

## RESEARCH ARTICLE

# From Innovation to Impact

## How Higher Education Can Evaluate Innovation's Impact and More Precisely Scale Student Support

by Mark Milliron, David Kil, Laura Malcolm, and Grace Gee

*Rigorously evaluating the impact of innovative student success initiatives is key in meeting institutional goals for student outcomes, resource allocation, and return on investment.*

### INTRODUCTION

INTEGRATED PLANNING STRATEGIES are being used more prominently by colleges and universities to track and assess progress toward their students' success and their own institutional goals. These strategies emphasize the importance of making decisions and justifying resource allocations through supportive data and a culture of evidence. To this end, institutional plans should not only state goals, rollout strategies, and resource needs, but also include commitments to assess both new and existing initiatives in order to measure what was actually accomplished and learn what could be done differently in the future, including potentially reallocating and reprioritizing existing resources.

*Institutional plans should not only state goals, rollout strategies, and resource needs, but also include commitments to assess both new and existing initiatives.*

The University of Arizona and Austin Community College provide prime examples of how institutions are making the assessment of initiatives and programs critical to institution-wide decision making. We recently worked with these institutions to assess the effectiveness of two campus-based service centers so that they could better understand the true impact of those services on student success and gather evidence to support continued investment.

Using statistically rigorous impact analysis, institutions are able to facilitate a deeper assessment of key initiatives and investments, including policy changes, student success initiative pilots, outreach strategies, and interventions. Equipped with this information, institutional leadership can better understand what is working, for whom, under what context, and at what time. This understanding will further lead to an awareness of how to proactively improve student success in the most effective and efficient manner, and, additionally, it will help focus limited resources.

*Equipped with this information, institutional leadership can better understand what is working, for whom, under what context, and at what time.*

We use prediction-based propensity score matching (PPSM), a methodology compliant with the U.S. Department of Education's What Works Clearinghouse's requirements, to systematically measure efficacy, ensuring that outcomes of students participating in the initiative being analyzed are compared to control students with similar propensity. This impact analysis illuminates the context and shows the covariates used by the propensity score matching model to find like pilot-control student pairs to compare against, the details of matching and model quality assurance like covariate distributions and calibration plots, and the statistical significance of measured results. The work is

transparent in terms of the quality of analysis—users can view detailed matching information such as percentage of pilot-control matches, top covariates/features used in matching, and pre-post matching distributions for the pilot and control groups, as well as model quality assurance details like ROC curves, calibration plots, and bias-variance trade-offs.

## AUSTIN COMMUNITY COLLEGE ACCELERATOR

Austin Community College (ACC) recently developed a high-tech learning laboratory called the ACCelerator headquartered at its Highland campus in Austin, Texas. The ACCelerator opened in Fall 2014 and is designed to offer technology-based instruction through adaptive online learning modules. There are 604 computer stations available to facilitate personalized learning and small group sessions. In addition, the ACCelerator has an extensive support network of faculty, counselors, advisors, tutors, librarians, and other staff to help students meet their educational needs and goals. The lab is spread over 32,000 square feet, with clusters of desktop computer stations surrounded by classrooms and study rooms.

While open to all students, many who leverage the ACCelerator do so to develop core skills and complete developmental education coursework. The ACCelerator offers an innovative developmental math course, Developmental Mathematics (MATD 0421), which provides students the opportunity to reach college-level math at their own pace. Other programs and services offered at the ACCelerator include tutoring in a variety of subjects, a first-year experience, group advising sessions, academic coaching, adult and continuing education programs, college readiness assessment and preparation, and student skills workshops.

After having the program in place for a full academic year, ACC was interested in understanding whether the ACCelerator was having the desired impact on student outcomes. Specifically, it was interested in knowing whether

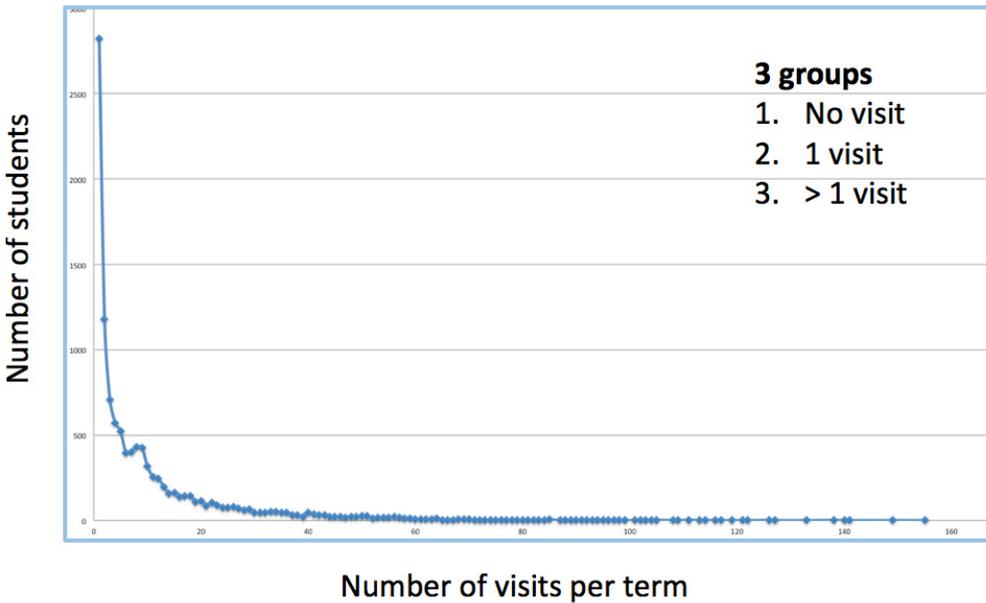
visits to the ACCelerator were correlated with higher persistence rates. The challenge with measuring correlation in this case is that participation in the ACCelerator programs is voluntary, and, as in many initiatives where students opt in, there can be many confounders in understanding whether a program is working, including selection bias. It's expected that students who make use of opt-in services, like drop-in support services, also exhibit other factors that make them more likely to persist anyway, so that comparing outcomes of students who voluntarily use such services to those who do not is not an accurate evaluation of the effectiveness of those services. To address this challenge, ACC worked with Civitas Learning to develop a study model that would more accurately measure the true impact of the ACCelerator.

