Abstract

This article provides a model for robust student affairs program assessment using diverse data sources, multiple outcomes, propensity score matching, and cost analysis. Contemporary outcome-based assessment in student affairs requires equity-minded approaches paired with methods that support causal inference and actionable results. As a guide for practitioners, we summarize the approach we used to assess a student success program at a large, research-intensive, Hispanic Serving Institution. We describe the team structure, data sources, and analytical framework for assessing the student journey from admissions through graduation. Essential project management templates such as a sample logic model, data tables, and a project calendar are also included. These tools and strategies can be successfully adapted to meet contemporary assessment needs in student affairs at institutions of all types.
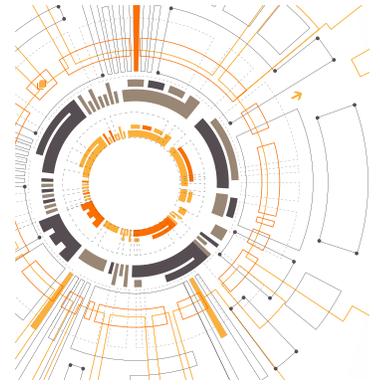
AUTHORS

Kendra Thompson-Dyck, Ph.D.
The University of Arizona

Michelle Sogge, MPA
The University of Arizona

Lucas Schalewski, Ph.D.
Columbia College Chicago

Alexandra Robie, EdD
The University of Arizona

Leveling Up Outcome-Based Assessment: Using Propensity Score Matching and Cost Analysis to Meet Contemporary Assessment Needs

*O*utcome-based assessment is at the forefront of student affairs practice (Biddix, 2018; Henning & Roberts, 2016; Schuh et al., 2016). Robust assessment is increasingly an expectation for student affairs professionals. There remains an ongoing need for high-quality reviews that support causal inferences between student participation and impact to ensure educational advancement for all (Henning et al. 2023, Horst et al., 2022; Montenegro & Jankowski, 2020). Outcome-based assessments are also critical evidence for university accreditation requirements (Gordon et al., 2019; Levy et al., 2018). Further, fiscal challenges in higher education make it pertinent to identify cost-effective strategies that support student success.

Unfortunately, methodological approaches have not kept pace with the changing landscape of assessment needs. Horst and colleagues' (2022) review of student affairs journal articles found program effectiveness claims were often not sufficiently supported in the corresponding methods. They suggest more rigorous assessment training for student affairs professionals. New approaches are needed to demonstrate credible evidence of success program effectiveness and produce useful insights for continuous improvement (Henning & Roberts, 2016).

To meet this need, we offer a recent success program study as an illustrative example. Our analytical approach combines descriptive, inferential, and cost analysis methods to examine program metrics across different domains. We describe our assessment

CORRESPONDENCE

Email
kthompd@arizona.edu

process and methodological choices using concrete examples and templates so that student affairs professionals can apply these tools to their work. We highlight the use of propensity score matching and cost analysis as advanced statistical techniques for assessing impact and cost.

Background

Outcome-based Assessment

Financial assistance is a critical component. This recognizes that socio-economic status is strongly correlated with college enrollment and completion and that need-based aid boosts student retention, while financial stress increases the likelihood of discontinuing college.

Outcome-based assessment within student affairs is a systematic approach to gathering, analyzing, and interpreting data to evaluate a program's effectiveness in meeting its goals and to use those findings to make improvements (Henning & Roberts 2016; Schuh et al., 2016). Outcome-based assessment is essential to ensure that research-based practices used to design new high-impact programs are effective for the specific context where they are implemented (Finney & Buchanan, 2021).

Bresciani Ludvik (2019) used the analogy of a mechanic to illustrate the purpose of outcomes-based assessment. Mechanics run diagnostics on a vehicle's performance indicators to discern if optimal performance is met. If this optimal state is not achieved, performance indicators may point towards further diagnostics or analysis. As assessment professionals may act as mechanics to student learning and success, outcomes-based assessment provides core data points that inform how students can perform optimally (e.g., student learning and success). Bresciani Ludvik (2019) also emphasized the importance of studying individual students' experiences and outcomes to improve overall university performance. Further, efforts to disaggregate performance indicators by student identity (i.e., race, gender, first-generation) support equity-driven practices.

