in-class and out-of-class contact was intended to steer students to the individual-specific campus services that they needed to be successful, thus turning into improved math class outcomes. If students improve their remedial and gatekeeper math outcomes, the logic model concludes that the retention rates will also improve.

The Beacon Program can be thought of as having two components. The first is a pure information piece: Volunteers delivered information to the classroom.[11] In addition, the Beacon Program's second component was that it provided students and instructors with a contact person who could refer students to services. The implementation study for the Beacon Program (see Visher, Butcher, & Cerna, 2010, chap. 2) surveyed Volunteers and students to see how the components of the Beacon Program

functioned on the ground. More than 90% of Volunteers and students reported that the Volunteers visited the classrooms to disseminate information. Only about 20% of the students reported meeting with their Volunteer one-on-one, but 50% of students reported some form of communication with their Volunteer outside of class. The Volunteers also indicated that communication with students outside of class, when it occurred, was generally over the phone or via email, with 60% of Volunteers indicating that they had three or more email contacts with students.[12] Like the students, Volunteers reported that one-on-one meetings were infrequent, with only about 10% of Volunteers indicating that they had three or more meetings with students outside of class. Finally, the survey of the Volunteers indicated that about half of them communicated with the instructors about students during the semester.

In sum, the implementation study supports that this was a "light-touch" intervention that was carried out with reasonable fidelity to the program concept. The backbone of the Beacon

Program was that Volunteers came into the classroom to deliver information to the students. When additional interaction between Volunteers and students took place outside the classroom, it was most often via email, and the content of the interaction was to provide referrals to appropriate campus services. Although one-on-one meetings between Volunteers and students were relatively rare, a follow-up survey[13] with students suggested the Beacon Program increased students' sense of connection to the institution: Program students were more likely (61% vs. 56%) than control students to agree with the statement that "at least one person who works at STC cares about how I do in math class."[14]

Data and Methodology

Data and Randomization Methodology

The data used in this study are mostly from administrative records maintained by the College. These data include transcript data, with information on all classes taken during a semester, and registration data for each semester of the study. We have basic demographic information maintained by the College. Finally, we have data from card-swipe entries into the Center for Learning Excellence, so we know which students visited the tutoring center. The data were reorganized from their raw form such that there is one observation per student.

As noted above, the evaluation was conducted using random assignment of whole classes to a treatment group (where classes were assigned a Volunteer) or to a control group (where classes were not assigned a Volunteer). All of the instructors teaching the targeted courses were required by their chairs and deans to participate in the Beacon Program, and all students enrolled in these courses participated. The random assignment methodology used the fact that instructors often teach more than one section of the same course, and "blocked" on faculty, so that half of each instructor's sections randomly received a Volunteer and half did not. As all students and faculty associated with the targeted courses are included in the study, issues of selection bias arising from students or faculty choosing to participate in the Beacon Program are not an issue in this evaluation. Furthermore,

as most faculty members taught more than one section, and half of these sections were in the program group and half in the control group, the two groups should be balanced in terms of faculty quality and the quality of students who elect to take classes with particular faculty.

There may be some concern that the blocking on faculty may have created the possibility of contamination from the program group to the control group, as faculty observed the Volunteers' presentations in the program classes, they might have contaminated the control students by presenting them with the same information. However, the implementation study indicates that many of the faculty wanted more class time for math instruction and needed to be convinced to accept a Volunteer. It is possible that they changed their minds after seeing a presentation and decided to bring that information to their control classes, but they were instructed not to do this and the implementation study suggests it was not an issue. If there was contamination, it would bias the estimates of the program impact toward zero.

The top panel of Table 1 shows descriptive statistics for background information on the 2,165 students in the study. The data contain information on the level of math class,[15] whether the individual is female, young (≤24 years old), a part-time student, and whether the individual is taking his or her class in the evening. Academic preparedness is captured by high school grade point average (GPA),[16] and whether a student's math placement test score was in the top or bottom quartile of the study sample.[17] There is information on whether the individual filled out a "Baseline Information Form" (BIF) survey, which indicates that this individual actively consented to individual random assignment, which will be discussed further in the section "Extensions: Differences in Program Impacts by Enrollee/Nonenrollee Status and by Teacher Quality." Finally, the data contain information on class size.

Table 1 shows the summary statistics separately for students in control classrooms, without a Volunteer, and those in program classrooms, who were assigned a Volunteer, as described above. The last column shows the *p*-value for the difference between the program and control classrooms. Importantly, none of these differences is statistically significant at conventional levels.[18] Whole-class randomization

placed a much smaller administrative burden on the College than individualized random assignment, and yet it created treatment and control groups that are very similar in terms of observable characteristics. This methodology also meant that all students taking these targeted courses and all faculty teaching in them were part of the study sample, so there should be no external validity issues arising from which types of faculty and students choose to actively enroll in a randomized-controlled trial.

The bottom panel of Table 1 gives summary statistics on the main outcomes examined: whether the student ever visited the tutoring center,[19] whether the student withdrew from his or her math class, whether the student passed his or her math class,[20] and whether the student registered for the following semester. Recall that the logic model presented above suggests that the Volunteers will help steer students to appropriate services, including tutoring, financial aid, counseling, and the like. These intermediate services will improve math class outcomes, reducing withdrawals and improving pass rates, which will in turn improve persistence in college. The only intermediate service for which we have administrative data is visits to the tutoring center, and we will investigate the program impact on this outcome. One should keep in mind, however, that the program may have affected the use of other services as well.[21]

The second panel of Table 1 shows the means of the outcome variables. The means for the control group are of interest here. About 18% of the control group students withdrew from their math classes and only 53% passed their math classes. Despite these poor outcomes in math class, only about 19% of control students visited the tutoring center, the on-campus service most targeted toward improving classroom performance. Finally, 59% of the control students registered for the next semester after the study took place. There are some program control differences evident in the mean values for these outcomes; however, our preferred methods for assessing the programs' effects will be explained in the following section.

Regression Methodology

This study is an evaluation of an intervention using a randomized-controlled design. The 83 different math class sections were randomly assigned to a treatment group that received a Volunteer or a control group that did not. The randomization methodology put similar classes into groups or "blocks," and then randomized within those blocks. Blocking helps to ensure that the program and control groups are the same in terms of observable and unobservable characteristics. As teachers' abilities vary and are hard to observe, teachers who taught more than one math class section formed a block, and within that block, each class section had an equal probability of being randomly selected to receive a Volunteer. Furthermore, evening and daytime sections each formed blocks, because students taking day and evening classes may be very different. To understand how randomization within blocks can balance unobservable characteristics, consider an instructor who teaches two classes and is more talented than other instructors. Without randomizing within a block, both of this instructor's sections might be randomly selected into the program group. If this instructor's students perform better than other students, teacher quality may confound the effect of the program. By "blocking" both of this instructor's classes together, and then randomly selecting one to get a Volunteer and one not to, the experiment can ensure that the treatment and control groups are balanced in terms of unobservable characteristics like "teacher quality" that may affect student outcomes. Using a linear regression, we compare outcomes for program and control students' in the same random assignment block. Controlling for random assignment block effects ensures that we are comparing the outcomes across program and control students who made similar choices regarding teachers and whether to take the class in the evening or the daytime.

Regression analysis also allows us to control for student background characteristics. Although as demonstrated in Table 1, these background characteristics are statistically identical for students in the two groups, controlling for background characteristics that are correlated with the academic outcomes of interest will allow for a more precisely estimated program impact.

The 2,165 students in the study received their classroom instruction in 83 different math class sections. Thus, it is possible that there are common elements in the classroom, common

"shocks," affecting students who participated in the same classrooms. These common shocks will affect the statistical inference, thus we report standard errors that are clustered at the class section level and are robust to these correlated shocks.[22]

Finally, this is an "intent-to-treat" analysis of the effect of the Beacon Program. The program effects are presented for all students who were included in the math classes on the day of the registrar's census. There are many changes that might arise from the Beacon Program: Volunteers' activities might change the composition of the class by affecting withdrawals or might have a direct effect on performance. For example, if Volunteers encourage struggling students to remain in the class, then students who remain in the program classes may be less mathematically talented, on average, than control students, leading, ceteris paribus, to worse performance in the program classes. However, the program may, in fact, improve students' outcomes by directing them to on-campus services that support their academic performance. If that is the case, then students receiving the program would perform better in class than control students, all else equal. The intent-to-treat analysis will yield insight into the net impact on math class outcomes of all of these changes potentially induced by the Beacon Program.[23]

In sum, we estimate the effects of the Beacon Program using a regression framework. The sample includes all the 2,165 students enrolled in the program and control classrooms on the day of the registrar's census. Specifications exploring alternate controls for individual characteristics and random assignment block effects are presented next. The standard errors are robust to class section common shocks.