Since this was an observational study, PPSM with baseline equivalence was used to control for selection bias and meet What Works Clearinghouse guidelines with reservation. PPSM is used to identify a comparable control group that is statistically indistinguishable from the pilot or participating group when a randomized controlled trial is not used to assess an initiative or program. PPSM matches pilot students to control students based on their similar likelihood to participate in the initiative (the propensity score)—in this case, use the ACCelerator services—and similar likelihood to achieve a certain outcome (the prediction score)—in this case, persist.

With the necessary student and usage data in hand, we analyzed the impact of visits to the ACCelerator across Fall 2014, Spring 2015, and Summer 2015. First, using attendance data from the institution's tracking system, we identified the list of students who visited the ACCelerator and considered those students the pilot or program participant group. Since access to the ACCelerator is open to all students, we considered the eligible control group (the group of students valid for PPSM matching) to be all other students who did not visit the ACCelerator during the same term. Based on this process, three groups were created for comparison: no visit (control), one visit (pilot 1), and > 1 visit (pilot 2).

Figure 1 shows the histogram of the number of visits to the ACCelerator by all ACC students. Students from the no-visit group (control) were matched for each of the pilot groups.

Figure 1 **Number of Visits to the ACCelerator Per Term By All Students**



The histogram of the number of visits has an exponential distribution as expected.

Next, a family of models was built to identify top predictors, divided into new- and returning-student covariates since new and returning students have different data footprints. The top predictors identified for these student segments were used to build prediction models for term-to-term persistence and propensity models for participation in the ACCelerator services.

Finally, the persistence prediction and propensity scores from these models were used to match pilot students to eligible control students. After the matched set of pilot and control students was identified, the groups were validated to be statistically indistinguishable enough for an impact analysis of ACCelerator services.

The prediction score probability density functions (PDFs) of the three groups before matching (figure 2) show significant differences between those students who visit the ACCelerator and those who don't. Immediately we see that the pre-match control group's PDF is shifted to the left, meaning that group

members are projected to perform poorly in comparison to the other two groups before matching, most likely due to selection bias. (Students who choose to go to the ACCelerator are more likely to persist anyway as demonstrated by other data factors.)

It is also interesting to note that students who visit the ACCelerator more than once fall more into the middle region of the PDF while those who visit only once have higher representations at the two extreme ends. This observation points to the importance of matching to ensure that selection bias doesn't contribute to overly optimistic program impact estimation.

Looking at the same prediction score PDFs post-matching (figure 3), we see a complete overlap in the PDF lines between the pilot and control groups. This indicates virtually indistinguishable differences in predicted persistence rates between pilot and control groups post-matching.

