

Establishing Legitimacy among STEM Intervention Programs: The Need for Evaluation

Casey E. George-Jackson and Blanca Rincon
University of Illinois at Urbana-Champaign

ABSTRACT

Intervention programs designed to improve undergraduate students' participation and success in the Science, Technology, Engineering, and Mathematics (STEM) fields exist in colleges and universities throughout the United States. These programs seek to attract and retain traditionally underrepresented students including women and minorities through supplemental services including tutoring, mentoring, research, and social support networks. This study examines the extent to which such programs conduct evaluations of their program or services, and how evaluations impact the legitimacy of intervention programs. Drawing upon interview data with 55 program directors and administrators at ten large, public, research universities, the findings suggest that many programs rely on anecdotal information to inform program changes due to financial and human resource constraints. The results also suggest a relationship between programs that do conduct evaluations and their long-term sustainability, ability to garner support from upper-level administrators and secure recurring funding, all of which impact the program's legitimacy. These findings have significant implications given the current economic climate, institutional cutbacks and reduction of services, and the need to increase the number of STEM degrees granted, particularly to underrepresented students. Recommendations, such as partnering with other departments and colleges to conduct evaluations, are offered.

Keywords: STEM, Evaluation, Student Retention

INTRODUCTION

Across the nation, there is a persistent call for increasing the number of science and math degrees awarded at the postsecondary level, for reasons of economic stability, international competitiveness, and scientific innovation. For groups that have been traditionally underrepresented in the Science, Technology, Engineering, and Mathematics (STEM) fields, including women, racial and ethnic minority students, and low-income students, success in the STEM fields is also related to issue of educational equity and social justice.

Intervention programs in the STEM fields that seek to recruit and retain students in science and math-based disciplines represent one effort to improve STEM education at the postsecondary level. Through a variety of services such as mentoring, tutoring, and providing research opportunities, these programs aim to improve the likelihood that students, particularly underrepresented students, will enter into, persist, earn a degree, and even enter the workforce in a STEM field.

While these programs serve an important role in our nation's colleges and universities, the current economic downturn and current calls for accountability represent significant challenges to the long-term sustainability of these programs. As programs face pressure to demonstrate their value and seek funding to continue their important services, evaluations become an important tool in the quest for legitimacy and stability. This paper investigates the extent to which STEM Intervention Programs (SIPs) at a sample of large, public, research universities are evaluated, and if so, how the evaluations are conducted and the results used. In addition, challenges to conducting evaluations are explored.

LITERATURE REVIEW

This paper focuses on the intersection of SIPs, evaluations of programs, and legitimacy. The topics of intervention programs and the practice of evaluation are briefly discussed, followed by a description of legitimacy theory, which is the theoretical framework guiding this study

STEM Intervention Programs

Postsecondary intervention programs typically aim to increase the recruitment and retention of students, particularly disciplines such as the STEM fields. The programs featured in this study seek to increase the representation and success rates of students from traditionally underrepresented groups, including but not limited to women, racial and ethnic minorities, first-generation, and low-income students. SIPs can be found in a variety of science and math-based disciplines, including the agricultural and biological sciences, where differences in representation by group may not be as severe as in other fields such as Engineering, but students still benefit from services offered by SIPs.

In the effort to increase STEM participation and success, SIPs provide important services to students through a variety of formats including living learning communities, academic bridge programs, and research opportunity programs. SIPs typically seek to provide the following elements: 1) academic and social integration; 2) knowledge and skill development; 3) support and motivation; and 4) advisement (Hrabowski & Maton, 1995; Maton, Hrabowski, & Schmitt, 2000). These services often supplement the formal education students receive within the classroom and major, and can impact students' sense of belonging, academic outcomes, and even career choices.

ASQ Advancing the STEM Agenda in Education, the Workplace and Society
Session 2-3

Although empirical-based literature on the outcomes of SIPs is currently limited, there is an increasing amount of scholarship that highlights the benefits of conducting and utilizing internal and external evaluations, such as redefining program goals and objectives to meet stakeholder's needs, increasing program effectiveness, developing logic models, and informing programmatic changes.

