

7

This chapter discusses the importance and common challenges of evaluating community college student success efforts; it includes a broad-based framework for carrying out effective evaluations.

Understanding the Effectiveness and Impact of Student Success Interventions on Campus

Bruce E. McComb, Jan W. Lyddon

How do we know an intervention or action is *high impact* or even *promising*? On what basis will an institution make a defensible judgment about the value of its efforts? It is more than guesswork and more than just measuring a series of numbers, but it is all too often left as an afterthought or relegated to the work of a (likely already overburdened) institutional research or institutional effectiveness staff to provide evidence to the effectiveness of student success initiatives. Perhaps worse yet, a practice might be considered effective because it carries a halo effect of having sponsorship from an individual or organization with considerable clout or power. If evaluation is not done well, an institution may waste resources on ineffective efforts, may fail to use formative information to improve them, and/or may fail to scale up effective ones (Berlin, 2014).

An evaluation's importance is not only for the benefit of institutional decision making. In a climate of decreasing state and federal support for higher education and increasing concerns about college affordability, community colleges are facing growing pressure to demonstrate the value of programs and interventions and justify business models with empirical evidence related to student success outcomes (e.g., completion rates, term-to-term and fall-to-fall retention, course success, and gainful employment after completion of college) (Berlin, 2014; Hess & Little, 2015). Institutions often adopt promising practices that have been developed by other colleges. Internally, community college administrators, faculty, and staff need to understand to what extent and under what conditions interventions add value and are appropriate for the college. At the same time, funders, accreditors, and those providing grants (e.g., government agencies and private

foundations) are increasingly requiring the evaluation of interventions as a condition of operations.

Over the past decade, there has been a shift in interest in improving student success, not just offering access. A leading nonprofit organization in student success improvement efforts is Achieving the Dream (ATD), which is now the largest nongovernmental reform movement in higher education (www.achievingthedream.org). ATD provides participating community colleges with a variety of services including the support of two coaches, a leadership coach and a data coach. As data coaches, we help institutions develop a “culture of evidence and inquiry” that in turn supports data-informed decision making to increase student success. A key component of the culture of evidence is evaluation, demonstrating that interventions are on track to accomplish the expected outcomes (called formative evaluation) or have achieved them (summative evaluation).

This chapter draws on our combined professional experience during the past decade of coaching more than 20 institutions. There are a number of common challenges that affect institutions working to evaluate their interventions as well as a number of promising practices to address them. In the following section, we discuss some of the most common challenges we have identified from both our experience and the literature. We then offer a broad-based guide to carrying out effective evaluation of student success efforts.

Common Challenges

One of the most common challenges that affects the selection and design of interventions and their evaluation, in our experience, is the “propensity to jump quickly on a solution before fully understanding the exact problem to be solved” (Bryk, Gomez, Grunow, & LeMahieu, 2015, p. 24). College faculty and staff may be overconfident that an effort will have an expected outcome. At the core of this, we have seen numerous behaviors that are likely to interfere with good evaluation, including but not limited to an inadequate or unclear definition of the problem, a failure to develop a solid theory of change¹ to inform the selection and development of the intervention, and/or delaying the development of an evaluation plan until after the intervention is implemented.

Another challenge is that institutions sometimes focus only on large outcome measures, also known as lagging indicators (Gendron & Traub, 2015). Common lagging indicators include course success, retention², or completion of degrees or certificates. Although important, these outcomes are difficult to predict as they are related to student characteristics as well as numerous programs and practices already taking place on campus. Lagging indicators are like looking in the rearview mirror of a car; you can see where you have been, but you may not be able to see where you are going. Leading indicators on the other hand, are metrics that are predictive of what

results are likely to be achieved and yet are often overlooked or given less attention. Leading indicators include metrics, such as attending classes, attending tutoring sessions, completing assignments on time, or participating in study groups.