Figure 2 Persistence Prediction Score Distributions Prior to Matching

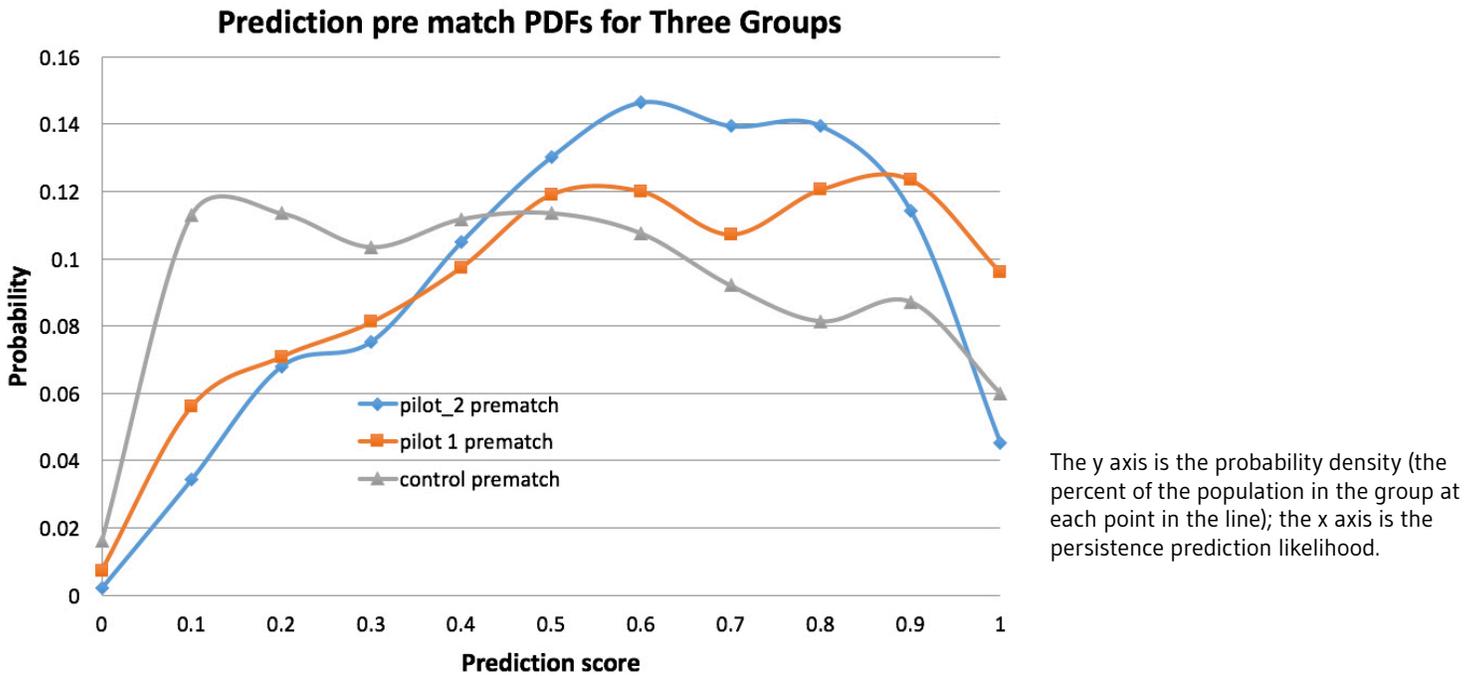
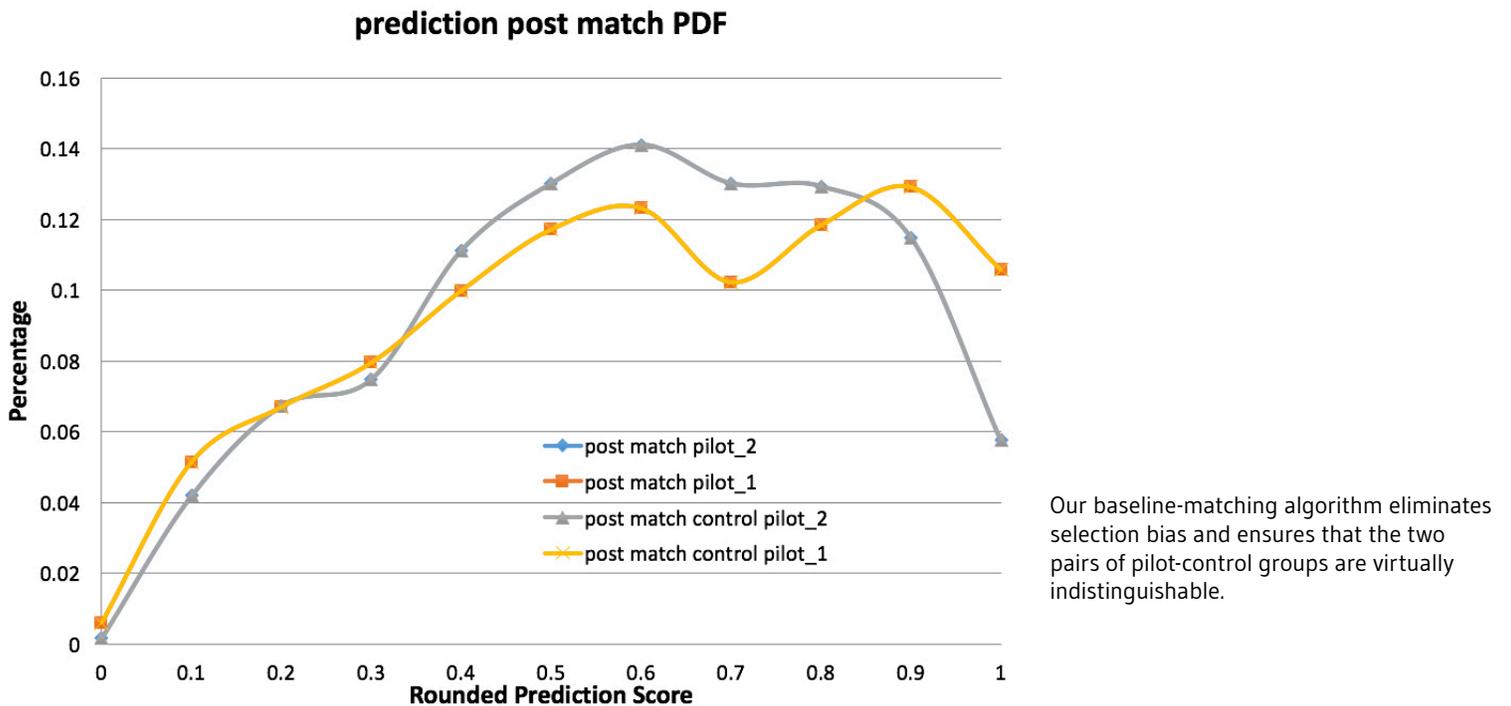
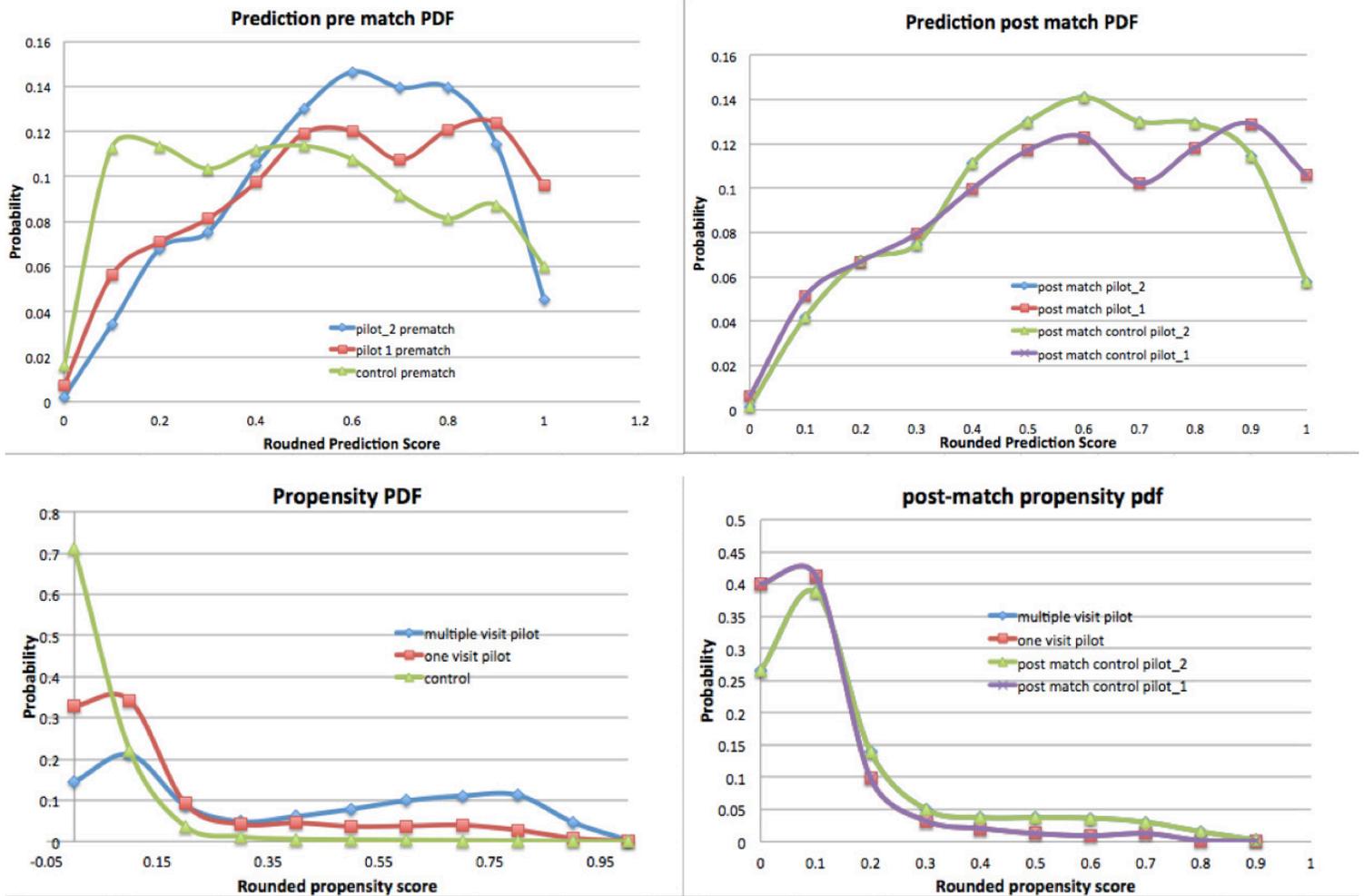