Evaluation

Evaluation is the systematic review of a program or policy, which uses various methodological approaches, to determine its merit, quality, worth, or value. In addition, evaluations can be used in response to the need for accountability, including efforts “to improve and better programs and society” (Alkin and Christie, 2004, p. 12). More specific to educational settings, evaluations aid in generating knowledge of what interventions and services work best for which students, and in informing decisions related to funding and programming. While individuals and groups informally participate in the process of evaluation every day, evaluation as a profession emerged in the 1960s and has been continually evolving since—mainly in Western countries—with multiple theorists and practitioners contributing to the development of the field. The 1960s and 1970s greatly influenced early evaluations due to massive increases in public funding for social and education programs and policies, and the need to evaluate the merit, worth, and success of such programs and policies. More recently, evaluation has been described as “assisted sensemaking” (Mark, Henry, and Julnes, 1999, p. 179), where systematic inquiries result in understanding values which assist in determining a program's merit or worth. Evaluations have also been described as a process or source of “enlightenment” that educates and informs policy-makers and program managers (Weiss, 2004, p. 161).

It is important to note that evaluation is inherently different from research. Although they are both mechanisms of inquiry and utilize multiple methodologies, the specific questions addressed through the process of evaluation are fundamentally different, as well as the purpose of each activity—namely that evaluation seeks to impact actions and decisions, while research is often conducted for the purpose of knowledge production.

Legitimacy

The theoretical framework used in this study to explore the relationship between SIP programs and evaluation is legitimacy theory. Legitimacy theory is a resource dependency theory that describes how organizations—or in this case, a SIP—gain acceptance due to their relationship with mainstream norms and values within an institution or society. As the SIP becomes legitimate, it sustains the flow of resources from the environment to the organization (Hannan & Freeman, 1989). Moreover, Parsons (1960) argues that organizations that pursue goals that are in alignment with societal values have a legitimate claim on resources as judged by stakeholders (Lounsbury & Glynn, 2001).

Here, SIPs seeking legitimacy may be influenced to incorporate structures and procedures that match widely accepted social norms, or norms found within a college or university. As such, SIPs may be encouraged not only to align their missions and goals with certain values, but also to demonstrate their value or worth to the college or university. By conducting evaluations, SIPs are able to demonstrate how their services and programming align with the norms of the college or university, as well as gain support and funding based on the demonstration of desired outcomes. As a result, programming, evaluation, staffing, funding, and even visibility can be

ASQ Advancing the STEM Agenda in Education, the Workplace and Society
Session 2-3

viewed as a cyclical process for SIPs. Ideally, as a SIP gains legitimacy, it will have more financial and human resources to conduct evaluations and deliver services to students. Conversely, SIPs that do not conduct evaluations may be less likely to be viewed as legitimate programs, and as a result, may struggle to secure adequate and recurring funding, support, and proper staffing.

DATA AND METHODOLOGY

The current study is a component of a larger study on underrepresented undergraduate students in the STEM fields at large, public, research universities. The overall project uses qualitative and quantitative data to investigate the many individual and institutional factors that impact the educational decisions and outcomes of women, students of color, and low-income students' participation within the STEM fields.

One component of the larger research project sought to explore how SIPs are designed, funded, staffed, and evaluated.¹ The latter topic is the focus of this particular study, which seeks to specifically examine the extent to which SIPs are performing evaluations of their programs, as well as examine the methodologies used to conduct evaluations and how evaluation results are used by program directors and administrators. Face-to-face interviews were conducted with SIP directors and administrators at ten large public, research universities in 2009 and 2010. The universities are part of a consortium, are collectively significant producers of STEM degrees, and have historically offered interventions programs designed to increase participation and success in the sciences. Potential participants were identified by searching each university's website for intervention programs in a broad set of STEM fields, and were then recruited to participate in the study. While the specific missions, goals, services, and targeted populations of each program included in the study varies, each aimed to increase the enrollment, persistence, and/or graduation rates of underrepresented students in the STEM fields. Examples of SIPs in the study include, but are not limited to: summer research programs; mentoring and tutoring programs; leadership development programs; and first-year experiences for underrepresented students in STEM (e.g., designated housing, first-year seminars, etc.).

Participants included administrators, faculty, and staff of SIPs. Of the 137 individuals invited to participate, 55 were interviewed, resulting in a response rate of 40 percent. Of the individuals who participated, 12 were male, 43 were female; 26 were Caucasian, 21 were African American, five were Latino, two were Asian American, and one was Native American. In terms of their education levels, 20 had obtained a Ph.D., five were enrolled in a Ph.D. program, 21 had earned Master's degrees, and three had earned Bachelor's degrees. The remaining six participants did not indicate their highest level of education. On average, interviews lasted approximately one hour, with two interviewers from the research team per interviewee. Interview teams included one principal investigator and one research assistant. Of the interviewees, three were male and four were female; five were African American, one was Latino, and one was Caucasian. Participants did not receive compensation for participating in the study.