Some metrics like course success can be both a leading indicator and a lagging indicator. For instance, course success is predictive of retention and completion. At the same time, as a lagging indicator, course success can often be predicted by leading indicators, such as class attendance, use of tutoring, or timely completion of assignments. By delaying evaluation to the end of the term, year, or several years, colleges are unable to understand the impacts of an intervention on leading indicators and are not able to make corrections during the course of the intervention. For evaluation purposes, exclusively using a lagging indicator, such as graduation rates or even retention rates to measure the impact of interventions such as new student orientation, is difficult at best. In the case of orientation, the extent to which students remain enrolled through the middle of the term as well as specific short-term behaviors like having a college email account set up before classes begin are likely better predictors of graduation or retention. In short, community colleges need to identify outcome measures that can be closely aligned with a particular intervention.

Community college faculty, staff, and administrators are increasingly doing more with less and are more often than not overburdened with responsibilities. This whirlwind of commitments can easily pull focus away from institutional student success goals. However, with the appropriate formative evaluation of data regarding leading indicators that are reasonably predictive of student success, colleges can maintain an ongoing focus on the intervention and make needed adjustments to maximize the impact on student success. As such, identifying and monitoring leading indicators and appropriate lagging indicators are key components of building a solid evaluation.

Some community colleges focus *only* on inputs or outputs without defining outcomes. However, community colleges engage in student success efforts in order to bring about some kind of positive change in one or more student outcomes (e.g., course success, retention/persistence, and degree completion). Outcomes must be focused, clearly defined, and measured using a methodologically rigorous research design. In our experience, colleges seldom use the gold standard for evaluation, described by MDRC President Gordon Berlin (2014) as experimental research designs that use random assignment to a treatment (i.e., a group that receives the intervention) and a control group. There are numerous intervening or extraneous variables that may have important relationships with student success outcomes (e.g., attending full time and placement in developmental education classes). Identifying as many of these variables as possible is essential in evaluating the intervention. The use of random assignment serves to control for these variables and provide a more valid measurement of the outcome.

However, using random assignment is difficult and may not always be required. Difficulties/issues in using random assignment to groups may include a random assignment process, lack of experience in using experimental design, time and effort required, sample size requirements, control of external variables, philosophical/ethical issues of random assignment, and a lack of understanding of the value of experimental research (Berlin, 2014). As an alternative, community colleges typically use a quasiexperimental design that does not require random assignment. The strength of the quasiexperimental design depends on the extent to which the comparison group is similar to the group participating in the intervention. A well-selected comparison group is a means of controlling for known extraneous variables. It is important to identify the comparison group at the onset of the intervention. Unfortunately, colleges may fail to adequately select a comparison group or so do too late. When a comparison group is not assigned prior to implementation, the college cannot ensure robust data collection on the comparison group.

Communication of the evaluation results, both formative and summative, is critically important and yet is often a challenge for evaluation teams. We advise taking care to share information with all stakeholders throughout the implementation and evaluation process and providing different types and amounts of information with different groups/stakeholders. For instance, individuals who are directly involved in implementing the intervention need detailed formative results in sufficient time to take corrective action, whereas administrators or faculty who are not directly involved may require only brief interim reports and/or a final summative report. Developing presentations and reports may require resources that are scarce (time, talent, and materials), and so communication of the evaluation is often limited to mandatory reporting. Fortunately, these challenges are not insurmountable.

Design the Evaluation While Designing the Intervention

The importance of designing the evaluation simultaneously with developing an intervention cannot be overstated. Doing so helps institutions to avoid jumping to a solution before a problem is well defined, begins with the end in mind, and gives more careful consideration to obtaining current and relevant data for both the formative and summative evaluations. Conducted while an intervention is ongoing, formative evaluation provides the feedback necessary to guide and improve the intervention. When the intervention has stabilized, summative evaluation serves to measure the overall value of the intervention (Rodriguez-Campos, 2005). Summative evaluation answers questions such as: Were the anticipated results achieved? Were the outcomes worth the cost and effort to achieve and sustain the results? Leading indicators, described earlier, are critical to the formative evaluation whereas lagging indicators are more useful for the summative evaluation.