Estimated Program Effects

All Students

Table 2 presents estimates of the impact of the Beacon Program on one of the key outcomes—students' probability of visiting the tutoring center—to investigate the differences in estimated impacts with alternate model specifications. These are linear probability models— the outcome is "1" if the individual visited the tutoring center and "0" otherwise. The top row of the table shows the coefficient on the dummy variable for the program group—coded as "1" if the individual was in a class with a Volunteer and "0" if the individual was in a class without a Volunteer. Pairs of columns present the results first without and then with adjusting the standard errors for potential correlated shocks at the classroom level. Moving across the columns shows what happens to the estimated impact as we control first for observable student characteristics, and then as we control for random assignment block effects in two different ways.

The first column gives the raw difference between the two groups, and is the same as calculated by subtracting the mean outcome for the control group from the mean outcome for the program group in Table 1. The raw difference indicates that the program group is 7.4 percentage points more likely to visit the tutoring center than the control group. The standard error of 0.0179 is the "raw" standard error and does not account for potentially correlated errors among students in the same class sections. The second column presents the results allowing for correlated shocks among students in the same class sections. The standard error increases by about 25% when these potential correlations are taken into account.

The third and fourth columns add controls for individual student characteristics[24] and class characteristics (course level and evening or day-time class). The addition of these controls does not change the estimated program impact, as expected, because Table 1 showed no statistically significant differences between program and control students along these observable dimensions. Comparing the third column with the fourth column shows that allowing for correlated errors increases the standard error of the estimated program impact.

The next set of columns (5 and 6) adds controls for random assignment block by including a dummy variable for each of the 40 random assignment blocks. The estimated impact of the program decreases in magnitude once these fixed effects are added, and the standard error is about the same. The Beacon Program is estimated to have increased the probability of visiting the tutoring center by about 5.8 percentage points, and the estimate is statistically significant at the

TABLE 2

Estimated Program Impact on Probability of Visiting the Tutoring Center.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Program group	0.0739	0.0739	0.0731	0.0731	0.0579	0.0579	0.0564	0.0564
	(0.0179)	(0.0224)	(0.0177)	(0.0200)	(0.0199)	(0.0189)	(0.0195)	(0.0183)
p-value	0.00004	0.00145	0.00004	0.000458	0.00371	0.00298	0.00391	0.00280
Constant	0.189	0.189	0.206	0.206	0.200	0.200	0.201	0.201
	(0.0126)	(0.0162)	(0.0285)	(0.0298)	(0.0498)	(0.0503)	(0.0498)	(0.0497)
Clustered standard errors[1]	No	Yes	No	Yes	No	Yes	No	Yes
Individual and class controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Random assignment block fixed effects	No	No	No	No	Yes	Yes	No	No
Teacher fixed effects	No	No	No	No	No	No	Yes	Yes
Observations	2,165	2,165	2,165	2,165	2,165	2,165	2,165	2,165
R^2	0.008	0.008	0.037	0.037	0.059	0.059	0.057	0.057

Note. See text for details. Estimates are from linear probability models. Individual and class controls include dummy variables for math class level (Math 0080 or 1414) and evening class, controls for being in the bottom or top quartile of the sample in math placement test score, a dummy for having a missing math placement test score, dummy variables for the individual being young (≤24 years old), female, and part-time student. [1]Standard errors are clustered at the level of each class section. Standard errors are shown in parentheses.

TABLE 3

Estimated Program Impacts on Probability of Visiting Tutoring Center, Withdrawal, Passing, and Registering for Subsequent Semester.

Variables	(1) Ever visit tutoring center	(2) Withdraw	(3) Pass	(4) Register for subsequent semester
Program group	0.0579	−0.0329	0.0152	−0.0101
	(0.0189)	(0.0146)	(0.0190)	(0.0200)
p-value	0.00298	0.0264	0.427	0.614
Part-time student	−0.0119	0.0112	0.00293	−0.0729
	(0.0190)	(0.0138)	(0.0219)	(0.0196)
p-value	0.534	0.420	0.894	0.0004
Constant	0.200	0.167	0.531	0.519
	(0.0503)	(0.0246)	(0.0488)	(0.0364)
Observations	2,165	2,165	2,165	2,165
R^2	0.059	0.106	0.171	0.035

Note. See notes to Table 2. The right-hand side variables are the same as in column 6, Table 2, and include individual and class characteristics as well as random assignment block fixed effects. Standard errors are clustered at the class section level. Standard errors are shown in parentheses.

1% level (*p*-value = 0.004) in column 5; in this case, clustering the standard errors at the class-room level increases the precision of the estimate, and in column 6, the standard error is slightly smaller.

The specification in column 6 of Table 2 is the preferred one and will be the basis for most of the subsequent analysis. However, in the section "Extensions: Differences in Program Impacts by Enrollee/Nonenrollee Status and by Teacher Quality," when we investigate whether the program impacts differ by a measure of teacher quality, our methodology requires estimates of teacher fixed effects. The methodology will be described in detail in the above-said section, but it is useful to know whether estimates that control for teacher fixed effects rather than random assignment block fixed effects are very different: Comparing Columns 5 and 6 with Columns 7 and 8 of Table 2 shows that one gets similar estimated program impacts when one controls for teacher fixed effects.[25]

The estimated program impact in column 6, our preferred specification, suggests that the Beacon Program increased the probability that a student visited the tutoring center by about 6 percentage points.[26] Table 1 shows that, on average, 19% of (control) students visited the tutoring center. Thus, the Beacon Program is estimated to have increased the use of this campus service by about 30%.[27] That is a large increase in student visits to the tutoring center, especially given the "light-touch" nature of the intervention. As discussed, the logic model for the program suggests that bringing Volunteers into the classroom to deliver information can guide students to campus services that address their needs. The evidence here indicates that, by the one available measure of campus services usage, this piece of the intervention was successful.

Next, we turn to examining whether guiding students to campus services had the hoped for effect on math class outcomes and college persistence. Table 3 presents results for the probability of ever visiting the tutoring center (for reference), the probability of withdrawing from the math class, the probability of passing the math class, and the probability of registering for the subsequent semester. The specifications are identical to those in column 6 of Table 2, and include random assignment block fixed effects, the full set of class and individual controls, and adjust the standard errors for common shocks within classrooms. Only the coefficients on the program group dummy and on the dummy for part-time status are reported in Table 3, but the full set of coefficients is given in Table A1 of appendix.

Results in column 2 of Table 3 show that the Beacon Program is estimated to have decreased math class withdrawals by 3.3 percentage points, which is statistically significantly different from zero at the 5% level. Given that about 18% of control students withdrew from their math class, this is nearly a 20% decline in the rate of math class withdrawal. Despite this significant impact on math class withdrawal rates, there is no statistically significant effect of the program on pass rates. Column 3 shows that although the point estimate for the program impact on passing is positive, the standard error is large. The 95% confidence interval (CI; not reported) ranges from −0.023 to 0.053.[28] It is worth noting that there is also no statistically significant impact on the probability of receiving a failing grade in the math class (not shown). The roughly 3 percentage point reduction in withdrawals split about evenly into 1.5 percentage points more passing and 1.5 percentage points more failing, but neither of these latter estimates is significantly different from zero. Finally, column 4 shows that there is no statistically significant effect of the program on the probability that students registered for the subsequent semester.

In sum, these results suggest that the Beacon Volunteers successfully guided students to the tutoring center. Students reduced their probability of withdrawing from their math class, either because visiting the tutoring center gave them confidence to stay in the class or because the Volunteers (or other academic advisers to whom Volunteers referred students) communicated the fact that passing the remedial and gatekeeper math courses was a prerequisite for progressing in college. However, there was no statistically significant effect of the Beacon Program on passing (or failing) and no significant effect on college persistence. The implications of these results will be discussed in the final section.