Figure 3 Persistence Prediction Score Distributions Post-Matching



Pre- and post-matching comparison plots for both prediction and propensity scores (figure 4) show again that there are complete overlaps between the pilot and control groups.

Figure 4 Pre- and Post-Matching PDF Comparisons for Prediction and Propensity Scores



Once we have controlled for selection bias, looking at the impact on term-to-term persistence, we see that the ACCelerator has a positive impact on all students who visited it at least once, with > 4 percentage points of improvement across terms. Figure 5 summarizes the impact measured in percentage point improvement in persistence for all students, including those enrolled in developmental education programs.

The persistence impact is particularly strong for students in developmental education programs who visit the ACCelerator more frequently, with > 10 percentage points of improvement in their persistence. The lone negative number for Summer 2015 students in developmental education programs with one visit is not statistically significant due to a very low sample size of 31.

Even for students not enrolled in developmental education programs, visiting the ACCelerator has a salient impact on

persistence. However, no dosage effect exists between one visit and multiple visits for these students. Furthermore, the

overall impact size, while positive, is much smaller than for students in developmental education programs.

Figure 5 Impact Analysis Results Summary

all students	prediction diff	prediction pilot	prediction control	pilot outcome	control outcome	N match pairs	possible matches	match rate	Impact % increase	std dev	z-score	p value
> 1 visit   Fall-2014	0.001449	0.645402	0.645342	0.740113	0.675141	1062	1796	0.59131	<b>6.497%</b>	0.01822	3.56513	<b>0.00018</b>
> 1 visit   Spr-2015	0.001665	0.559065	0.559012	0.633373	0.574163	1672	3024	0.55291	<b>5.921%</b>	0.01547	3.82684	<b>0.00007</b>
> 1 visit   Sum-2015	0.001775	0.578089	0.578156	0.621749	0.580378	846	1798	0.47052	<b>4.137%</b>	0.01929	2.14499	<b>0.01598</b>
1 visit   Fall-2014	0.000829	0.672866	0.672692	0.694845	0.635052	485	590	0.82203	<b>5.979%</b>	0.02483	2.40854	<b>0.00801</b>
1 visit   Spr-2015	0.000912	0.574544	0.574595	0.603871	0.554839	775	912	0.84978	<b>4.903%</b>	0.02216	2.21282	<b>0.01346</b>
1 visit   Sum-2015	0.001103	0.575061	0.575106	0.582524	0.580097	412	529	0.77883	<b>0.243%</b>	0.02724	0.08910	<b>0.46454</b>
deved students	prediction diff	prediction pilot	prediction control	pilot outcome	control outcome	N match pairs	possible matches	match rate	Impact % increase	std dev	z-score	p value
> 1 visit   Fall-2014	0.00201	0.61647	0.61636	0.76620	0.64507	355	763	0.46527	<b>12.113%</b>	0.03210	3.77323	<b>0.00008</b>
> 1 visit   Spr-2015	0.00256	0.54976	0.54975	0.64990	0.52314	497	1294	0.38408	<b>12.676%</b>	0.02848	4.45062	<b>0.00001</b>
> 1 visit   Sum-2015	0.00367	0.62458	0.62483	0.77477	0.63964	111	619	0.17932	<b>13.514%</b>	0.05577	2.42311	<b>0.00769</b>
1 visit   Fall-2014	0.00176	0.57060	0.57039	0.63855	0.49398	83	105	0.79048	<b>14.458%</b>	0.06356	2.27470	<b>0.01146</b>
1 visit   Spr-2015	0.00232	0.52763	0.52739	0.59794	0.51546	97	142	0.68310	<b>8.247%</b>	0.06293	1.31061	<b>0.09500</b>
1 visit   Sum-2015	0.00387	0.63348	0.63242	0.64516	0.77419	31	54	0.57407	<b>-12.903%</b>	0.10255	-1.25822	<b>0.10420</b>
non deved students	prediction diff	prediction pilot	prediction control	pilot outcome	control outcome	N match pairs	possible matches	match rate	Impact % increase	std dev	z-score	p value
> 1 visit   Fall-2014	0.00117	0.65993	0.65989	0.72702	0.69024	707	1033	0.68441	<b>3.678%</b>	0.02213	1.66205	<b>0.04826</b>
> 1 visit   Spr-2015	0.00129	0.56300	0.56293	0.62638	0.59574	1175	1730	0.67919	<b>3.064%</b>	0.01843	1.66254	<b>0.04834</b>
> 1 visit   Sum-2015	0.00149	0.57107	0.57111	0.59864	0.57143	735	1179	0.62341	<b>2.721%</b>	0.02054	1.32474	<b>0.09264</b>
1 visit   Fall-2014	0.00064	0.69398	0.69381	0.70647	0.66418	402	485	0.82887	<b>4.229%</b>	0.02692	1.57069	<b>0.05814</b>
1 visit   Spr-2015	0.00071	0.58126	0.58135	0.60472	0.56047	678	770	0.88052	<b>4.425%</b>	0.02367	1.86902	<b>0.03081</b>
1 visit   Sum-2015	0.00088	0.57031	0.57044	0.57743	0.56430	381	475	0.80211	<b>1.312%</b>	0.02825	0.46452	<b>0.32115</b>