Equity-minded Assessment

Historically, many assessment methods have failed to examine different outcomes, needs, and experiences between groups. Leaders in the field encourage assessment that is rooted in theoretical frameworks such as cultural competency and deploys meaningful data disaggregation (McNair et al., 2020; Montenegro & Jankowski, 2020). Bourke (2017) encourages using data to support action for social justice. Framing data and results in an equity-minded manner can produce novel critical questions to shape campus dialogues. Such frameworks are particularly important for minority serving institutions, including Hispanic Serving Institutions (HSIs). While the federal designation for HSIs is based on thresholds of Hispanic and low-income student enrollment, the concept of what it means for institutions to serve Hispanic students, referred to as "Servingness," embraces a holistic approach (Garcia 2020). A Servingness framework centers culturally affirming, transformative educational experiences for students that lead to positive academic and non-academic outcomes (Garcia 2020). Assessment is key to measuring and improving institutional capacity to serve Hispanic students equitably (Franco & Hernandez, 2018; Garcia et al., 2019). Equity-mindedness should be woven throughout the entire assessment process to the greatest extent possible (Montenegro & Jankowski, 2020).

Example Student Success Program

Program History

In 2008, the university president established the student success program (SSP) to improve educational attainment and upward mobility for low-income residents. Table 1 shows the program logic model which includes high-impact practices such as financial, academic, social, and emotional support associated with student retention (Collings et al., 2014; Nora & Crisp, 2007; Snowden & Hardy, 2012).

Financial assistance is a critical component. This recognizes that socioeconomic status is strongly correlated with college enrollment and completion (Engberg & Wolniak, 2014; Palardy, 2013; Wilbur & Roscigno, 2016) and that need-based aid boosts student retention (Bettinger, 2004; Bettinger, 2015; Millea et al., 2018), while financial stress increases the likelihood of discontinuing college (Britt et al., 2017). At its inception, funding was expansive.

Table 1
Simplified Student Success Program Logic Model

Resources and Inputs	Activities	Outputs	Outcomes
Operational Staff	Invite eligible students	Distributed funds to participants	**Short and Mid-Term:** Increased retention
Funding	Award financial support	Program participation	Increased financial, social, and academic support
Program Staff	Engage program participants in affiliated programs	Advisor meetings	Greater sense of belonging on campus and engagement
Program Partners	Renew participation based on program, financial, and academic requirements	Academic outcomes (GPA 2.0 or greater)	**Long Term:** Increased graduation
		FAFSAs completed	Less debt at graduation
			Better job and graduate school placement
			Upward mobility
			Invest in state and communities

A combination of federal and institutional grant aid covered tuition and fees, housing, food, and books. Gift aid was awarded up to the cost of attendance, resulting in a significant reduction in student loan borrowing for participants. Funding at the time of the program review shifted to a flat amount of $10,000 per student per year, combined with other federal, institutional, or private aid a student received.

Tailored support services are required for students to receive continued funding. First-year students receive one-on-one peer mentoring and workshops on adjusting to college, building academic skills, and making connections. Identity-specific mentoring is offered to students who seek community based on their racial, sexual, or first-generation college student identity, which is a successful strategy supporting historically minoritized groups (Queener & Ford, 2019). Second-year programs emphasize campus involvement, leadership development, and experiential learning while third- and fourth-year options focus on career development, graduate/professional school planning, and preparing for life after college. This comprehensive suite of services was designed with a culturally responsive, asset-based philosophy that recognizes and honors the vast knowledge and skills students bring with them to college (González et al., 2006).

Prior Assessment Efforts

Beginning in 2013, success program staff collected and stored extensive student-level data annually including participant sex, race/ethnicity, high school, major, and a detailed record of support programs usage. These data were regularly used to analyze retention and graduation annually, but a holistic review of the program efficacy or cost-effectiveness was needed.

Analytical and Methodological Approach to the Comprehensive Review

Program Assessment Team Structure

Cochran et al. (2018) recommend an assessment team with internal and external perspectives. This composition can reduce bias while also retaining integral program context and buy-in from those who are close to the program. A multi-person review team can bring diverse personal and professional experiences and identities to the assessment process which supports equity-mindedness.

Our program review team included four on-campus staff members: two internally situated and two externally situated in relation to the program. The internal team provided an in-depth understanding of the program's history, context, and data. The other two team members were from a centralized assessment office for university student affairs and student services. This composition offered nuanced and balanced viewpoints, which we recommend as standard practice for comprehensive assessment whenever possible.

Assessment that Reflects the Student Journey: Beyond Single Outcomes

Henning and Roberts (2016) emphasize that student success programs with broad goals should not be narrowly assessed. Since the cohort-based student success program (SSP) was designed to support access, retention, graduation, and post-graduation outcomes, we designed an assessment to examine program participants' experiences and outcomes throughout the undergraduate journey, starting with admissions and culminating with post-graduation metrics. Since the program provided direct financial aid, campus leaders were interested in understanding the impact of substantial aid and the associated costs and benefits.