Results by Part-Time/Full-Time Status

One of the main mechanisms through which the Beacon Program may have affected student outcomes was by providing information to students that can help them with their academic work, access financial aid, or obtain advising.

Part-time students may lack such information about these campus resources: Time constraints may have led them to attend school part-time rather than full-time, and these same time constraints may limit their ability to gather information by visiting administrative offices or going to informational meetings. They tend to spend less time on campus due to work or family obligations.[29] Similarly, they may have fewer student peers who can pass information along to them. The literature on part-time students suggests they are particularly at risk for dropping out of college.[30] For these students, a Volunteer who comes into their classroom a few times a semester to "dose" them with timely information, and having a "go-to" person for information, may be especially helpful.[31]

Table 4 shows the estimated impact of the Beacon Program for part-time and full-time students by including a dummy variable for part-time and an interaction term between the program group indicator and the indicator for part-time. The program impacts are estimated for the probability of ever visiting the tutoring center, withdrawal, passing the math class, and registering for the subsequent semester. The specifications are identical to those in Table 3, with the addition of the Part-Time × Program Group interaction term. With this addition, the estimated impact of the program for full-time students is now just the coefficient on the program group indicator, given in the first row of the table. The coefficient on the interaction term shows the difference in the estimated program effect for part-time students from full-time students. The bolded entries show the calculated impact of the program for part-time students (calculated by summing the coefficients for the program group and the interaction term) and the *p*-value for the null hypothesis that the program effect is zero for part-time students.

Column 1 of Table 4 shows the estimated program effects for the probability of visiting the tutoring center. The interaction term between the program group and part-time is not statistically different from zero, indicating that part-time and full-time students increased their probability of visiting the tutoring center by about 5.5 to 6 percentage points. The results for math class outcomes, however, are statistically different for part-time and full-time students. The

TABLE 4

Estimated Program Impacts for Full-time and Part-time Students on Probability of Visiting Tutoring Center, Withdrawal, Passing, and Registering for Subsequent Semester.

Variables	(1) Ever visit tutoring center	(2) Withdraw	(3) Pass	(4) Register for subsequent semester
Program group	0.0603	−0.0107	−0.0129	−0.0141
	(0.0271)	(0.0178)	(0.0277)	(0.0249)
p-value	0.0287	0.548	0.642	0.571
Program Group × Part-time	−0.00543	−0.0496	0.0627	0.00890
Student	(0.0373)	(0.0268)	(0.0415)	(0.0386)
p-value for interaction term	0.884	0.0680	0.135	0.818
Program group coefficient + interaction term coefficient	0.0549	−0.0603	0.0498	−0.0052
***p*-value for program group + interaction term = 0**	0.0355	0.0067	0.0860	0.8658
Part-time student	−0.00927	0.0353	−0.0275	−0.0772
	(0.0254)	(0.0186)	(0.0271)	(0.0276)
p-value	0.716	0.0610	0.314	0.00643
Constant	0.199	0.155	0.546	0.521
	(0.0505)	(0.0251)	(0.0478)	(0.0376)
Observations	2,165	2,165	2,165	2,165
R^2	0.059	0.107	0.172	0.035

Note. See notes to Tables 2 and 3. The specifications are the same as in Table 3, with the addition of an interaction term between the dummy variables for part-time student status and the program group. Regressions include individual and class characteristics as well as random assignment block fixed effects. Standard errors are clustered at the classroom level. Standard errors are shown in parentheses.

calculated impact of the program on math class withdrawal for part-time students is about a 6 percentage point decrease, and this estimate is statistically different from zero and statistically different from the estimated impact for full-time students. In addition, for part-time students, the decrease in math class withdrawals translated into an increase in the pass rate.[32] Part-time program students were about 5 percentage points more likely to pass than part-time control students, and this estimate is statistically significantly different from zero at the 10% level.[33]

Part-time students, who comprise nearly half the sample and who ex ante one might suppose would be particularly in need of information, or in need of connection with someone on campus, appear to be the students whose math class performance benefited from the Beacon Program. The estimated impact on math class pass rates for part-time students represents a nearly 10% increase in math class pass rates.[34] To put the estimated 5 percentage point increase in pass rates in perspective, it is about 70% of the size of the estimated coefficient of being in the top 25% of the math placement test distribution relative to being in the middle of the distribution; for part-time students, the Beacon Program had about the same estimated impact as a substantial increase in math preparedness.

Extensions: Differences in Program Impacts by Enrollee/Nonenrollee Status and by Teacher Quality

In this section, we will discuss extensions to the main results. These include estimates of whether the Beacon Program's impacts are heterogeneous by whether a student is the "type" who actively enrolls in random assignment studies and by an available measure of teacher quality. These results are somewhat speculative, but we think they are interesting and bring up issues that have important implications for other research or highlight important avenues for future research.

A persistent question in the random assignment evaluation literature is whether there may be differences in treatment effects for those people who volunteer to participate in experimental evaluations, thus, potentially, affecting the external validity of the experiment (see Shadish, Cook, & Campbell, 2002, for a thorough discussion of external validity in research design). If individuals who choose to participate in studies are more motivated than the general populace, for example, are they also better able to take advantage of the treatment offered in a given intervention? If they are, then it is possible that the experiment lacks external validity, and that the treatment effects would be much different if the program was extended to a broader population, rendering the experiment's results less relevant for policymakers who must consider what would happen if they extended the program to all individuals.

The original design for this study was to randomize individuals into treatment and control groups rather than randomizing class sections. Several weeks after registration was opened, the decision was made to randomize whole classes rather than individuals, in part to reduce burden on the college. Thus, students who registered before this change in design were asked whether they would consent to the random assignment procedure. Those students who agreed to random assignment[35] were asked to sign an informed consent form to comply with human subject rules and to complete a "BIF."[36] We use the information on whether the individual filled out a BIF to indicate that this group of students, 532 out of the 2,165 in the study sample, actively agreed to participate in an experiment, and we call them "Enrollees." In all, 75% of study students registered after the decision was made to switch to whole-class random assignment. These students did not go through enrollment for an individual random assignment study and did not explicitly volunteer for such a study. Among these "Nonenrollees," some would have likely declined and some would have likely agreed to individualized random assignment. Unfortunately, we cannot compare the program impacts for students who we know enlisted in a random assignment study with impacts for those who we know explicitly declined, but we can compare the effects for those who actively enlisted with the rest of the students. We investigate whether the program effect differs for the Enrollee subgroup, indicated by whether the student filled out a BIF survey. Importantly, Table 1 shows that there is no statistical difference in the likelihood of filling out a BIF survey for students in the program and control groups.

Table A2 of the appendix presents differences in characteristics for those students who volunteered for random assignment, "Enrollees," and those students who were never asked to volunteer for the study, "Nonenrollees." For most variables, Enrollees are not statistically different from Nonenrollees. However, the Enrollees were less likely to be in the lowest level developmental math class, were more likely to be female, and were more likely to be older than 24 years old. It is important to keep in mind that these are also individuals who registered for classes earlier, before the switch was made to whole-class random assignment and that is why they were given the opportunity to accept or decline individualized random assignment. Older people, women, and those who are more academically prepared may be those who are also more likely to register early, and they may be the types who are more likely to volunteer for random assignment studies. We believe analyzing differences in program impacts by Enrollee and Nonenrollee status can tell us something about the external validity of this intervention, however, as Table A2 of the appendix indicates, being an active consenter or not is not the only difference between these groups. At a minimum, Enrollees (who registered earlier, are more female, older, and are in higher level math classes) comprise an interesting subgroup for which to investigate whether there are heterogeneous effects of the intervention. Potentially, this analysis goes further and gives insight into whether there are heterogeneous effects of the program for those types of people who volunteer for experiments compared with the population of math students at large.

TABLE 5

Estimated Impact of Program by "Enrollee" Status on Probability of Visiting Tutoring Center, Withdrawal, Passing, and Registering for Subsequent Semester.