The overall impact percentage points show clearly that the ACCelerator program helps students in developmental education programs (double-digit increases in persistence).

Further, we performed drill-down impact analyses, creating multiple segments and matching at the drill-down segment level. Key findings are as follows:

- » The lower the persistence prediction scores for students in developmental education programs, the higher the impact of visiting the ACCelerator. Impact improves from 10.25 percentage points to 12.44 percentage points to 17.75 percentage points as the persistence prediction score ranges are lowered from 0.6–1.0 to 0.4–0.6 to 0–0.4.
- » Part-time students in developmental education programs who frequent the ACCelerator improve their persistence by 14.07 percentage points versus full-time developmental education students who improve by 9.01 percentage points.
- » There is a dosage effect in terms of visit frequency for students enrolled in developmental education programs. In this case, students who visit more than four times improve their persistence by 15.49 percentage points.
- » The statistically significant persistence improvement benefits apply to students of all experience levels, ranging from 11.33 percentage points to 13.87

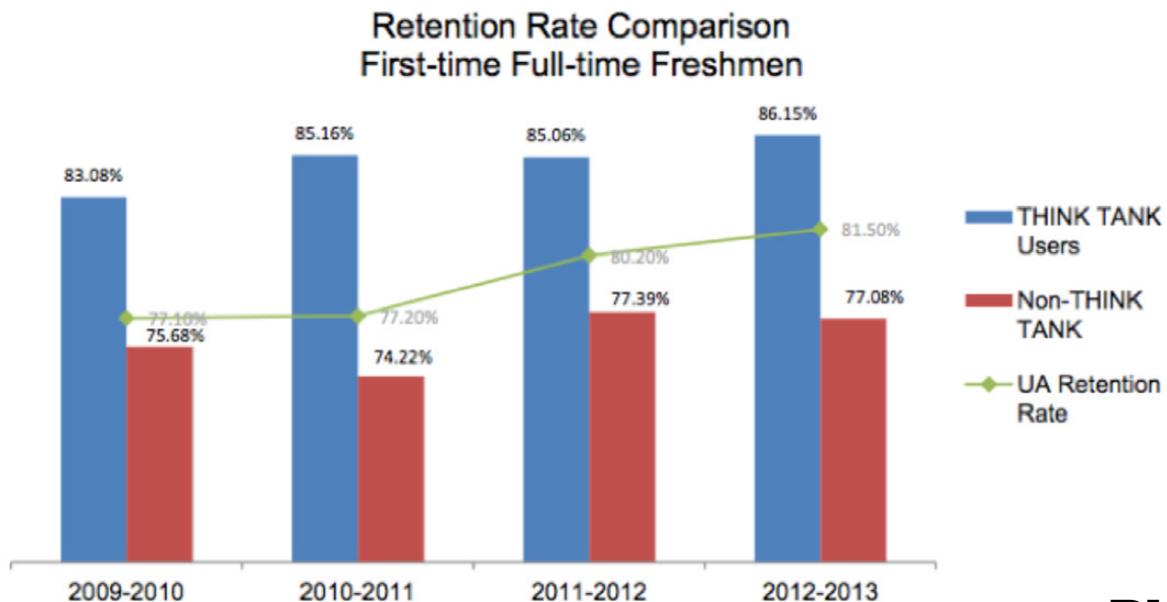
percentage points in a consistent manner across all terms completed.

Given the results of this analysis, ACC has leaned into its investment in the ACCelerator. The college has expanded the program’s offerings and services in order to continue to grow student success. In addition, it has increased marketing of the ACCelerator and targeted the students who benefit most. Growth in student participation has been dramatic and will continue to provide data for analysis of the ACCelerator’s ongoing benefit.

### UNIVERSITY OF ARIZONA THINK TANK

The University of Arizona (UA), located in Tucson, serves over 50,000 students and provides a large centralized student support center called THINK TANK that offers several academic support services including a Writing Center, math tutoring, and supplemental instruction. Most of these services are free for students, and to ensure convenience and encourage participation, four physical locations on campus as well as online services are available. The THINK TANK has been open for several years, and descriptive data indicate that the services are working (figure 6).

Figure 6 Descriptive Statistics of Year-to-Year Retention



Specifically, first-time full-time freshmen who used THINK TANK services showed 4–8 percent higher persistence rates than the institutional average and 7–11 percent higher persistence rates when compared to first-time full-time freshmen who did not make use of the service. UA was particularly interested in seeing the impact on this group of students since a majority of THINK TANK services are focused on supporting them.

However, these descriptive statistics do not account for selection bias, the likely possibility that the students being compared against each other, in this case THINK TANK users and non-users, are too different to do a true apples-to-apples comparison.