> *A multi-person review team can bring diverse personal and professional experiences and identities to the assessment process which supports equity-mindedness.*

We collected and analyzed data from across the institution and student journey (e.g., admissions yield, post-graduation outcomes) as shown in Table 2. We began with traditional descriptive and comparative analyses common in student affairs assessments before adding predictive models and cost-effectiveness analyses to level up our approach (Schuh et al., 2016). This framework could be adapted for teams with varying capacities and time constraints.

Data Collection and Integration

Rather than collect new data, we maximized existing data and leveraged local data steward expertise. This strategy reduces campus survey fatigue that can lead to low response rates (Porter et al., 2004). Montenegro and Jankowski (2020) view this practice as equity-minded since it reduces the data collection burden on students, particularly historically minoritized groups.

However, historical program-level participation datasets are not always clean and readily available. Having designated data stewards who manage the responsible acquisition, cleaning, storing, and use of data and metadata is essential for a comprehensive program review (Plotkin, 2014; Rosenbaum, 2010). Our strategy was to collate siloed departmental data into an expansive dataset across domains and functional areas (e.g., institutional student data records, admissions, financial aid, summer transition programs, fraternity and sorority programs, leadership programs, academic support, campus recreation, student engagement, and career development). This was possible because our team had trusted relationships with other student data stewards across the university through a student data coalition that meets monthly. These colleagues provided essential context for data analysis (e.g., historical trends in data coding/collection) and situated our findings in relation to departmental and university policies and practices.

> *Having designated data stewards who manage the responsible acquisition, cleaning, storing, and use of data and metadata is essential for a comprehensive program review*

First, we made an explicit request to each functional area partner and discussed the larger purpose of the project. We scheduled 'Collegial Check-In' meetings after receiving the data to ensure that our use and interpretations were correct. For example, we worked closely with the financial aid data team to clarify our understanding of financial aid award codes during cleaning and coding, as well as during the reporting phase when we translated data insights into recommendations. These steps were crucial to ensure validity and improve trust in information-sharing among partners before sharing findings with campus leaders. Setting this expectation with partners for collaboration at the beginning of an assessment can increase buy-in and dispel trepidation about assessment findings.

Propensity Score Matching: Determining Differential Impact

Experimental or quasi-experimental approaches can improve the credibility of educational research. These methods account for the counterfactual, the control condition where the program experience or 'treatment' is not administered, to ascertain the average treatment effect of the program among participants (Murnane & Willet, 2010; Horst et al., 2022). It is often not feasible or desirable to utilize a true randomized control trial in higher

Table 2

Assessment Domains and Methods Used

Domain	Analysis	Methodology	Inferences
Admissions	Yield rate of students selected for SSP and comparison groups (e.g., eligible, applied not selected, eligible, did not apply)	Descriptive statistics	Descriptive
Demographic, Academic and Financial Aid	Student background / financial aid comparison participants vs. eligible non-participant peers	Pearson's chi-square or t-test measures of association by subgroup	Descriptive; Comparative
First Year Program *within the affiliated support programs* for SSP Participants	Participation rates, disaggregated by student characteristics (race/eth., first gen, gender). Retention by program choice	Pearson's chi-square measures of association by subgroup	Descriptive; Comparative
Participation in Other Cocurricular Activities	Rates of engagement in other programs, SSP participants vs. eligible non-SSP peers	Pearson's chi-square measures of association by subgroup	Descriptive; Comparative
Retention and Graduation – Descriptive	YR1 Retention, YR2, YR3 Persistence, YR4, YR5, YR6 Graduation rates of SSP participants and non-SSP peers	Rates among pooled cohorts (multiple years)	Descriptive
Retention and Graduation – Propensity Score Matching	YR1 Retention, YR4, YR5, YR6 Grad rates between SSP and statistically matched comparison group within non-SSP peers	PSM to produce rates among pooled cohorts, disaggregated by subgroups (e.g., first gen Latinx female)	Predictive
Drivers of First Year Retention	Factors that predict retention among SSP participants (demographic, financial, academic, participation)	Logistic regression driver analysis	Predictive
Cost-Effectiveness Analysis	Dollars spent per additional student retained and graduated due to program participation	Cost-effectiveness ratio analysis	Cost Analysis
Loan Debt among Graduates	Average loan accumulation of SSP vs. eligible peers	Descriptive statistics	Descriptive; Comparative
Post-Graduation Career and Graduate School	Rates of employment, continuing education of SSP vs. peers	Descriptive statistics	Descriptive; Comparative
University Foundation Fund-Development and Endowments	Foundation funds to support endowments and scholarships	Descriptive statistics	Descriptive

education programs. One method we increasingly recommend for program assessment is Propensity Score Matching, a quasi-experimental approach.