Responses to questions regarding the evaluation practices and results are of particular interest in this paper. Interviewees were specifically asked if the program had been formally evaluated, either internally or externally. If the program had been evaluated, participants were then asked about the focus and results of the evaluation. After the audio recordings were transcribed, the

¹ The full interview instrument is available at: http://stepup.education.illinois.edu/Data_Request

ASQ Advancing the STEM Agenda in Education, the Workplace and Society
Session 2-3

researchers utilized an open coding strategy to organize data into broader themes and issues. Open coding is concerned with describing, identifying, and categorizing the phenomena of interest (Stauss and Corbin, 1990). Upon development of coding, pattern-matching logic was used to identify distinct characteristics of institutional programs and common themes and issues across all the campuses. Pattern matching logic compares empirically based patterns with theoretical and conceptual frameworks (Yin, 2003).

FINDINGS

Approximately half of the programs featured in the study had been formally evaluated, while another 42 percent had not been formally evaluated, and 9 percent did not provide an answer to the question. Of the programs that had been evaluated, 42 percent had been formally evaluated through internal means, 18 percent were formally evaluated through external means, 15 percent had been evaluated both internally and externally, and an additional 6 percent did not specify the type of evaluation conducted. An additional 18 percent of programs had been informally evaluated through internal efforts. The focus of the evaluations conducted ranged from investigating the climate of a department or college, student outcomes, student attitudes, retention, academic transitions, and the extent that a program's mission is being met.

A variety of data collection techniques were used in the completed evaluations, including pre- and post-tests, online and written surveys, exit interviews, focus groups, students' self-evaluation, faculty rankings, observations, mid-program evaluations, and even the inclusion of comparison groups. In addition, some participants cited informally following students and their outcomes on social networking sites such as Facebook, and gathering anecdotal information through informal conversations with students. As expected, external evaluations conducted appeared to be better organized and were more formal as compared to internal evaluations. Of the program directors and administrators who had not conducted an evaluation on their program indicated that they would like to in the future.

An overview of the internal and external evaluations that had been conducted is discussed first, followed by common themes found within the data, including: 1) Evaluation as a Requirement; 2) Use of Evaluation Results; 3) Evaluation Expertise; and 4) Resource Constraints.

Internal Evaluations

As indicated above, the majority of program administrators who had evaluated their program did so internally, indicating that a staff member of the SIP conducted the data collection and assessment. Some of the internal evaluations gathered data through formal means such as exit interviews with seniors and pre- and post-tests to assess students' experiences and knowledge of STEM. However, internal evaluations also relied on anecdotal evidence to determine the program's success. For example, one administrator indicated that she would "hear stories" about former students in her program and their continuing involvement in the STEM field, including presenting their research at conferences.

Some administrators who performed internal evaluations of their programs benefited from having had training or a background in evaluation. For example, these administrators recognized the value of having a well-designed evaluation and were able to collect data from comparison groups for each cohort in their program. Being familiar with evaluation methods and design enabled some internal evaluators to incorporate formal aspects of evaluation into their

ASQ Advancing the STEM Agenda in Education, the Workplace and Society
Session 2-3

assessment efforts. As such, there was a variety of approaches, methods, and results from the internal evaluations conducted on SIPS.

External Evaluations

SIP administrators whose programs were formally evaluated were able to do so because of partnerships with other departments on campus, by hiring part-time graduate students who were earning degrees in evaluation, or by hiring an evaluator from outside the university. On-campus partnerships were typically with colleges of education who offer degrees in evaluation methods. Evaluators who were hired from outside the university were paid through funds that had been reserved for the specific purpose of evaluating the program. The National Center for Women in Information Technology was also mentioned as a resource for free evaluation services and assistance, including guidelines for conducting evaluations, examples of survey instruments, and best practices for data collection.