To account for selection bias in the measurement of impact on student success, UA's Student Affairs Assessment and Research team had previously conducted analyses using propensity score matching (PSM). PSM is a statistical matching technique used to estimate the effect of a treatment, service, or other intervention by controlling for the covariates that may predict receiving the treatment or service. Prior PSM analyses had focused on the measurement of supplemental instruction programs in support of course success. In the case of the analysis of THINK TANK services, there was interest in applying, enhancing, and automating this methodology to control for selection bias.

Therefore, the next step in UA's analysis—in partnership with Civitas Learning—was automating the ability to run PSM with the added rigor of using a prediction as part of propensity score matching—PPSM. The propensity and prediction scores used for matching are determined by several covariates, or representative data about the students, including demographic information, financial aid, socioeconomic status, academic performance, incoming factors including test scores and transfer credits, student behavior including online activity, and more. These covariates are collected and derived from multiple institutional systems such as the

student information system (SIS), learning management system (LMS), management and tracking systems, and more.

In order to ensure an accurate analysis of efficacy can be measured for a program like the THINK TANK, it's crucial that not only a breadth of data representing the student and his/her academic career is available for creating covariates and building prediction and propensity models, but that detailed data about the program or initiative itself is collected so that the analysis is done in the appropriate context. UA was able to quickly and rigorously measure the impact of its THINK TANK services since its Student Affairs Assessment & Research group had student services usage data readily available through a dedicated tutoring management system.

Using this data we analyzed the impact of three services within the THINK TANK—the Writing Center, math drop-in tutoring, and supplemental instruction—across four years from Spring 2011 to Fall 2015. First, we identified the list of students who used THINK TANK services each term and considered those students the pilot or program participant group. Since the Writing Center and math tutoring services were available for any student to take advantage of voluntarily, we considered the eligible control group, or the group of students valid for PPSM to match against, to be all students who did not use THINK TANK services during that same term. Since supplemental instruction was only offered for certain courses, the eligible control group for those services was defined more narrowly to be students in those courses who did not participate.

After identifying the pilot and eligible control groups for each of the THINK TANK services, we built several models for the different student segments identified based on data availability and similarity of top predictive covariates. For example, new students and continuing students typically have vastly different data footprints and should have separate models for improved accuracy and robustness. The top predictors identified for these student segments were used to build prediction models for term-to-term persistence

and propensity models for participation in THINK TANK services.

Finally, the persistence prediction and propensity scores from these models were used to match pilot students to eligible control students. After the matched set of pilot and control students was identified, the two groups were validated to be statistically indistinguishable enough for impact analysis

of THINK TANK services. The covariates used in matching included demographic, census, financial aid, socioeconomic, academic performance, course load and degree pathway, test score, transfer, and behavioral data. As an example, figures 7 and 8 show the pre-matching and post-matching distributions of the prediction and propensity scores for the pilot and control groups identified for the Writing Center analysis.

Figure 7 Pre-Matching and Post-Matching Persistence Scores for the Writing Center

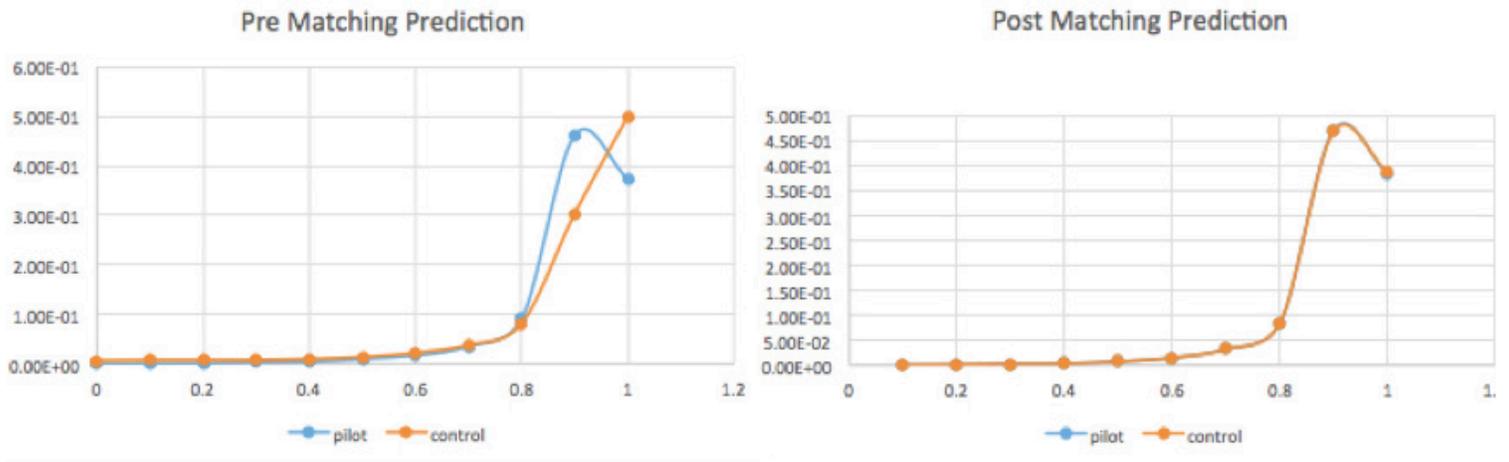
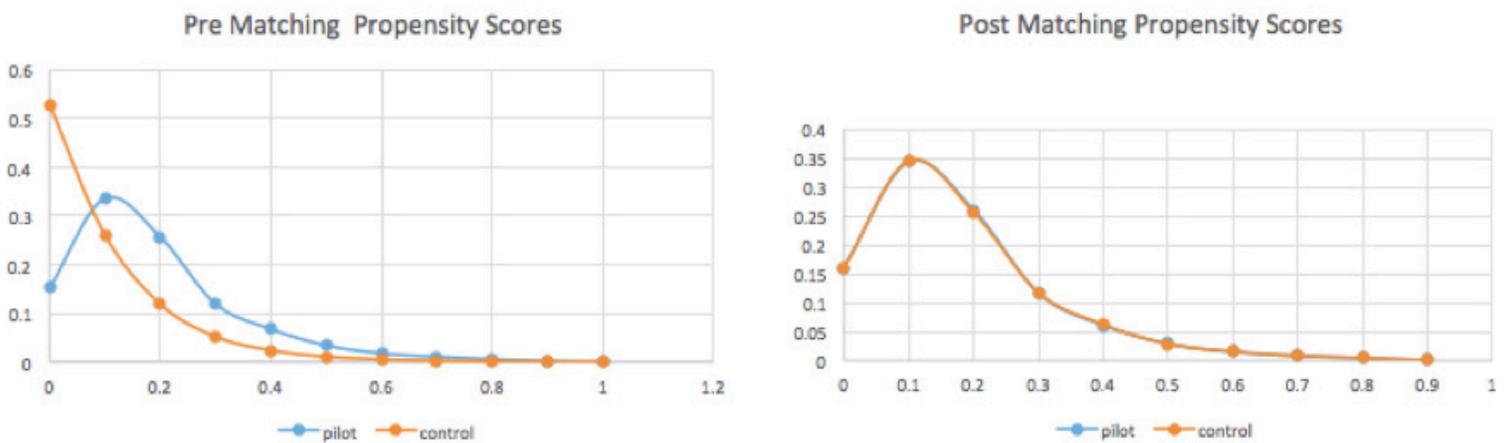