Propensity score matching (PSM) is an alternative to randomized control trials that creates a control condition using statistics to compare outcomes between groups of participants and non-participants (see Harris and Horst's 2016 article for a step-by-step guide). Specifically, PSM generates a balanced comparison group by matching students on variables or covariates that are predictors of self-selecting into the program and the outcome(s) of interest. Propensity score values are generated by a logistic regression model predicting participation and reflect the probability of students participating in a program. These scores are then used to match participants with non-participants. The program's effect is evaluated by comparing the average treatment effects for the participant group in relation to the statistically similar non-participant group. This method has been used to demonstrate the impact of fraternity and sorority membership (Holmes & Bowman, 2017), an engineering grading program (Novak et al., 2016), and honors program participation (Keller & Lacy, 2013) on student success. Propensity score matching can demonstrate the differential impact of the program overall and by subgroup (e.g., first-generation, Black, male students).

Specifically, PSM generates a balanced comparison group by matching students on variables or covariates that are predictors of self-selecting into the program and the outcome(s) of interest.

Program evaluations with a cost component are relatively uncommon in higher education but provide value to decision-makers for contextualizing the return on institutional investments.

The first and most critical step within PSM is to identify the appropriate covariates. A core assumption is that all potential cofounding variables that are related to both the selection into treatment and our intended outcome have been included. In higher education, common covariates include student demographics, academic preparation and achievement, and campus engagement indicators. Critically, these must have a hypothesized relationship with either program participation, the outcome of interest, or both (Harris & Horst, 2016).

The PSM model first conducts a logistic regression analysis predicting participation in treatment from the covariates. These scores are used to generate balanced treatment and comparison groups, which the analyst verifies using post-estimation commands. Once balanced groups are generated, the outcome analysis is performed to demonstrate the difference in average treatment effects between the treated and untreated groups.

In most cases, a one-to-one nearest neighbor matching method is used so that each program participant is matched with a statistically similar non-participant peer. A caliper threshold (e.g., 0.2) can be set by the analyst to limit the absolute distance between propensity scores suitable to be matched to ensure a high-quality comparison group (Austin, 2009; Stuart, 2010; Harris & Horst, 2016). A one-to-one nearest neighbor matching approach with no replacement where non-participants are matched only once has been shown to reduce bias between various PSM techniques (Austin, 2014; Caliendo & Kopeinig, 2008).

In our assessment, we first limited the analytical dataset to program participants as well as non-participants who met the initial eligibility criteria and could have participated but did not. Then, we identified covariates associated with program participation and key outcomes of retention and graduation including student demographics, academic background, financial need, and financial aid award package indicators. A model with 10 covariates using one-to-one nearest neighbor matching, a caliper width of 0.2, and no replacement created an appropriate, balanced comparison group. We then compared the average treatment effects of participants to non-participants overall for retention and graduation rates, reported as a percentage point difference. In applying an equity lens to our assessment, we provided supplemental breakouts by subgroups, such as for first-generation, Hispanic, female students, which indicated even greater returns from the program for students from typically marginalized communities or identities. Appendix A shows sample results as an example using synthetic figures.

Cost Analysis

We then conducted a cost analysis to identify and monetize the various inputs required to support the program. Program evaluations with a cost component are relatively uncommon in higher education but provide value to decision-makers for contextualizing the return on institutional investments (Henning & Roberts, 2016). The scope can vary from costs and benefits borne within an organization, across a national program, or for society at large. For example, Levin and Garcia (2018) considered the cost-savings to the taxpayer for investment in community college programs in the state of New York, whereas Bowden and Belfield (2015) examined the Talent Search TRIO programs nationwide, and Walcott et al. (2018) framed undergraduate research initiatives in terms of students' earnings potential in relation to the university's costs.

Cost-effectiveness identifies and monetizes program implementation costs in cases where the benefits are difficult to monetize or where analysts want to compare alternate interventions seeking to impact a similar outcome (Cellini & Kee, 2015). This method distills costs and outcomes into a cost-effectiveness ratio with costs as the numerator and the unit of effectiveness as the denominator such as "dollars per dropout prevented," (Cellini & Kee, 2015, p. 637).

We used a cost-effectiveness framework to identify the costs of retaining and graduating participants using the procedural approach described by Cellini and Kee (2015). In our case, we applied an organizational lens to consider costs and benefits to the university, acknowledging that this does not account for alternative costs and benefits to other stakeholders (e.g., students, program staff, and state).

Conceptually, costs were divided into direct financial aid award costs from program-specific scholarships and grants, and program implementation costs (e.g., staff salaries, employee-related expenses, and materials/training). In consultation with the budget office and

campus partners, we determined that indirect costs such as facility space and services provided by ancillary units should be excluded since those would not be reappropriated or realize any cost savings if the program were to be discontinued. Direct aid and program cost totals across years were combined to provide a total cost of program administration for seven cohorts in their first year of program participation.