Evaluation Models—No One Size Fits All

Various methods and models exist for evaluation, and each varies somewhat in complexity and structure. However, not every method fits every institution's needs. The experience and preferences of those involved in the design and evaluation process may lead some institutions to use one method or model whereas others may prefer to use a different, but similarly effective method. As such, we briefly describe three models and provide references for further exploration.

Prominent among evaluation models is the logic model. The logic model can serve as a planning tool for an intervention strategy, helps organize the planning and the evaluation functions, and can be useful to the institution in identifying and collecting data needed for monitoring and evaluation (W.K. Kellogg Foundation, 2006). Logic models typically include four primary sections: inputs, activities, outputs, and outcomes. Data are arranged from left to right in a table often displayed on a single page with accompanying detail on separate pages. A logic model may show the interaction of these sections as a visual flowchart, which captures the essence of the intervention. In other cases, the model simply lists items under each section. The W. K. Kellogg Foundation (2006) *Logic Model Development Guide* does an excellent job of detailing how to develop logic models and provides templates, checklists, and examples. Another useful resource is Rincones-Gomez (2009) who explains and provides examples of logic models.

Another model, improvement science, is an integral component of the Carnegie Foundation's highly successful Community Colleges Pathways developmental math initiative. The improvement science model asks three deceptively simple questions to guide the development of *both* the intervention and its evaluation (Bryk et al., 2015, p. 114). The questions include: (1) "What specifically are we trying to accomplish? (2) What change might we introduce and why? (3) How will we know that a change is actually an improvement?" Langley et al. (2009, p. 5) note, "These three questions are combined with the Plan-Do-Study-Act (PDSA) Cycle to form the basis of the model."

The use of the PDSA Cycle builds in formative evaluation, ensuring that evaluation is an ongoing process and not a one-time event or action. A strong feature of the work in adopting an intervention under the guidance of the Carnegie Foundation's networked improvement communities' (NICs) approach is consistent use of the PDSA cycle within the classroom throughout the term. Conditions in one institution differ from others, so the formative evaluation (which is discussed in more detail later) using a PDSA approach is essential (Bryk et al., 2015).

A third model for evaluation is Friedman's Results Accountability Method that "starts with ends and works backward, step by step, to means" (Friedman, 2005, p. 12). The method combines intervention design with the evaluation design. The model begins with an understanding of the

conditions the college wants and then identifies indicators, (i.e., how to measure those conditions) baseline outcome(s), and what success looks like if the college performs better than the baseline. The method also identifies strategies and performance measures at this stage. Performance measures in the Results Accountability Method are derived from three primary questions that address both the intervention and its evaluation: (a) How much did we do, (b) How well did we do it, and (c) Is anyone better off?

Designing the Evaluation

It is worth repeating that it is inefficient and sometimes counterproductive to separate evaluation from the intervention; practitioners implementing the intervention need formative information to be able to make adjustments and improvements throughout implementation. We advise that staff and faculty who are directly involved with the intervention be included as part of the evaluation design team. Moreover, the method of data collection should be as clear and simple as possible. Data that have already collected by the institution should be used if and when appropriate. When designing the intervention and its evaluation, Bryk et al. (2015) recommend:

Improvement requires measures that (a) operationalize a working theory of improvement; (b) are specific to the work processes that are the object of change; (c) have formative value signaling subsequent action useful to consider; (d) are framed in a language that is meaningful to those engaged in the work; (e) produce data accessible in a timely manner; and (f) are embedded in social routines that secure the trust and openness necessary to sustain meaningful change efforts. (p. 101)

Although the evaluation and intervention plans may be interwoven, it is important to ensure that the summative evaluation is conducted independent from the intervention. More specifically, we strongly advise that someone with little or no direct stake in the outcome lead the summative evaluation. Becoming too engaged in or invested in the intervention may compromise the integrity of the summative evaluation. For some projects, particularly those supported by external grant funding, an independent evaluator is essential (i.e., someone outside the institution or at least from outside the unit who is implementing the intervention).