Figure 8 Pre-Matching and Post-Matching Propensity Scores for the Writing Center



Note the difference in the score distributions prior to matching, indicating selection bias and inherent differences in outcomes that may contribute to overly optimistic impact estimation. In other words, the 7–11 percent persistence rate

difference shown in the descriptive data between students who used THINK TANK services and those who did not is most likely an overestimate of the impact of those services.

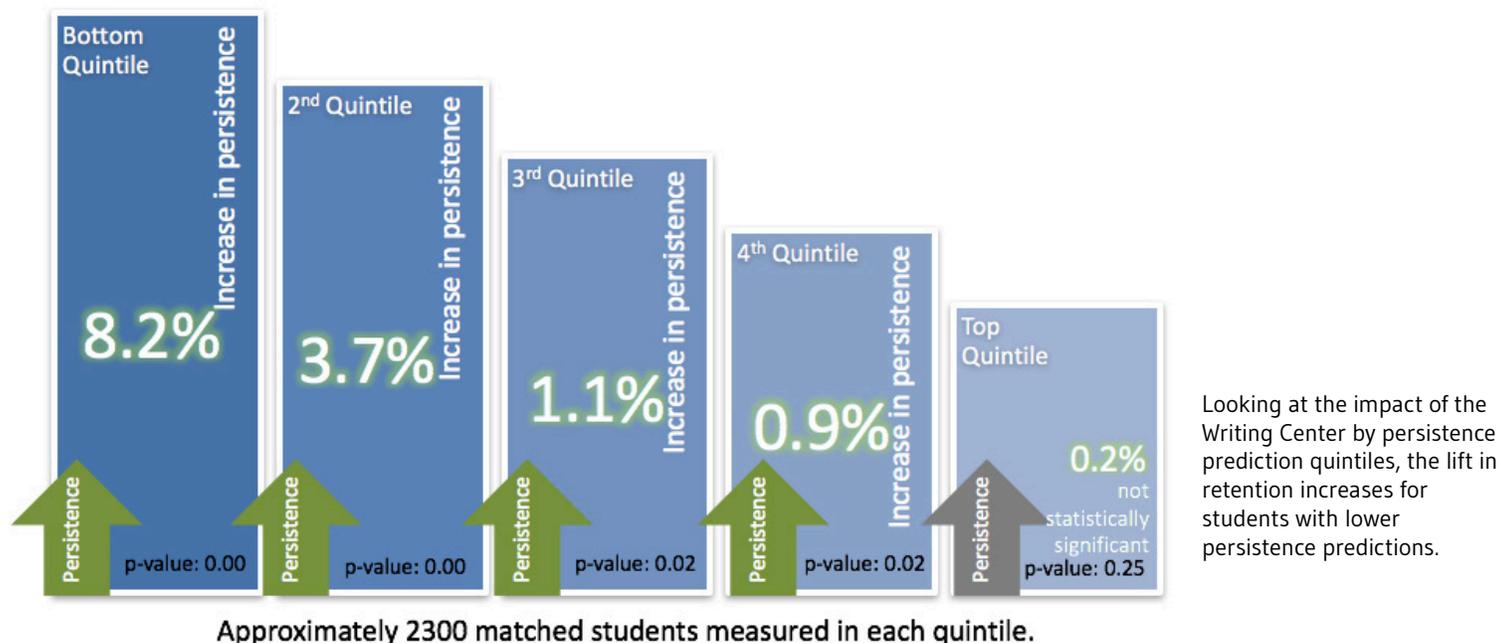
However, after a subset of pilot and control student matches was identified through PPSM, the post-matching prediction and propensity score distribution graphs show that the two student groups are nearly indistinguishable; in other words, an apples-to-apples comparison of their outcomes can be made. Comparison of the outcomes between these two matched groups will provide a much more accurate measure of the impact of the Writing Center services.

After using PPSM to analyze four years of data across the Writing Center, math drop-in tutoring, and supplemental instruction THINK TANK services, we measured a statistically significant ( $p\text{-value} < 0.05$ ) 2.3 percentage point increase in persistence for students who took advantage of those services versus similar students who did not, a stark contrast to the 8 percentage point increase indicated with purely descriptive data. This 2.3 percentage point

improvement equated to an additional 587 students persisting and at least \$3.3 million in additional term tuition as well as an estimated \$7.5 million in additional tuition over those persisting students' academic careers.

Using the student segments identified earlier, we also did a drill-down analysis to determine the impact of THINK TANK services on specific types of students to better understand how the services could be more effectively focused or promoted. For example, we discovered that when comparing the impact by persistence prediction quintiles, which evenly distribute the student population into five groups based on their persistence predictions from highest risk to lowest risk, THINK TANK services were most impactful for the highest-risk students—up to an 8.2 percentage point statistically significant increase in persistence for the students with the lowest persistence likelihood, as shown in figure 9.