Since returning students increase tuition revenue, we provided a conservative adjustment to the costs based on Federal Pell Grant monies. Nearly all participants received a Pell Grant, so continued enrollment among this population translates into tuition dollars from federal grant funding. In consultation with the budget and financial aid offices, we reduced the total costs for the university by the average of the maximum Pell Grant amount across the years of study, multiplied by the number of additionally retained students. Although additional revenue streams through increases in student persistence do occur (i.e., Housing & Residential Life, Bookstore, Athletics), these amounts were inestimable and were not included in the revenue calculation.

An example table in Appendix A shows the high-level reported cost calculations, though more detailed year-by-year breakdowns were included in the full report. The key figure divides the total cost by the number of additionally retained students, determined by the PSM modeling, to obtain a calculated cost per student. Given the sensitivity of aligning a dollar amount with a specific program and intervention for the first time without any comparison program figures, our evaluation team cross-checked our cost model at each stage of the process with the program director, budget office, scholarship and financial aid, and other leadership members before reporting our findings internally.

Audiences will read public or widely shared reports with varying degrees of background and data literacy, necessitating clear deliverables that communicate insights, not just data.

While the cost-effectiveness analysis produced useful and actionable insights, it also risked causing inadvertent sticker shock for those unfamiliar with the level of institutional investment required to administer a success program of this size. The program review team was also apprehensive of conducting this novel assessment on a program that primarily serves low-income, first-generation, students of color. We acknowledge that programs and services designed to attract and retain high-socioeconomic students exist and likely carry similar, or even greater, price tags. Therefore, we recommend that program cost reporting should always be contextualized with other similar programs to reduce sticker shock and prevent the unintentional targeting of cuts to programs that support historically underserved populations.

Communication and Use of Findings

We used a multifaceted, targeted communication plan to ensure insights were used for program improvement (Bourke, 2017). It is critically important to bridge insight-to-action, especially with lengthy data reports (Henning & Roberts, 2016). High-level decision-makers often have limited program-specific knowledge and time to digest nuanced information. Audiences will read public or widely shared reports with varying degrees of background and data literacy, necessitating clear deliverables that communicate insights, not just data. This step reduces the chances of misinterpretation or misuse.

The review team created one public and one confidential report. The public report was widely available to provide transparency (Montenegro & Jankowski, 2020). The confidential version was for university-affiliated members only and included sensitive details on the budget (e.g., salary, fundraising). The report purposefully integrated context-specific divisional and institutional language informed by the strategic plan, mission statements, and metrics that leadership has top of mind (Henning & Roberts, 2016). Importantly, we authored substantive recommendations based on the results (e.g., optimize award allocation with larger awards vs. larger cohort). Data visualization components were added to enhance communication (Evergreen, 2020). We sent tailored email memos to concisely share the most relevant and useful findings for decision-makers with the full reports attached.

To maximize the use of insights, we invited senior leaders, student affairs administrators, budget office representatives, and program staff members to a virtual presentation and debrief. The goal was to talk through the assessment process, highlight key recommendations, answer questions, and discuss the recommendations with the added context and perspectives provided by the audience. We distributed the confidential report two weeks in advance.

Leveraging key partnerships and engaging in thoughtful, strategic communication of assessment findings were key to maximizing the use of data-informed insights.

These communication strategies facilitated productive discussion by decision-makers on the report's evidence and aligned recommendations. Following those conversations, a debrief meeting with the program review team occurred where nearly all our recommendations were adopted in practice. Table 3 is the model that was used during this meeting, which ties each recommendation to a policy and practice plan for action.

Table 3

Template Model for Demonstrating Use of Assessment Findings

Recommendation	Additional Information	Proposed Plan
Restate the evidence-based recommendation here.	Space to provide additional institutional context and information to broaden understandings.	Identify the actions taken based on the recommendation moving forward.

Project Timetable

Considerable time and staff resources are required to execute a comprehensive review successfully. In our case, we had the advantage of clean historical data and an established collegial network of data stewards. In this best-case scenario, the project unfolded over 12 months. A detailed timetable located in Appendix B provides the month-by-month breakdown of activities which may be useful for replication or modification based on institutional needs and staff capacity.

Limitations

As with all assessment projects, there are limitations. In most cases, mixed-methods approaches are ideal to support equity-mindedness and integrate student voices (Henning et al. 2023). Due to time, resources, and decision-maker priorities, we used exclusively quantitative methods. Future comprehensive program reviews are expected to incorporate learning assessment as well as qualitative data. We recommend aligning methodological choices with institutional priorities, integrating equity-centered strategies in data collection and reporting, and using mixed methods when time and resources allow.