One program administrator mentioned several benefits to having her SIP externally evaluated, including the level of expertise provided by the external evaluators, the ability to demonstrate the program's value to others, and how the results could be used:

If you're doing a good job and you set up a good evaluation plan, which is why we have outside evaluators—which I think is good, because they know what they're doing much better than I do—I think if you have a good evaluation plan, which we needed for our NSF review, and if the program is successful and you show that to the college and university, then I think that's very strong. If it's not successful, that means you have to change something. And I'm okay with that. If it's not working, I don't want to keep it going the way it is, if it's not working... [the university] wants [the program] to succeed and they're gonna help make it sustainable.

This quote also demonstrates a number of the common themes found in the data, including the first theme: Evaluation as a Requirement.

Evaluation as a Requirement

A number of participants indicated that evaluations were performed to meet the requirements established by a funder, such as the National Science Foundation (NSF) and the National Institutes of Health (NIH), or by another stakeholder. In these cases, the evaluation was not viewed as optional but as a requirement for continued funding or support. One participant noted that the evaluation requirement made their staff "good stewards of that money" used to fund their program. Another participant discussed the desire of their advisory board to know how the funds were being used and what type of impact the program was having on students, while another mentioned that their "college wants to make sure that they're getting their money's worth."

Some funders, including corporations, NSF, and NIH, required programs to be evaluated and evidence of specific outcomes in order for funding to be renewed. For these programs, a portion of the program's budget is allocated for the purpose of conducting the evaluation. One participant actually criticized this practice arguing that the funds set aside for evaluation purposes should instead go towards serving the students, and that the funder should pay for the evaluation separate from the program's budget.

ASQ Advancing the STEM Agenda in Education, the Workplace and Society
Session 2-3

Despite requiring evaluations, the expectations of the funders and stakeholders were not always clearly articulated. One participant felt pressured by upper-level administrators within the university to demonstrate the value of the intervention program, but when asked what the evaluation should specifically focus on, the administrators were unable to provide an answer, other than they wanted the program to be evaluated. This particular program director preferred more direction and information before designing and conducting the evaluation, and at the time of the interview, was still waiting for instructions on what the evaluation should focus on. Although an external stakeholder desired an evaluation to be completed, without clear guidelines and direction, the evaluation was on hold. Similarly, another administrator described “vague pressure to evaluate our programs” which resulted in “a lot of resistance and pushback” due to the lack of clearly defined expectations.

Use of Evaluation Results

Of programs that had been evaluated, participants indicated that the results were used to make decisions about and inform changes made to the program. As one participant described, the purpose of the evaluation was “for us to figure out what we’re doing right and wrong.” Another participant indicated that the evaluation results allowed her to determine “where our students fit in and where they’re lacking, because that also helps me develop program. If we notice that students are struggling in a science and a math in the same semester, so we started a program, just a general workshop about studying for the sciences and math.” Other administrators indicated the need to use evaluations to report to offices within their department, college, or university’s diversity office. Findings from evaluations were also shared with funders who wanted assess the impact of a particular program and its services. Some participants indicated that the most important result of the evaluation was to show an increase in the number of students being recruited or retained in the STEM fields. With the bottom line being the numbers, other outcomes such as an increased sense of belonging for students, improved departmental climate, and increased opportunities to conduct research were considered to be less important than being able to demonstrate an actual increase in the number of students entering and succeeding in STEM.

Perhaps most important in terms of establishing legitimacy, is the use of evaluation results to provide evidence of the success of their intervention to others. Doing so enables program directors and administrators to secure recurring funding as well as evidence with which to approach new funders for additional support. One participant indicated that evaluation results “play an important piece to how we prove that we’re successful and can secure dollars” to provide additional services to students. Another participant described the use of evaluation in pursuing additional support: “as we work on building some grant opportunities, evaluation is a critical piece of that so we’re doing more of it.” One program in particular was able to demonstrate the worth and value of the program to their university, that the program was endowed with \$350,000 of funding per year, resulting in the program being institutionalized and sustained from year-to-year.

Evaluation Expertise

As mentioned above, a number of SIPs that had been externally evaluated were conducted through partnerships with local experts of evaluation, including individuals from their university’s college or school of education, while others were conducted by supervised graduate

ASQ Advancing the STEM Agenda in Education, the Workplace and Society
Session 2-3

students in evaluation or education degree programs. Many program directors spoke of wanting to perform internal evaluations, talking about the need for evaluation with their staff, and including evaluations in future plans for their program.