Formative evaluation is a different story. Formative evaluation should occur with greater frequency—perhaps as often as weekly—so as to allow for adjustments in the intervention in a timely way. It is probably best, and in some cases essential, to have those involved in the intervention responsible for collecting and analyzing data for the formative evaluation, as they are in a position to make adjustments to the intervention as needed. However, it is also good practice to keep track of and report the results and changes made to a person or group independent of the work.

Many Achieving the Dream colleges create data teams composed of faculty and staff members who are comfortable with data-informed decision making, including one or more members with statistical analysis capabilities. This team provides assistance to the institutional research or institutional effectiveness office in a variety of ways including assisting intervention teams with the evaluation design. The team or selected members can provide an independent perspective on interpreting the evaluation data and provide feedback to the intervention team.

Steps in Creating an Evaluation Model

We have presented three evaluation models from the literature. Regardless of the method an institution uses, there are essential steps or components for developing and evaluating an intervention. This model supports most evaluation methods. The first step is for the intervention team to identify a theory of change that includes outlining the situation or problem to be resolved. It can be useful to develop a clear picture of the future reality to be achieved. For instance, a college may seek to increase fall-to-fall retention. Results that are desired from the intervention should be identified in terms of observable and measurable outcomes (e.g., increase retention by 5 percentage points within 3 years). The team should work together to develop a clear description of *what* has to change to move from the current to the future reality. The team should then determine *how* to bring about the change by identifying one or more strategies or practices that are expected to generate the change and achieve the desired outcomes. An example of practices that may positively influence retention is mandatory advising for new students combined with the use of an early alert system. Large-scale initiatives may use systems improvement maps and driver diagrams in order to determine the changes needed. Excellent examples of improvement maps and driver diagrams can be found in Bryk et al. (2015). It is important to identify the basic premises the strategies are based on (i.e., research findings and theoretical grounding) and why the college expects the strategies to be effective.

Next, the team should work together to design the implementation plan for the new mandatory advising and early alert initiative. For the purposes of this chapter we emphasize the evaluation aspects of the work. At this point, the team must make a decision whether to use an experimental design with random assignment and a control group or a quasiexperimental design using a comparison group. Within this, the team should identify and prioritize specific performance metrics based on the work done in step 1. Narrowing them to the right balance or combination of what is essential and available is vitally important. Qualitative and quantitative data should be collected and available *from the start of the implementation of the intervention* to support the formative evaluation. It is possible that as the implementation proceeds, some formative data will not be as helpful as originally

thought and new formative data may need to be collected. As described later, metrics should include both quantitative and qualitative elements and broadly divide into outputs and outcomes.

The team must also develop evaluation questions and identify how the qualitative and quantitative performance metrics will be used to answer each of the questions. The team should consider whether it would need additional metrics to fully answer the questions. According to Friedman (2005), evaluation questions may include (a) How much did we do? (b) How well did we do it? (c) Regarding effort: How hard did we try and what was our quality of effort? (d) Regarding effect: What change did we produce? What modifications are necessary? Is it time to scale the intervention?

The final step is to develop an evaluation plan in tabular (or another visual) form. The team should list the evaluation questions on the left side and then in succeeding columns and rows specify the following information:

1. The data and information necessary to answer the evaluation questions and to support formative evaluation
2. Sources and procedures for collecting the data and information
3. Analysis methods
4. Responsibilities for carrying out the formative and summative evaluations
5. Due dates (i.e., how frequently the formative evaluation will occur and when the summative evaluation will take place)

At the bottom of the form specify the comparison group(s), that is, the group(s) with which the students in the intervention will be compared to understand whether change has occurred.

Outputs and Outcomes

Both outputs and outcomes mentioned in these steps are essential metrics for an evaluation. Outputs are leading indicators that the intervention will have some effect. The targeted level of activity or reach to be achieved for the intervention to work properly should be specified so the evaluation plan can examine target to actual performance for outputs. Outputs are units of service (Friedman, 2005) such as activity counts (e.g., number of participants or instructors trained) and can provide guidance in answering the question “How much did we do?” Outputs may also be measures of reach, such as 30% of advisors were trained or 40% of math students used the math lab, and can help answer the question, “How well did we do it?”