Conclusion

In this article, we reviewed the process used to execute a comprehensive outcome-based assessment project on a long-standing student success program at a large, Research I, Hispanic Serving Institution. We detailed various project management and methodological considerations to provide fellow assessment professionals with a valuable roadmap to replicate similar work on their campuses. Leveraging key partnerships and engaging in thoughtful, strategic communication of assessment findings were key to maximizing the use of data-informed insights. We support Henning and Robert's (2016) claim that assessment professionals are not passive evaluators, but agents charged to work with stakeholders to facilitate planning and action. Student affairs and higher education stakeholders will continue to require data driven causal inferences to determine the impact of student success program participation. Assessment methodological approaches should continue to evolve to meet the profession's current and future needs.

References

Austin, P. C. (2009). Some methods of propensity-score matching had superior performance to others: Results of an empirical investigation and Monte Carlo simulations. *Biometrical Journal*, *51*(1), 171-184. https://doi.org/10.1002/bimj.200810488

Austin, P. C. (2014). A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine*, *33*(6), 1057-1069. https://doi.org/10.1002/sim.6004

Bettinger, E. (2004). How financial aid affects persistence. In C. Hoxby (Ed.), College choices: The economics of where to go, when to go, and how to pay for it (pp. 207-238). University of Chicago Press. https://www.nber.org/papers/w10242

Bettinger, E. (2015). Need-based aid and college persistence: The effects of the Ohio College Opportunity Grant. *Educational Evaluation and Policy Analysis*, *37*(1), 102S-119S. https://doi.org/10.3102/0162373715576072

Biddix, J. P. (2018). Research methods and applications for student affairs. Jossey-Bass.

Bourke, B. (2017). Advancing towards social justice via student affairs inquiry. *Journal of Student Affairs Inquiry, Improvement, and Impact*, *3*(1), 1-18. https://doi.org/10.18060/27837

Bowden, A. B., & Belfield, C. (2015). Evaluating the Talent Search TRIO program: A benefit-cost analysis and cost-effectiveness analysis. *Journal of Benefit-Cost Analysis*, *6*(3), 572-602. https://doi.org/10.1017/bca.2015.48

Bresciani Ludvik, M. J. (2019). What makes a performance indicator an equity-driven, high performance indicator? *Assessment Update*, *31*(2), 1-2, 15-16. https://doi.org/10.1002/au.30163

Britt, S. L., Ammerman, D. A., Barrett, S. F., & Jones, S. (2017). Student loans, financial stress, and college student retention. *Journal of Student Financial Aid*, *47*(1), 3. https://eric.ed.gov/?id=EJ1141137

Caliendo, M. & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31-72. https://doi.org/10.1111/j.1467-6419.2007.00527.x

Cellini, S. R. & Kee, J. E. (2015). Cost-effectiveness and cost-benefit analysis. In: Newcomer, K. E, Hatry, H. P., & Wholey, J. S. (Eds.), *Handbook of practical program evaluation* (4th ed., pp. 636-672). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119171386.ch24

Cochran, M. F., Shefman, P. K., & Hettiarachchi, M. M. (2018). Assessing the assessors: Views from the inside and outside. *Journal of Student Affairs Inquiry, Improvement, and Impact*, *4*(1), 1-18. https://doi.org/10.18060/27909

Collings, R., Swanson, V., & Watkins, R. (2014). The impact of peer mentoring on levels of student wellbeing, integration and retention: A controlled comparative evaluation of residential students in UK higher education. *Higher Education*, *68*, 927–942. https://doi.org/10.1007/s10734-014-9752-y

Engberg, M. E. & Wolniak, G. C. (2014). An examination of the moderating effects of the high school socioeconomic context on college enrollment. *The High School Journal*, *97*(4), 240–263. https://doi.org/10.1353/hsj.2014.0004

Evergreen, S. (2020). *Effective data visualization: The right chart for the right data* (2nd edition). Sage Publications.

Finney, S. J., & Buchanan, H. A. (2021). A more efficient path to learning improvement: Using repositories of effectiveness studies to guide evidence-informed programming. *Research & Practice in Assessment*, *16*(1), 36-48. https://eric.ed.gov/?id=EJ1307022

Franco, M. A., & Hernández, S. (2018). Assessing the capacity of Hispanic serving institutions to serve Latinx students: Moving beyond compositional diversity. *New Directions for Institutional Research*, *2018*(177), 57-71. https://doi.org/10.1002/ir.20256

Garcia G. A. (Ed.). (2020). *Hispanic serving institutions (HSIs) in practice: Defining "Servingness" at HSIs*. Information Age Publishing.