Variables	(1) Ever visit tutoring center	(2) Withdraw	(3) Pass	(4) Register for subsequent semester
Program group	0.0566	−0.0386	0.0133	0.00835
	(0.0204)	(0.0154)	(0.0200)	(0.0261)
p-value	0.007	0.0139	0.507	0.750
Enrollee × Program group	0.00686	0.0208	0.00889	−0.0697
	(0.0488)	(0.0291)	(0.0454)	(0.0504)
p-value for interaction term	0.889	0.478	0.845	0.171
Program group coefficient + interaction term coefficient	0.0635	−0.0178	0.0222	−0.0614
***p*-value for program + interaction = 0**	0.1561	0.5213	0.5980	0.1074
Enrollee	0.0334	−0.0371	0.0316	0.0785
	(0.0303)	(0.0211)	(0.0318)	(0.0386)
p-value	0.275	0.0819	0.324	0.0453
Constant	0.189	0.177	0.520	0.500
	(0.0496)	(0.0253)	(0.0492)	(0.0385)
Observations	2,165	2,165	2,165	2,165
R^2	0.060	0.107	0.172	0.037

Note. See notes to Tables 2 and 3. The specifications are the same as in Table 3, with the addition of an interaction term between the dummy variables for "Enrollee" status and the program group. Regressions include individual and class characteristics as well as random assignment block fixed effects. Standard errors are clustered at the class section level. Standard errors are shown in parentheses.

Similar to the subgroup estimates for part-time students described above, we investigate whether there is evidence supporting heterogeneous treatment effects for the Beacon Program by experiment enrollment status by including an indicator for "Enrollees" and an interaction term between this variable and the indicator for being in the program group. The specification used to estimate impacts of the program for the Enrollees and Nonenrollees is the same as that in Table 3, with the addition of the Enrollee indicator and interaction term. The results are presented in Table 5.

The coefficient on the Enrollee status variable indicates that control group Enrollees are less likely to withdraw from math class and are more likely to persist into the next semester in college than Nonenrollees, and these differences are statistically significant at the 10% and 5% levels, respectively. To the extent that there are statistically significant differences in outcomes between these two groups, the types of students who actively sign up to take part in experiments

tend to have better outcomes. It would give some concern for the results of individualized random assignment studies, then, if any beneficial effects of the program were accruing solely to the students who are the types to actively participate in experiments. The main thing to note is that none of the Enrollee × Program Group interaction terms is statistically different from zero at conventional levels of significance. In addition, at least in terms of the point estimate on withdrawal rates, the Beacon Program appears to have reduced math class withdrawal more for Nonenrollees than for Enrollees (although again this difference is not statistically significant). These estimates do, at least, suggest that the program's impact on withdrawal rates was not solely driven by the Enrollees. For this intervention, the evidence is suggestive that an experiment based only on Volunteers for a random assignment study would underestimate the efficacy of the program in reducing withdrawal rates.

Program Impacts by Teacher Quality

As mentioned above, this study is rare in that it used whole-class random assignment, and all the instructors of the targeted math classes were required to participate in the study. Most studies that use randomization rely on instructors, as well as students, to voluntarily enroll in the study and then randomize among those who are interested participants. As with students, this fact may cause concern for the external validity of the results because even if a program is found to have an impact, perhaps it is just that these are the better instructors who know how to take advantage of any innovation. Perhaps the program would have much different effects if prescribed for all instructors. Because of the design of the evaluation of the Beacon Program, we have an opportunity to investigate whether the program's impact is different for different types of teachers. Unfortunately, unlike with individual students in this study, instructors were never asked to actively consent to the program, so we have to investigate heterogeneity along a different dimension: teacher "quality."

Investigating whether the Beacon Program has heterogeneous impacts by teacher quality is important for two reasons: First, as mentioned above, one worries that high-quality teachers are more likely to participate in (other) studies and are those who generate any positive effects of the intervention; and second, students in classes taught by low-quality teachers comprise a particularly interesting subgroup for the Beacon Program. Program administrators hoped that steering students to appropriate services, like the tutoring center, might help compensate for having a lower quality instructor. Alternatively, better teachers may be better precisely because they are able to make better use of available resources, and the Beacon Volunteers may be just one more such resource that they understand how to deploy to maximum benefit, and thus the program may increase differences in student outcomes between high- and low-quality teachers.

Of course, to investigate heterogeneous program effects by teacher quality, we have to attempt to measure teacher quality. Studies of the effect of teachers on student outcomes are much more common in the K–12 literature than in higher education, and we borrow from that literature here. In the K–12 literature, it is common to estimate teacher fixed effects in value added models with standardized test scores as the outcome and use those coefficients as a measure of teacher quality (see Aaronson, Barrow, & Sanders, 2007). We can do something similar: Students in developmental math (0080 and 0085) took a common final exam, thus we can use final exam score as the outcome variable and estimate the "effect" each teacher (in control group classrooms) has on that test score as a measure of teacher "quality."

It is important to recognize that whether teacher effects on test scores capture what we intend by "teacher quality" is a matter of controversy. As students are not typically randomly assigned to teachers, in this study or in most others, whether the teacher "fixed effects" capture something about teacher quality or about the match quality with students or student sorting[37] is often unclear.[38] Nonetheless, we use this measure as it is frequently used in the K–12 literature and is a logical starting point for an empirical measure of teacher quality in a higher education context.

To estimate teacher effects on the final exam score, we use the sample of 325 students (with valid final exam scores) in control group classrooms, taught by teachers who also had a class in the program group.[39] Table A3 of the appendix presents the regression coefficients. These regressions control for class and individual characteristics.[40] Twelve dummy variables are included to capture the fixed teacher effects for the 13 teachers in this subsample. All of the coefficients on the teacher dummy variables are negative, indicating that Teacher 1, the omitted category, had students who scored the highest on the final exam (controlling for observable characteristics). Note that these teacher fixed effects are jointly statistically significant at (at least) the 1% level, and that the estimated effects are large: The difference between the best teacher (the omitted category) and the worst teacher (Number 12) is 32 points on the final exam, and that is a bigger effect than moving from the bottom quartile of the math placement test to the top quartile. Again, these cannot be considered causal estimates of the teacher effects, but, as in the K–12 literature, teacher fixed effects are highly correlated with student outcomes.

Whether these teacher fixed effects truly measure teacher "quality" or not, we can nonetheless use them to investigate whether the Beacon Program has a differential effect for teachers whose control students score better on the final exam. We use the coefficients on the teacher fixed effects to rank teachers. Teacher Number 1, the teacher with the highest scoring students, is coded as having "0" for teacher quality. The other teachers are coded as having the coefficient on their indicator variable as their "teacher quality" measure (all negative). We then divide teachers into "top half" and "bottom half" of the "teacher quality" distribution. We estimate equations with the program group dummy, a "Top 50% Teacher Quality" dummy × Program Group interaction term, and individual and class characteristics on the right-hand side. Teacher fixed effects are included and these will absorb the main effect on outcomes of a teacher who is in the top/bottom half of the quality distribution.[41]

The interaction term between "Top 50% Teacher Quality" and the program group indicates whether the program effect was different for students in classes with the teachers whose control group students did the best on the common final exam. Note that we can only estimate these results for the students who were taught by these 13 teachers who had classes in the program and control groups, so the sample size falls to 1,045.

The results are reported in Table 6. The *p*-value for the interaction term between "Top 50% Teacher Quality" and the program group indicates the interaction term is not statistically different from zero for four out of the five dependent variables.[42] The interaction term is statistically significant at the 10% level for registering for a subsequent semester, and indicates that the program effect for students with higher quality teachers was lower than the program effect for students with lower quality teachers, although the overall program effect is not statistically different from zero for either group.

Although the estimated program impacts lack precision when divided by teacher quality, the point estimates for most outcomes are opposite signed for the high- and low-quality teachers. Again, although these results are inconclusive, the point estimates suggest that any beneficial

results of the Beacon Program were not solely driven by high-quality teachers.

One should not push the implications of the results in this section for other studies too far, because the estimates here are imprecise and because the context of any given intervention and study will be different and likely matter a great deal. Nonetheless, if it were only those students who actively enlisted in the study or those students who were taught by better teachers who drove any estimated benefits from the intervention, one might worry more about evaluations where one only gets to see results for teachers and students who actively enroll in random assignment studies.