Similar to the causal chain presented by Brinkerhoff (2006), performance metrics can be thought of as a causal chain; outputs lead to awareness, satisfaction, knowledge and/or skill development, which in turn lead

to observable and measurable change in behavior that influences student success outcomes/impacts. Measures of awareness, satisfaction, knowledge and skills development, and observable and measurable behaviors (e.g., class attendance, attending tutoring, seeing an advisor, and grades) can be very useful for the team's formative evaluation. Ongoing monitoring of activities and behaviors should be used to trigger immediate corrective action during implementation.

In terms of summative evaluation, outcomes/impacts may include measures such as course completion, successful achievement of learning objectives, semester or year retention, degree completion, or gainful employment. Outcomes/impacts can be thought of as a second causal chain where short-term outcomes lead or influence longer term outcomes (referred to as impacts). For instance, achieving learning objectives may lead to a short-term outcome (e.g., course success), which may lead to a longer term outcome such as retention, which in turn influences completion of a certificate or degree, which may lead to gainful employment. Outcomes or impacts such as being a better citizen or having a well-rounded education could also be considered although these may be harder to measure. It is important to note that not every intervention will or should be evaluated on all of the outcomes listed here. Rather, some interventions will be expected to affect only part of the causal chain.

Analysis Methods

Descriptive statistics are the first step in analyzing quantitative data collected for an evaluation. Statistics such as percentages and means describe and compare the characteristics, behaviors, and/or outcomes of the participants (both the students receiving the intervention and the control/comparison group). In many cases, descriptive statistics suffice for evaluating an intervention. Inferential statistics (e.g., *t*-test and ANOVA) can be used to assess whether differences between groups are practically or statistically significant and were unlikely to occur by chance. More advanced techniques such as regression may also be useful in cases where students participated in multiple interventions or to control for other variables related to the outcome. Qualitative data can be extremely valuable in providing a rich description of the students' perceptions and behaviors related to the intervention. Analysis of qualitative data may involve identifying patterns or themes in the data to answer "how" or "why" questions.

We suggest including someone with a solid statistical background (e.g., research analyst or faculty member) as part of the team or in an advisory role. This person will be able to suggest types of data that need to be collected to answer the evaluation questions and provide formative feedback. An analyst will also be able to suggest when further, more complex statistical analyses are needed. We caution, however, to always keep the analysis methods as simple as possible to avoid costly and complex evaluations

in relation to the scope and magnitude of the intervention. This might be thought of as a trade-off between the “good enough” and the perfect scientific evidence. Realistically, “good enough” may be appropriate for many interventions.

Often-Overlooked Evaluation Topics

Although the impact of the intervention on student success should be the primary consideration in evaluation, return on investment (ROI), sustainability, and scalability are evaluation issues that warrant mention. Return on investment requires consideration of the value of the effort, usually measured in monetary terms, compared with the savings or the actual revenue gain resulting from student retention. No community college has unlimited resources, and if the cost of an intervention greatly outstrips the value of the end results, it may need to be redesigned or reconsidered altogether. There are also other “value” metrics, including the value of positive relations with the community and long-term benefits to students’ well-being. In our experience, ROI is, however, an often-overlooked consideration in developing and implementing evaluations. Additional information on return on investments in student success along with an Excel-based calculator can be found at: <http://www.jff.org/publications/calculating-cost-return-investments-student-success>.

Related to ROI are sustainability and scalability. Grant-funded interventions generally require institutions to develop a plan to continue the actions after the grant funding ends, but even for internally funded interventions, this practice is an important consideration in evaluation. Can the institution continue the quantity and quality of effort over the longer term? If an intervention is reliant on the good will of volunteers, efforts might not be sustainable. On the other hand, if an intervention can be embedded as part of a continuing institutional process or program, it is more likely to be sustained.