Garcia, G. A., Núñez, A. M., & Sansone, V.A. (2019). Toward a multidimensional conceptual framework for understanding "servingness" in Hispanic-serving institutions: A synthesis of the research. *Review of Educational Research*, *89*(5), 745-784. https://doi.org/10.3102/0034654319864591

González, N., Moll, L. C., & Amanti, C. (Eds.). (2006). *Funds of knowledge: Theorizing practices in households, communities, and classrooms*. Routledge.

Gordon, S. R., Shefman, P., Heinrich, B., & Gage, K. (2019). The role of student affairs in regional accreditation: Why and how to be included. *Journal of Student Affairs Inquiry, Improvement, and Impact 5*(1), 1-25. https://doi.org/10.18060/27916

Harris, H. & Horst, S. J. (2016). A brief guide to decisions at each step of the propensity score matching process. *Practical Assessment, Research & Evaluation*, *21*(4), 1-11. https://doi.org/10.7275/yq7r-4820

Henning, G. W., Rice, A., Heiser, C., & Lundquist, A. E. (2023). Equity-centered assessment practices: Survey findings and recommendations. *Research & Practice in Assessment*, *18*(2). https://www.rpajournal.com/dev/wp-content/uploads/2024/03/Equity-centered-Assessment-Practices-RPA.pdf

Henning, G. W., & Roberts, D. (Eds). (2016). *Student affairs assessment: Theory to practice*. Stylus Publishing. https://doi.org/10.4324/9781003447139

Holmes, J. & Bowman, N. (2017). A quasi-experimental analysis of fraternity or sorority membership and college student success. *Journal of College Student Development*, *58*(7), 1018-1034. https://doi.org/10.1353/csd.2017.0081

Horst, J. S., Finney, S. J., Prendergast, C. O., Pope, A. M., & Crewe, M. (2022). The credibility of inferences from program effectiveness studies published in student affairs journals: Potential impact on programming and assessment. *Research & Practice in Assessment*, *16*(2), 17-32. https://eric.ed.gov/?id=EJ1348828

Keller, R. R. & Lacy, M. G. (2013). Propensity score analysis of an honors program's contribution to students' retention and graduation outcomes. *Journal of the National Honors Council*, *14*(2), 73-84. https://files.eric.ed.gov/fulltext/EJ1082022.pdf

Levin, H. M. & Garcia, E. (2018). Accelerating community college graduation rates: A benefit-cost analysis. *The Journal of Higher Education*, *89*(1): 1-27. https://doi.org/10.1080/00221546.2017.1313087

Levy, J., Hess, R., & Thomas, A. (2018). Student affairs assessment and accreditation: History, expectations, and implications. *Journal of Student Affairs Inquiry*, *4*(1), 1-19. https://doi.org/10.18060/27888

McNair, T. B., Bensimon, E. M. & Malcom-Piqueux, L. (2020). *From equity talk to equity walk: Expanding practitioner knowledge for racial justice in higher education*. Jossey Bass.

Millea, M., Wills, R., Elder, A., & Molina, D. (2018). What matters in college student success? Determinants of college retention and graduation rates. *Education*, *138*(4), 309-322. https://aalhe.scholasticahq.com/article/24575-equity-in-assessment-the-grand-challenge_and-exploration-of-the-current-landscape

Montenegro, E., & Jankowski, N. A. (2020). A new decade for assessment: Embedding equity into assessment praxis. National Institute for Learning Outcomes Assessment. https://www.learningoutcomesassessment.org/wp-content/uploads/2020/01/A-NewDecade-for-Assessment.pdf

Murnane, R. J., & Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research*. Oxford University Press.

Nora, A., & Crisp, G. (2007). Mentoring students: Conceptualizing and validating the multidimensions of a support system. *Journal of College Student Retention: Research, Theory & Practice*, *9*(3), 337–356. https://doi.org/10.2190/CS.9.3.e

Novak, H., Paguyo, C., & Siller, T. (2016). Examining the impact of the engineering successful/unsuccessful grading program on student retention: A propensity score analysis. *Journal of College Student Retention*, *18*(1), 83-108. https://doi.org/10.1177/1521025115579674

Palardy, G. J. (2013). High school socioeconomic segregation and student attainment. *American Educational Research Journal*, *50*(4), 714-754. https://doi.org/10.3102/0002831213481240

Plotkin, D. (2014). Data stewardship: An actionable guide to effective data management and data governance. Elsevier Inc. https://doi.org/10.1016/C2012-0-07057-3

Porter, S. R., Whitcomb, M. E., & Weitzer, W. H. (2004). Multiple surveys of students and survey fatigue. *New Directions for Institutional Research*, *2004*(121), 63-73. https://doi.org/10.1002/ir.101

Queener, J. E., & Ford, B. A. (2019). Culturally responsive mentoring programs: Impacting retention/graduation rates of African American males attending predominately White institutions. In *Overcoming challenges and creating opportunity for African American male students* (pp. 120-132). IGI Global.