Discussion and Avenues for Future Research

This evaluation of the Beacon Program at STC has several important findings and implications. The Beacon Program was designed to address a problem that faces the majority of U.S. community colleges: high-failure rates in remedial and gatekeeper math courses. The program was designed to be low cost and reach a large number of students by recruiting Volunteers from among existing employees to go to math classrooms and deliver information about available services on campus; these presentations lasted 5 to 10 minutes and occurred a few times during the semester. Students also had the opportunity to follow up with Volunteers. It was hoped that by proactively providing students with information about available services, like the tutoring center, their math class performance would improve and this would yield improved college persistence.

In addition to the program itself being low cost, the evaluation methodology used whole-class random assignment, rather than individual-level random assignment, to program and control groups. This significantly reduced the administrative burden of the evaluation design and did not interfere with normal registration. The results indicate that whole-class random assignment delivered program and control groups that were balanced in terms of observable characteristics. In addition, all students taking the targeted courses and all instructors teaching these courses were part of the study, so selection into the random assignment study is

TABLE 6

Estimated Program Impact Interacted With Teacher Quality Measure.

Variables	(1) Final exam score	(2) Ever visit tutoring center	(3) Withdraw	(4) Pass	(5) Registered for subsequent semester
Program group	2.751	0.0569	−0.0412	0.0469	0.0534
	(2.328)	(0.0295)	(0.0243)	(0.0386)	(0.0383)
p-value	0.244	0.0609	0.0982	0.232	0.168
Program group × Top 50% Teacher Quality	−2.463	−0.0270	−0.0311	−0.0151	−0.0916
	(2.889)	(0.0402)	(0.0379)	(0.0530)	(0.0544)
p-value for interaction term	0.399	0.505	0.418	0.777	0.0992
p-value for Program Coefficient + Interaction Term = 0	0.8406	0.2817	0.0160	0.3835	0.2837
p-value for null hypothesis that teacher effects are jointly zero	0.0000	0.1188	0.0009	0.0000	0.0550
Teacher indicators					
Teacher Indicator 2	−25.46	0.0527	−0.0204	0.310	0.0110
p-value	0.000	0.264	0.514	0.000	0.0586
Teacher Indicator 3	−21.87	0.0214	0.0274	−0.266	−0.0296
p-value	0.000	0.709	0.430	0.000	0.550
Teacher Indicator 4	−21.77	0.142	0.0170	0.135	−0.0093
p-value	0.000	0.0308	0.809	0.0783	0.858
Teacher Indicator 5	−11.01	0.0120	0.0994	−0.0622	0.0618
p-value	0.000	0.817	0.0244	0.391	0.149
Teacher Indicator 6	−5.967	0.0653	0.0399	−0.117	0.0137
p-value	0.450	0.202	0.519	0.308	0.102
Teacher Indicator 7	−6.485	0.0746	0.0417	−0.0747	−0.0109
p-value	0.166	0.263	0.381	0.286	0.884
Teacher Indicator 8	−10.34	−0.0390	0.0789	−0.117	0.0303
p-value	0.0003	0.456	0.101	0.0251	0.674
Teacher Indicator 9	−8.057	−0.0301	−0.0929	0.0007	0.0449
p-value	0.0262	0.520	0.0330	0.988	0.426
Teacher Indicator 10	−10.94	−0.0124	−0.0643	0.0192	0.0270
p-value	0.000	0.841	0.0276	0.783	0.568
Teacher Indicator 11	−6.911	−0.0069	−0.0877	0.0523	0.0242
p-value	0.0064	0.931	0.0019	0.265	0.556
Teacher Indicator 12	−30.48	0.0425	−0.0514	−0.0169	−0.0337
p-value	0.000	0.454	0.215	0.849	0.493
Teacher Indicator 13	−23.95	0.0536	−0.0661	0.126	−0.0061
p-value	0.000	0.469	0.267	0.105	0.928
Observations	672	1,045	1,045	1,045	1,045
R^2	0.231	0.040	0.061	0.155	0.029

Note. See notes to Table 2 and text for details. Specifications are the same as Table 2, column 8, with the addition of the interaction term between "Top 50% Teacher Quality" and program group indicator. Final exam score (column 1) is only estimated for the 672 students who stayed in the class and had valid final exam scores. The other columns are "intent-to-treat" estimates. Regressions include a constant. Standard errors are clustered at the level of each class section. Standard errors are shown in parentheses.

not an issue and bolsters the external validity of the study.

The study finds that this modest intervention was very successful at affecting the first link in the chain of effects the administration hoped for: Program students were about 30% (6 per- centage points) more likely to visit the tutoring center than control students. For students over- all, however, this did not translate into a statisti- cally significant increase in math class pass rates. This suggests that the activities in the tutoring center may not have been sufficiently

well designed to increase student success in math; however, this study does not have information on what took place in the tutoring center. Understanding the activities that take place in tutoring centers, a nearly ubiquitous feature of institutions of higher education, and their impacts on academic outcomes is an important goal for future research.

The Beacon Program also significantly decreased withdrawal rates for program students compared with control students by about 20%. Some of these students who were induced to stay in the course by the Beacon Program went on to pass and some went on to fail with neither of these outcomes rising to the level of statistical significance. How should one view a reduction in the withdrawal rate without an increase in the pass rate? Some college leaders viewed completing the math class, even with a failing grade, as beneficial to students. They argued that many students take the class multiple times to pass and that getting students to stay in the course and learn as much as possible the first time through should lead to an improvement in their pass rates the next time around. Investigating whether that claim is true is outside the scope of this study, but if true, it suggests the reduction in withdrawal rates, even without a significant increase in pass rates in the study semester, was a good outcome.

The Beacon Program did substantially improve math class withdrawal and passing rates for part-time students, a group that comprises almost half the students in the study and a group that is considered at higher risk of not earning a credential while enrolled at college. The program was designed to deliver important information about the college to students in their classrooms rather than relying on students to find the information they need. Part-time students are likely to spend less time on campus and likely to be more time constrained than full-time students. It makes sense that an intervention that went into the classroom to deliver information would have a larger effect for this important group of students. Bringing information about services into the classrooms, where the students can be found, may be a simple way to ensure that students are informed about college services, and this may be particularly important for students who face more time constraints.

In sum, it is important to note that the Beacon Program was an innovative homegrown intervention that brought information directly to students in the classroom rather than waiting for them to seek it. It was designed to cost STC very little by harnessing the commitment, creativity, and knowledge of faculty, administrators, and staff already on campus. It was a "light-touch" intervention that increased visits to the tutoring center by about 30%, reduced withdrawal rates from math class for all students by about 20%, and increased the pass rates for part-time students by 10%.

Appendix

TABLE A1

Estimated Program Impacts on Probability of Visiting Tutoring Center, Withdrawal, Passing, and Registering for Subsequent Semester.

Variables	(1) Ever visit tutoring center	(2) Withdraw	(3) Pass	(4) Register for subsequent semester
Program group	0.0579	−0.0329	0.0152	−0.0101
	(0.0189)	(0.0146)	(0.0190)	(0.0200)
p-value	0.00298	0.0264	0.427	0.614
Math 0080	−0.0988	−0.0828	0.0254	−0.0895
	(0.0253)	(0.0232)	(0.0473)	(0.0407)
p-value	0.000193	0.000603	0.593	0.0307
Math 1414	0.414	0.176	0.0903	0.531
	(0.0887)	(0.105)	(0.375)	(0.0677)
p-value	*0.000*	*0.0978*	*0.810*	*0.000*

(continued)

TABLE A1 (CONTINUED)

Variables	(1) Ever visit tutoring center	(2) Withdraw	(3) Pass	(4) Register for subsequent semester
Evening class	0.197	−0.0308	−0.161	0.0527
	(0.0422)	(0.0330)	(0.0586)	(0.0525)
p-value	0.000	0.353	.00750	0.318
Top 25% on	−0.0236	−0.0865	0.0699	−0.0511
placement exam	(0.0251)	(0.0190)	(0.0265)	(0.0308)
p-value	0.349	0.000	.00989	0.101
Bottom 25% on	0.0469	0.0567	−0.112	−0.0115
placement exam	(0.0271)	(0.0200)	(0.0248)	(0.0252)
p-value	0.0867	.00585	0.000	0.648
Missing placement	−0.0139	−0.0602	0.0938	−0.00215
exam	(0.0275)	(0.0267)	(0.0370)	(0.0407)
p-value	0.614	0.0266	0.0132	0.958
Young (≤24 years)	−0.0989	0.0141	−0.119	0.0160
	(0.0245)	(0.0186)	(0.0249)	(0.0255)
p-value	.000122	0.452	0.000	0.531
Female	0.0383	−0.0406	0.0737	0.0653
	(0.0174)	(0.0157)	(0.0222)	(0.0200)
p-value	0.0307	0.0114	0.00133	0.00163
Part-time	−0.0119	0.0112	0.00293	−0.0729
	(0.0190)	(0.0138)	(0.0219)	(0.0196)
p-value	0.534	0.420	0.894	0.000363
Constant	0.200	0.167	0.531	0.519
	(0.0503)	(0.0246)	(0.0488)	(0.0364)
p-value	.000151	0.000	0	0
Random assignment block fixed effects	Yes	Yes	Yes	Yes
Observations	2,165	2,165	2,165	2,165
R^2	0.059	0.106	0.171	0.035

Note. These results are estimated with random assignment block fixed effects and the standard errors are clustered on the class section (see column 6, Table 2). These results show the coefficients for the other regressors (except the random assignment block fixed effects) that underlie the program group estimates presented in Table 3 (see text for details). Standard errors are shown in parentheses.