Finally, there is scalability. Offering valuable services to a small percentage of the student population is unlikely to “move the needle” in student success metrics. On the other hand, scaling up from a small pilot group to the entire population can be challenging, especially when the demographics of the pilot program are not representative of the larger population. Examples of this are abundant in other fields, notably medicine where the effectiveness of a medication is found to be different for different groups (e.g., females and males). Likewise, the impact(s) of an intervention on a group of students who are motivated enough to volunteer to participate may be very different results when the intervention is expanded to a broader group of students (e.g., issue of selection bias). Moreover, an intervention that is scaled up may not have any overall significant effect on success outcomes. An intervention that fails to demonstrate significant impacts should not necessarily be abandoned, however.

Instead, perform additional analyses to identify whether the intervention was more or less effective for particular subgroups (e.g., first generation and racial/ethnic groups) that may identify the value in a more targeted approach. Additional information on sustainability and scaling of interventions is available in a guide to scaling community college interventions at <http://www.publicagenda.org/files/CuttingEdge2.pdf>.

Final Thoughts

Evaluation is an essential part of student success efforts and should to be designed at the same time teams design the intervention(s) aiming to improve student success. Evaluations, when well done, can assist community colleges in staying on track of intervention goals/objectives through consistent and frequent formative evaluation and improvement. A well-designed summative evaluation can provide the basis for decisions to continue or scale up interventions by providing data regarding particular elements of the intervention that had the desired impact. The combination of formative and summative evaluations is a series of value judgments based on well-defined, agreed-upon, defensible criteria that can set an institution's interventions on a path toward realization of improved student success outcomes.

Ultimately, we feel that it is important to find and follow a method of evaluation that fits the needs of the college, is not more complex than needed to answer the evaluation questions, and does not present unrealistic data collection or analysis challenges. Moreover, development and follow-through on an evaluation plan with timelines, shared responsibility, and relevant data elements are vitally important.

Notes

1. Interventions or programs designed to improve student success are based on a theory of change, that is, if the institution does X it will result in Y effect. Further detail is provided in the later part of the chapter.

2. In this chapter retention means term-to-term retention and/or year-to-year retention.

References

- Berlin, G. (2014). *Impact on a large scale: The importance of evidence*. New York, NY: MDRC. Retrieved from http://www.mdrc.org/sites/default/files/Impact%20on%20a%20Large%20Scale%20_web.pdf
- Brinkerhoff, R. O. (2006). *Telling training's story: Evaluation made simple, credible, and effective*. San Francisco, CA: Berrett-Koehler.
- Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Cambridge, MA: Harvard Education Press.
- Friedman, M. (2005). *Trying hard is not good enough*. Victoria, BC, Canada: Trafford Press.

- Gendron, S. A., & Traub, S. (2015). *Leading and lagging indicators to inform systemwide change*. Rexford, NY: International Center for Leadership Education. Retrieved from <http://leadered.com/pdf/ICLELeadingLaggingIndicatorsSB.pdf>
- Hess, F. M., & Little, B. (2015). *Moneyball for education: Using data, evidence, and evaluation to improve federal education policy*. Washington, DC: American Enterprise Institute. Retrieved from <https://www.aei.org/wp-content/uploads/2015/03/Moneyball-for-education.pdf>
- Langley, G. J., Moen, R. D., Nolan, K. M., Nolan, T. W., Norman, C. L., & Provost, L. P. (2009). *The improvement guide: A practical approach to enhancing organizational performance*. San Francisco, CA: Jossey-Bass.
- Rincones-Gomez, R. J. (2009). *Evaluating student success interventions: Principles and practices of student success*. Silver Spring, MD: Achieving the Dream. Retrieved from <http://www.achievingthedream.org/resource/177/evaluating-student-success-interventions-principles-and-practices-of-student-success>
- Rodriguez-Campos, L. (2005). *Collaborative evaluations: A step-by-step model for the evaluator*. Tamarac, FL: Lumina Press.
- W. K. Kellogg Foundation. (2006). *Logic model development guide*. Battle, Creek, MI: Author. Retrieved from <https://www.wkkf.org/resource-directory/resource/2006/02/wk-kellogg-foundation-logic-model-development-guide>

BRUCE E. MCCOMB and JAN W. LYDDON, PhD, are principals of Organizational Effectiveness Consultants (OEC). They are both Data Coaches with Achieving the Dream.