Rosenbaum, S. (2010). Data governance and stewardship: Designing data stewardship entities and advancing data access. *Health Services Research*, *45*(5p2), 1442-1445. https://doi.org/10.1111/j.1475-6773.2010.01140.x

Schuh, J. H., Biddix, J. P., Dean, L. A., & Kinzie, J. (2016). *Assessment in student affairs*. (2nd edition). Jossey-Bass.

Snowden, M. & Hardy, T. (2012). Peer mentorship and positive effects on student mentor and mentee retention and academic success. *Widening Participation and Lifelong Learning*, *14*, 76–92. https://doi.org/10.5456/WPLL.14.S.76

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, *25*(1), 1-21. https://doi.org/10.1214/09-STS313

Walcott, R. L., Corso, P. S., Rodenbusch, S. E., & Dolan, E. L. (2018). Benefit–cost analysis of undergraduate education programs: An example analysis of the freshman research initiative. *CBE—Life Sciences Education*, *17*(1). https://doi.org/10.1187/cbe.17-06-0114

Wilbur, T. G., & Roscigno, V. J. (2016). First-generation disadvantage and college enrollment/completion. *Socius*, *2*. https://doi.org/10.1177/2378023116664351

Appendix A
Example Data Tables

The following data are artificial and are examples of how results may be shared. Actual results are not presented given data restrictions.

Example Table of PSM Results with Artificial Data

First Year Student Retention

Cohorts 2013-2019	Retention YR1-YR2	% Point Diff.	# of Additional Students
Participants (n=2,005)	87.8% (n=1,760)	+9.1***	+182
Matched Non-Participants (n=2,005)	78.7% (n=1,578)		

***p<.000

First Year Retention by Subgroup

Cohort 2013-2019	First Gen White Female	First Gen White Male	First Gen Hispanic Female	First Gen Hispanic Male
Participants	86.3%	77.4%	86.1%	88.1%
Matched Non-Participants	74.0%	74.8%	78.0%	74.7%
% Point Difference	12.2***	2.6	8.1***	13.4***
Participant (n=)	200	112	780	402
Non-Participant (n=)	200	112	780	402

***p<.000, **p<.01, *p<.05

Example Table of Cost-Effectiveness Ratio Calculation for Program Impact on YR1 Retention

Direct Aid Financial Aid in Cohort Entry Year	$ Total A
Program Costs 2013-2019	$ Total B
Total YR1 Costs 2013-2019	$ Total C = (A+B)
Additional Students Retained due to Participation	
Participants Retained (87.8%)	1,760
Matched Peers Retained (78.7%)	1,578
Additional Participants Retained (9.1%)	182
Federal Pell Grant $ Gain from Additionally Retained Students	
Avg. Max Pell Grant 2014-2020 due to Retention ($ Avg. X 182)	$ Total D
Calculated Costs Per Student	
Total YR1 Costs Minus Federal Pell Grant $ Gained	**$ Total E = (C – D)**
Cost per Additional Participant Retained (182)	**$ Per Student (E / 182)**

Cost-Effectiveness Ratio = Costs / Units of Effectiveness

Dollar per additional student retained = $ Total Costs E / 182

**Appendix B
Program Review Timeline**

	Assessment Activity Timetable for Comprehensive Review
October	Form assessment team. Develop early framework for analyses, context, and data needs.
November	Hold information gathering meetings with campus stakeholders (Financial Aid, Enrollment Management, Program Directors, Foundation) to determine context, eligibility criteria for invitation, logic model, historical changes, cost framework.
December	Provide detailed financial aid data request and justification.
January	Conduct preliminary analyses. Collegial check-in meetings on financial aid data context, use, implications.
February	Conduct analyses. Solicit data tables from partners in enrollment, career development, and foundation.
March	Preview preliminary data analyses and insights with all contributing data stewards and stakeholder departments.
April	Develop recommendations section in consultation with Program Director.
May	Present written final report and PowerPoint presentation to leadership committee including Provost.
Summer	Attend follow-up meetings with leaders to clarify understanding of insights and provide additional data points, as requested. Final report without cost data distributed to campus.
Fall	Leadership committee subgroup presentation of program changes that were adopted based on the comprehensive review, with modifications tied specifically to recommendations in the report.