TABLE A2
Differences Between Students Who Actively Enrolled in a Random Assignment Study (Enrollees) and Those Who Did Not (Nonenrollees).

Variables	Enrollees	Nonenrollees	*p*-value for difference
Math 0080 (lowest level remedial math course)	0.117	0.172	0.029
Math 0085 (remedial math)	0.430	0.420	0.787
Math 1414 (lowest level college math course)	0.460	0.410	0.210
Female	0.660	0.548	0.000
Young (≤24 years)	0.626	0.734	0.000
Top 25% on placement test[a]	0.231	0.224	0.754
Bottom 25% on placement test[a]	0.235	0.220	0.511
Missing placement test score[a]	0.087	0.083	0.830
Part-time student	0.443	0.483	0.162

TABLE A2 (CONTINUED)

Variables	Enrollees	Nonenrollees	p-value for difference
Evening class	0.139	0.183	0.210
Program group	0.501	0.491	0.811
Class size	25.25	26.35	0.272
Observations	532	1,633	

Note. Students who actively volunteered for individual random assignment filled out a Baseline Information Form (BIF). The project eventually used class-level random assignment for which consent was not needed (see text for details). This table shows the differences in observable characteristics between those who actively enrolled in the random assignment study and those who did not.
[a]Top 25%, bottom 25%, and missing placement test score are all derived from information on (in-sample) math placement test scores. Not all the students have placement test scores, and among those who do, not all took the same test. Missing placement test is a dummy equal to 1 if there is no placement test data, and 0 otherwise. Top 25% and bottom 25% are defined within this sample and are equal to 1 if the students' placement test score fell within those percentiles for a given test.

TABLE A3
Teacher Fixed Effects for Remedial Math Final Exam Score.

	(1)
	Final exam score
Math 0080	5.654
	(1.416)
Evening class	10.29
	(1.855)
Top 25% on placement test[a]	1.465
	(4.473)
Bottom 25% on placement test[a]	−11.45
	(2.783)
Young (≤24 years)	−12.36
	(2.392)
Female	4.439
	(2.699)
Part-time student	−2.465
	(1.736)
Teacher Indicator 2	−27.36
	(2.195)
Teacher Indicator 3	−24.81
	(3.137)
Teacher Indicator 4	−24.91
	(2.480)
Teacher Indicator 5	−15.85
	(2.077)
Teacher Indicator 6	−20.79
	(2.074)
Teacher Indicator 7	−9.063
	(3.473)
Teacher Indicator 8	−14.67
	(2.761)
Teacher Indicator 9	−7.918
	(5.405)
Teacher Indicator 10	−16.33
	(2.605)
Teacher Indicator 11	−6.174
	(2.052)
Teacher Indicator 12	−32.09
	(2.236)
Teacher Indicator 13	−20.87
	(2.339)
p-value for null hypothesis that teacher effects are jointly zero	0.000
Constant	74.24
	(3.31)
Observations	325
R^2	0.269

Note. The model is estimated for the 325 Math 0080/0085 students with valid scores who were in classes in the control group that were taught by an instructor who also taught a class in the program group. A dummy variable indicating placement test missing is also included in the regression. Standard errors are clustered on the math class section. Standard errors are shown in parentheses.
[a]Top 25%, bottom 25%, and missing placement test score are all derived from information on (in-sample) math placement test scores. Not all the students have placement test scores, and among those who do, not all took the same test. Missing placement test is a dummy equal to 1 if there is no placement test data, and 0 otherwise. Top 25% and bottom 25% are defined within this sample and are equal to 1 if the students' placement test score fell within those percentiles for a given test.

to the staff and students of South Texas College, in particular Dr. Luzelma Canales. All errors are our own.

Notes

1. South Texas College (STC) referred to the volunteers as "Mentors" or "Beacon Mentors"; to emphasize the difference in this program from traditional mentoring programs, we will call the individuals who presented information to the classrooms "Volunteers" or "Beacon Volunteers."

2. In particular, STC's student body has a much higher fraction Hispanic than is the average at U.S. community colleges. It is worth noting that the Hispanic population in the United States is growing; discovering which programs have promise and which do not for improving higher education outcomes for this important group is worthwhile, even if those findings do not necessarily generalize to all other U.S. subgroups.

3. It is common for single-site experiments to result in publications that then spur further replication and study. For example, there are now a large number of replication studies underway of "performance-based scholarships" or other programs that use monetary incentives to improve students' performance. Early studies in this area examined experimental results from one location and resulted in important publications. For example, Angrist, Lang, and Oreopoulos (2009) use data from a field experiment at a single institution in Canada. Barrow, Brock, Richburg-Hayes, and Rouse (in press) examine outcomes from an experiment at three sites, but all were in New Orleans, Louisiana. One might justly have asked whether the results from these experiments fielded at particular sites would yield results that would generalize to other locations and subgroups. These early publications have generated a rich, multisite research agenda designed to answer precisely that question (see, for example, Patel & Richburg-Hayes, 2012).

4. Enrollments climbed steadily in the years leading up to the intervention and evaluation (Source: U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System [IPEDS]).

5. Aside from the percentage of Hispanic, STC students are similar to those at other community colleges in terms of percentage of male and percentage of students attending full-time. Student retention and graduate rates are also similar to national averages.

6. Nationally, about 60% of an incoming cohort is referred to developmental education, and the majority of those are referred to at least one math class (Adelman, 2004; Attewell, Lavin, Domina, & Levey, 2006; Bailey, Jeong, & Cho, 2009; Perry, Bahr, Rosin, & Woodward, 2010). According to the Achieving the Dream institutional database, about 85% of incoming students at STC in 2007 were referred to developmental math.

7. According to a recent study of 57 community colleges, only one third of students whose scores placed them in the lowest level of math completed the developmental sequence within 3 years (Bailey et al., 2009). In the face of statistics like that, there is a lively policy debate around whether developmental courses provide needed academic remediation for students or erect additional barriers without improving their academic outcomes. Research to date on this question is mixed. For example, Martorell and McFarlin (2011) use the regression discontinuity around the pass threshold on placement exams to compare later academic and labor market outcomes for those students who just barely fail and those who just barely pass. These results suggest that remedial education has little beneficial impact on students' outcomes. Bettinger and Terry Long (2009), however, using a different identification strategy that relies on differences in placement policies and proximity to different institutions, find students in remediation are more likely to persist in college than similar students who were not remediated. These differences in results may not be as much in conflict as appear at first glance, and may be due to the fact that the empirical strategies allow for different comparisons. The first paper only allows for comparisons between those on the margin of remediation for whom remedial education may pose a barrier but not much help. The second paper employs an empirical strategy that may allow for comparisons between those students with lower skills, where some received remedial assistance and some did not. Both of these comparisons are interesting and important, but each informs a different dimension of the policy question. Other researchers and policymakers have taken a different tact, arguing for a redesign

of the developmental education curriculum to try to make it more engaging and more aligned with desired core competencies (Adams, 2003; Adams, Miller, & Roberts, 2009; Bryk & Treisman, 2010; Jenkins, Speroni, Belfield, Shanna, & Edgecombe, 2010).