Figure 9 Persistence Rate Increases by Quintile for Students Using the Writing Center



Additional drill-down insights were uncovered in the analysis process, such as:

» There was a 2.3 percentage point statistically significant increase in persistence for STEM majors who used

the Writing Center, showing the importance of these services even for majors outside of the liberal arts.

» There was a 2.7 percentage point statistically significant increase in persistence for first-time full-time students who used the Writing Center and a 1.9 percentage

point statistically significant increase in persistence for those who used the math drop-in tutoring, both results slightly higher than for non-first-time full-time students.

Based on these insights, UA has plans to create strategic campaigns to reach out to students and encourage use of THINK TANK services based on their persistence predictions, particularly as part of the first-year experience. Findings will also be shared with academic colleges and faculty to advocate for the referral of students to THINK TANK services and the expansion of supplemental instruction, another THINK TANK service with demonstrated impact, to more courses. Finally, overall, these findings provide support for the continued investment in services that have proven to help students and improve outcomes for wider institutional leadership.

## CONCLUSION

Providing this type of statistically rigorous impact analysis to more accurately determine return on investment for these types of services will be critical for future institutional planning efforts—e.g., planning that brings together academic, financial, and facilities planning—and for fostering a culture of optimization in making decisions to help reach institutional goals, particularly around student outcomes. Whether an institution is focused on realizing big gains from a single initiative or looking to achieve smaller but still meaningful gains from several different programs or outreach efforts, this type of impact analysis provides the timely, relevant, and strong signals needed to more confidently take action and dramatically affect student outcomes.

*Providing statistically rigorous impact analysis to more accurately determine return on investment will be critical for future institutional planning efforts.*

## AUTHOR BIOGRAPHIES

MARK MILLIRON, PH.D., is an award-winning leader, author, speaker, and consultant who has worked with universities, community colleges, K–12 schools, foundations, corporations, associations, and government agencies across the country and around the world. As co-founder and chief learning officer at Civitas Learning, he is involved in all areas of the company's development, especially in helping to catalyze a thriving learning community around analytics and student success initiatives with partner institutions. He serves on numerous corporate, nonprofit, and education boards and advisory groups, including the Texas Student Success Council. In 1999, The University of Texas at Austin's College of Education named him a Distinguished Graduate for his service to the education field. In 2007, the American Association of Community Colleges presented him with its National Leadership Award; in 2011, the National University Technology Network named him the recipient of the Distinguished Service Award; and in 2013, he was inducted into the United States Distance Learning Association's Hall of Fame.

DAVID KIL has more than 20 years of experience in building various analytics apps and solutions spanning nonlinear time-series analysis to predictive analytics, outcomes research, and user experience optimization. He and his team are working on (1) improving predictive algorithms to provide much more actionable insights, (2) adding new capabilities to automate ROI and outcomes analyses as part of action analytics, and (3) making the Civitas Learning analytics platform self-learning and more intelligent over time. He holds 13 U.S. patents, is the author of a book on pattern recognition and predictions, and has published a number of articles in journals. He currently serves as chief data scientist with Civitas Learning.

LAURA MALCOLM has nearly 25 years of education and product design experience in building technology products to help people learn and achieve their educational goals. With an MA.Ed. from Stanford in learning, design, and technology, she has spent more than 15 years in executive leadership roles directing the design and development of innovative educational technology products. She is a two-time CODiE Award recipient for product design. As co-founder and senior vice president for outcomes and strategy, she focuses on helping Civitas Learning's partner community leverage signals from their data, take informed action, and achieve measurable student success results.

GRACE GEE directs product management at Civitas Learning to help institutions deploy statistically rigorous measurement to understand program, practice, and policy efficacy. With experience as a product analyst, data scientist, and engineer, she specializes in data analytics, algorithms, machine learning, and AI. She received a master's in electrical engineering from Stanford University.

## Planning for Higher Education

Society for College and University Planning

[www.scup.org](http://www.scup.org)

© 2017 by the Society for College and University Planning

All rights reserved. Published 2017.

ISSN 0736-0983

Indexed in the Current Index to Journals in Education (ERIC), Higher Education Abstracts, and Contents Pages in Education.

Also available from ProQuest Information and Learning, 789 E. Eisenhower Parkway, P.O. Box 1346, Ann Arbor, Michigan 48108.

### ABOUT PLANNING FOR HIGHER EDUCATION JOURNAL

*Planning for Higher Education* is a publication of the Society for College and University Planning, an association of professionals devoted to planning at academic institutions. This journal seeks to transmit the knowledge, ideas, research, and experience most likely to advance the practice of higher education planning and policy making. SCUP members receive a complimentary subscription. The society offers both individual and institutional group memberships.

For contributor guidelines, subscription information, or permission to share from the journal, visit [www.scup.org/phe](http://www.scup.org/phe) or contact [managing.editor@scup.org](mailto:managing.editor@scup.org). Correspondence about membership should be sent to [membership@scup.org](mailto:membership@scup.org).

### ADVERTISING IN THE JOURNAL

Thank you to the organizations that have chosen to advertise with SCUP in this publication. Please remember to consider these organizations when you are looking for additional planning services.

Interested in advertising with SCUP? Please visit [www.scup.org/advertise](http://www.scup.org/advertise) or contact [advertise@scup.org](mailto:advertise@scup.org).

### ABOUT THE SOCIETY FOR COLLEGE AND UNIVERSITY PLANNING (SCUP)

The Society for College and University Planning is a community of higher education planning professionals that provides its members with the knowledge and resources to establish and achieve institutional planning goals within the context of best practices and emerging trends. For more information, visit [www.scup.org](http://www.scup.org).

### WHAT IS INTEGRATED PLANNING?

Integrated planning is the linking of vision, priorities, people, and the physical institution in a flexible system of evaluation, decision making and action. It shapes and guides the entire organization as it evolves over time and within its community.