8. They also received a few additional hours of follow-up training once the program got underway.

9. It is worth noting that the novelty of having someone besides the instructor take control of the class periodically may have its own independent effect on student outcomes. The experimental design does not allow us to distinguish whether it was the type of information delivered or the presence of the messenger that had an effect.

10. Near the end of the semester, the Volunteers informed students about Priority Registration—early registration available to a targeted group of students, including those in the Beacon Program—and encouraged them to begin the process of registration for the following semester with the help of an academic adviser.

11. There is a large literature, across diverse settings, indicating that interventions that solely provide information can change behavior. For example, Daponte, Osborne, and Taylor (1999) test whether providing information about food stamp eligibility can affect take up of that program, and Bertrand and Adair (2011) test whether illuminating loan terms can affect the probability of individuals using high-cost "payday" loans. Both find evidence that information alone can change behavior.

12. It is possible that when students contacted Volunteers outside of class, the conversation and advice may have strayed from a discussion of campus services. However, it is unlikely that the Volunteers were providing additional mathematics instruction as this was neither their background nor included in their brief training. The training included the basics of academic advising and information on available campus services.

13. Note that these results should not be thought of as an "impact of the experiment" because not all students in the experiment responded to the follow-up survey. In the results section, we focus on outcomes from administrative records, which allows us to know outcomes for (virtually) all students who were initially assigned to program and control classrooms.

14. Research suggests that such perceptions may be particularly important for success in higher education among Hispanic students (see, for example, Bordes, Arredondo, Robinson, & Rund, 2011; Hurtado & Ponjuan, 2005; Laden, 1999; Sedlacek, 2010; Sedlacek, Benjamin, Schlosser, & Sheu, 2007). The fact that the STC has mostly Hispanic students may make the finding that the program conveyed a sense of connection to the students particularly important for the outcomes. Given that Hispanic students are a growing group and are disproportionately represented among those students referred to developmental education, it is critical to investigate programs that may improve their higher education outcomes.

15. Three levels of math courses were included in the study—Math 0080, the lowest level of developmental math; Math 0085, the next level of developmental math; and Math 1414, the first level of college math.

16. High school grade point average (GPA) is expressed as a fraction of a student's high school's maximum GPA, because not all high schools have a maximum GPA of 4.0. This information is only available for 1,464 students. It is not used later in the study due to the high rate of missing.

17. There were two math placement tests that the students could have taken and each is scored differently. To use the information consistently for all students, we created an indicator variable for whether the students scored in the top or bottom quarter of the study sample. Placement test scores are missing for about 8% of the sample, and we created a "missing" indicator that is equal to 1 for these students (and 0 otherwise).

18. The fact that class sizes are the same across the program and control groups bears special mention. At the time of registration, students were not told whether they were in a program or a control classroom but could have learned this before the final "add/drop" deadline, which occurred a week after classes began. Volunteers made their first visit to their assigned classroom prior to the final add/drop date. The class "census," identifying which students were in which (program or control) classrooms, was generated after the last add/drop date and after the initial Volunteer visit. Thus, if the Volunteers' visits induced switching out of program classes prior to the census date, that would not be observable in the data. We do not think this was a problem in practice, however, because if switching out of Volunteer classes was common, one would expect the program and control math classes to be different sizes. Class size was calculated for the 83 different classes, and the average class size was 26.14 students for the control group and 26.02 students for the program group. The *p*-value for this difference is .890, indicating that the two classes are statistically the same in size. If exactly the same number of students preferred no Volunteer as preferred a Volunteer, and students were able to switch their schedules according to their preferences, then we might still end up not only with the same class size but also with unobservable differences in preferences between the control and program classes. If those unobservable preference differences were correlated with math abilities, for example, then

that would pose a problem for the validity of the experimental results. Qualitative data collected as the experiment progressed indicated no evidence for switching away from or toward a Volunteer (see Visher, Butcher, & Cerna, 2010, for full description of the qualitative findings).

19. Note that data on visiting the tutoring center also come from administrative records. Students had to swipe a card when they entered the center and their visit was recorded. We also have information on the number of visits. In theory, one could have information on the length of the visit, but many students neglected to log out and thus the duration of visit is dramatically overestimated.

20. Passing means receiving a "D" grade or above in the class. Withdrawing counts as "not passing," so the same students are included in the "passing" and "withdrawal" samples.

21. We have survey data for whether students visited the financial aid offices or met with advising staff, and these survey data show no statistically significant differences between program and control students on use of these other services. However, only about two thirds of students in the experiment responded to the follow-up survey. We concentrate on the outcomes for which we have administrative records for all students.

22. There are alternative methodologies, for example, hierarchical linear models (HLM). Angeles and Mroz (2001) compare methodologies and conclude that linear regression with clustered standard errors delivers valid, robust estimates.

23. There is one final caveat to the intent-to-treat analysis: The classroom-level random assignment makes it difficult to guarantee a pure intent-to-treat analysis. As explained above, the registrar's census of students in classes took place after the initial meeting of Volunteers with the program classrooms. Thus, it is possible that some students exited the class between that initial meeting and the census. Although the summary statistics do not suggest that this happened, this potential problem provides further reason to examine the program impact using regression analysis: We can hold constant characteristics, like math placement test scores, to ensure that we are only estimating the program impact from differences in program and control students who are similar.

24. These individual controls include female, young (≤ 24 years old), part-time status, bottom or top quartile of math placement test scores (within sample), and an indicator for missing placement test score.

25. There are 35 different teachers in the sample.

26. Other estimates of the impact of the program on the number of visits suggest this is not just a one-time mechanical increase stemming from the fact that

some Volunteers showed their classes the tutoring center (Visher, Butcher, and Cerna, 2010).

27. (5.79 percentage points/18.9 percentage points) $\times 100 = 30.60\%$.

28. Ex ante power calculations were done with a range of assumptions for intraclass correlations: .01, .05, and .10. The measured intraclass correlation for pass rates is .13 in this sample. Minimum detectable effect sizes using the ex post sample characteristics are about 0.22 of a standard deviation. The top of the 95% confidence interval (CI) for the estimated program impact on passing is 0.053, which is about 0.10 of a standard deviation of the pass rate for the control group. This is close to the simulated minimum detectable effect size that had the intraclass correlation been .01 (the lowest level considered in the ex ante calculation) rather than .13.

29. Campus staff often referred to part-time students as "PCP" students: parking lot-classroom-parking lot.

30. A finding that is borne out in this sample: Results in column 4 of Table 3 show that part-time students are about 8 percentage points less likely to register for a subsequent semester than full-time students.

31. Note that part-time students make up nearly half of the sample; there are part-time students in every class section, so the experimental design is valid for examining outcomes for this subgroup.

32. Note that it is impossible to tell whether it is the same students who were induced to stay in the course by the program who also increased the pass rate.

33. The coefficient on the interaction term is not statistically different from zero at conventional levels, indicating that although the estimated impact on pass rates for part-time students is statistically different from zero, it is not statistically significantly different from the impact on pass rates for full-time students, which is imprecisely estimated.

34. In the control group, there were no statistically significant differences between part-time and full-time students in probability of visiting the tutoring center, withdrawing, or passing, so we use the overall control group mean, shown in Table 1, to interpret these coefficients as percentage changes. Control group part-time and full-time students were significantly different in their probabilities of registering for the subsequent semester: 62% of full-time students registered for the next semester, and part-time students were 7.9 percentage points less likely to do so (*p*-value for difference: 0.004).

35. Anecdotally, only a handful of students declined individualized random assignment.

36. The Baseline Information Form (BIF) survey collects a rich array of demographics and background information. As we do not have this information for all study participants, it is not used in the analyses.

However, to the extent that BIF information overlaps with the administrative records, the information is the same in both.

37. The type of sorting that would cause a problem for using fixed effects estimates as measures of teacher quality is if a teacher has a reputation for being nice to students with poor math preparation, and thus that teacher has a class disproportionately comprised of students who are likely to perform poorly on the final exam, whether or not she teaches them well. As we can control for math placement test scores in the regressions below, and some other personal characteristics, any sorting that would bias the estimated teacher effects must be along other (unobservable) dimensions.

38. Note that even in the case where students are randomly assigned to teachers, it is controversial whether teacher fixed effects in exam score regressions capture quality. Carrell and West (2010) find evidence that teachers whose students perform better on the final exam in that particular class may be sacrificing long-term understanding and problem-solving skills in favor of the better short-term outcome, potentially because these teachers "teach to the test."

39. Note that using test scores to create a teacher quality measure requires us to limit the sample to students who did not withdraw from the course prior to the final exam, and thus it is estimated for a select sample of students.

40. This teacher quality measure is very similar whether or not student characteristics are held constant, suggesting that if different types of students selected into classes taught by teachers of differing quality, this selection was along unobservable student characteristics.

41. Note that there are many different potential ways to define quality with the available data. One could use the probability of withdrawal as the outcome and designate those teachers whose students are least likely to withdraw as the highest quality; as mentioned above, using test score outcomes follows the previous literature. There are also many different ways to investigate the Program × Quality interaction. With enough teachers, a continuous variable capturing quality might make sense, for example. Given the small sample of teachers, we simply divide the sample between top half and bottom half of the quality measure.

42. The final exam score outcome in column 3 is a self-selected sample comprised of the 672 students who stayed in the class and took the common final exam.

References

Aaronson, D., Barrow, L., & Sanders, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics, 25*, 95–135.

Adams, W. W. (2003). Developmental mathematics: A new approach. *The Mathematical Association of America*. Retrieved from http://www.maa.org/features/112103devmath.html

Adams, P. S., Miller, G. R., & Roberts, A. (2009). The accelerated learning program: Throwing open the gates. *Journal of Basic Writing, 28*(2), 50–69.

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

Angeles, G., & Mroz, T. (2001). *A guide to using multilevel models for the evaluation of program impacts* (Measure Evaluation Working Paper Series, WP-01-33). Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill.

Angrist, J., Lang, D., & Oreopoulos, P. (2009). Incentives and services for college achievement. Evidence from a randomized trial. *American Economic Journal: Applied Econometrics, 1*, 136–163.

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

Bailey, T., Jeong, W. D., & Cho, S.-W. (2009). *Referral, enrollment, and completion in developmental education sequences in community colleges* (CCRC Working Paper No. 15). New York, NY: Community College Research Center, Teachers College, Columbia University.

Barrow, L., Brock, T., Richburg-Hayes, L., & Rouse, C. (in press). Paying for performance: The impacts of a community college scholarship program for low-income adults. *Journal of Labor Economics*.

Bertrand, M., & Adair, M. (2011). Information disclosure, cognitive biases, and payday borrowing. *Journal of Finance, 66*, 1865–1893.

Bettinger, E., & Baker, R. (2011). *The effects of student coaching in college: An evaluation of a randomized experiment in student mentoring* (National Bureau of Economics Research Working Paper 16881). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w16881

Bettinger, E. P., & Terry Long, B. (2009). Addressing the needs of underprepared students in higher education: Does college remediation work. *Journal of Human Resources, 44*, 736–771.

Bordes, V., Arredondo, P., Robinson, K. S., & Rund, J. (2011). A longitudinal analysis of Latina/o students' academic persistence. *Journal of Hispanic Higher Education, 10*, 358–368.

Brock, T. (2010). Young adults and higher education: Barriers and breakthroughs to success. *Future of Children, 20*, 109–132.

Brock, T., & Richburg-Hayes, L. (2006, May). *Paying for persistence: Early results of a Louisiana scholarship program for low-income parents attending community college*. Retrieved from http://www.mdrc.org/sites/default/files/full_472.pdf

Bryk, A. S., & Treisman, U. (2010, April 18). Make math a gateway, not a gatekeeper. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/Make-Math-a-Gateway-Not-a/65056/

Carrell, S. E., & West, J. E. (2010). Does professor quality matter: Evidence from random assignment of students to professors. *Journal of Political Economy, 118*, 409–432.

Daponte, B., Osborne, S. S., & Taylor, L. (1999). Why do low-income families not use food stamps? Evidence from an experiment. *Journal of Human Resources, 34*, 612–628.

Deil-Amen, R., & Rosenbaum, J. E. (2003). The social prerequisites of success: Can college structure reduce the need for social know-how. *Annals of the American Academy of Political and Social Science, 586*, 120–143.

Goldrick-Rab, S. (2010). Challenges and opportunities for improving community college student outcomes. *Review of Educational Research, 80*, 437–469.

Hoachlander, G., Sikora, A. C., & Horn, L. (2003). *Community college students: Goals, academic preparation, and outcomes*. Washington, DC: National Center for Education Statistics.

Hurtado, S., & Ponjuan, L. (2005). Latino educational outcomes and the campus climate. *Journal of Hispanic Higher Education, 4*, 235–251.

Jenkins, D., Speroni, C., Belfield, C., Shanna, S. J., & Edgecombe, N. (2010). *A model for accelerating academic success of community college remedial English students: Is the accelerated learning program (ALP) effective and affordable?* (CCRC Working Paper No. 21). Retrieved from http://ccrc.tc.columbia.edu/media/k2/attachments/remedial-english-ALP-effective-affordable.pdf

Knapp, L. G., Kelly-Reid, J. E., & Ginder, S. A. (2009). *Enrollment in postsecondary institutions, Fall 2007; Graduation rates, 2001 & 2004 cohorts; and financial statistics, fiscal year 2007: First look*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Laden, B. V. (1999). Socializing and mentoring college students of color: The Puente Project as an exemplary celebratory socialization model. *Peabody Journal of Education, 74*, 55–74.

Martorell, P., & McFarlin, I., Jr. (2011). Help or hindrance? The effects of college remediation on academic and labor market outcomes. *The Review of Economics and Statistics, 93*, 436–454.

Patel, R., & Richburg-Hayes, L. (2012). *Performance based scholarships: Emerging findings from a national demonstration*. New York, NY: MDRC.

Perry, M., Bahr, P. R., Rosin, M., & Woodward, K. M. (2010). *Course-taking patterns, policies, and practices in developmental education in the California community colleges*. Mountain View, CA: EdSource.

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

Scrivener, S., & Weiss, M. J. (2009). *More guidance, better results?: Three-year effects of an enhanced student services program at two community colleges*. New York, NY: MDRC.

Sedlacek, W. E. (2010). Noncognitive measures for higher education admissions. In P. L. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (3rd ed., pp. 845–849). Amsterdam, Netherlands: Elsevier.

Sedlacek, W. E., Benjamin, E., Schlosser, L. Z., & Sheu, H. B. (2007). Mentoring in academia: Considerations for diverse populations. In T. D. Allen & L. T. Eby (Eds.), *The Blackwell handbook of mentoring: A multiple perspectives approach* (pp. 259–280). Malden, MA: Blackwell.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Visher, M. G., Butcher, K. F., & Cerna, O. S. (2010). *Guiding developmental math students to campus services: An impact evaluation of the Beacon Program at South Texas College*. New York, NY: MDRC.

Visher, M. G., Schneider, E., Wathington, H., & Collado, H. (2010). *Scaling up learning communities: The experience of Six Community Colleges*. New York, NY: MDRC.

Visher, M. G., Wathington, H., Richburg-Hayes, L., & Schneider, E. (2008). *The learning communities demonstration: Rationale, sites, and research design* (An NCPR Working Paper). New York, NY: National Center for Postsecondary Research.

Visher, M. G., Weiss, M. J., Weissman, E., Rudd, T., & Wathington, H. (2012). *The effects of learning communities for students in developmental education: A synthesis of findings from Six Community Colleges*. New York, NY: MDRC.

Weiss, M. J., Visher, M. G., & Wathington, H. (2010). *Learning communities for students in developmental reading: An impact study at Hillsborough Community College*. New York, NY: MDRC.

Zachry, E. M., & Schneider, E. (2008). *Promising instructional reforms in developmental education: A case study of three Achieving the Dream colleges*. New York, NY: MDRC.

Zachry, E. M., & Schneider, E. (2011). *Unlocking the gate*. New York, NY: MDRC.

Authors

KRISTIN F. BUTCHER is the Marshall I. Goldman Professor of economics at Wellesley College. Her research is in the areas of health, education, and labor economics.

MARY G. VISHER is a senior associate at MDRC. She works in both the K–12 and Young Adults and Postsecondary Education Policy Areas.