The program directors and administrators who did evaluations themselves or used an external evaluator spoke the language of evaluation, demonstrating a level of expertise on the subject. They described using formative and summative evaluations, and recognized the difference between using a mid-program evaluation as feedback and conducting an evaluation at the end of a program to determine its overall value and outcomes. Comparison groups, and pre- and post-tests were also mentioned in relation to evaluations, providing further evidence of familiarity with evaluation methods and techniques. In this sense, there appears to be a knowledge gap between those who are familiar with evaluation and evaluative methodologies and those who are not. One participant indicated “I’m not a statistician. I don’t know how to design a questionnaire. I don’t know how to do that... I’m a community organizer.” Another person acknowledged that evaluation was not her background and she did not identify herself as an evaluator but recognized that “evaluation is an important strength, and I know it’s something where we really need to do a better job of.”

Resource Constraints

The greatest challenge to conducting evaluations cited by participants was a lack of resources, including qualified staff trained in evaluation methodologies, funds, and time. The lack of resources could affect the ability to conduct an evaluation, as well as analyze the results of data that was collected. Perhaps most frustrating to one director was that data had been gathered from students who had participated in the program for the past five years, but the program lacked the resources to analyze the data, resulting in an untapped source of information that could be used to inform and improve upon the program. Other respondents indicated that evaluations, particularly external evaluations, were simply too expensive to conduct, and with many programs already facing scarce resources, providing students with programming and services took precedence over conducting evaluations. One participant weighed the options of spending money on an evaluation or servicing students: “I like to get feedback, and I don’t trust myself to do it well. But I don’t want to pay \$50,000 to someone to do it either cause it’s taking money from students.” This comment represents the intersection of the lack of resources programs face and a lack of evaluation expertise of program staff. By finding ways to hire staff who have evaluation training, or by partnering with others to perform an evaluation, the challenge of scarce resources may be less burdensome. Of the participants lacked the resources necessary for an evaluation, a number indicated that they would like to conduct evaluations in the future, if they had the resources to do so. These participants recognized the value of and need to conduct evaluations but simply did not have enough resources to design and conduct the evaluation.

SUGGESTIONS FOR BEST PRACTICES

The findings suggest that evaluations and their results are important tools for SIPs, including the ability to demonstrate the program’s value and worth to others, to seek additional funding and support, and to inform changes to an existing program. Given the reality of resource constraints that many programs, departments, colleges, and universities are currently facing, partnering with colleges of education and evaluation departments is one way to approach evaluations. Such partnerships will likely provide programs with the needed expertise of evaluation to develop an

ASQ Advancing the STEM Agenda in Education, the Workplace and Society
Session 2-3

appropriate design for the specific goal and needs of the evaluation and program. The availability of graduate student researchers trained in evaluation methods, who are typically seeking out a topic of study for thesis or dissertation purposes, represents one way in which the cost of evaluation can be minimized while not sacrificing the quality of the evaluation. Evaluations can also be budgeted within requests for funding, whether or not an evaluation is required. Integrating resources, such as funds, into the everyday operations of the program can increase the likelihood that evaluations will be conducted, rather than viewed as being optional.

Regarding staff and the need for evaluation expertise, programs should begin to seek training opportunities for conducting evaluations, “learning the language” of evaluation, and seek to hire staff who have a background in evaluation. In addition, evaluation results should be used to inform decisions about the program’s services, mission, goals, and students served, as well as seek to inform the creation of new programs designed to improve students’ access to and success in the STEM fields. Finally, SIPs with limited resources should seek to partner with other, similarly-designed SIPs to create shared data collection instruments from which to evaluate their programs. This would not only combine efforts and potentially minimize the cost associated with developing and administering the instrument, but also may allow for comparisons to be made between similar programs.

CONCLUSION

The findings presented in this paper indicate that evaluations and their results can serve an important role in demonstrating the value and worth of STEM intervention programs and the services they provide to funders and other stakeholders. By demonstrating the program’s value and worth, programs are in turn able to seek additional funding and support, which can increase overall legitimacy. Evaluation results can also be used to improve the program itself, enabling it to further contribute to the goal of increasing student access to and success in the STEM fields. While some programs are challenged by a lack of resources and knowledge of evaluation, evaluations can still be pursued through collaborations, purposeful staffing, and budgeting for future evaluations. As the current state of the economy and calls for accountability are likely to persist for the near future, evaluations are an important and necessary tool for STEM intervention programs and efforts to continue to expand the nation’s STEM education capacity and workforce.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 0856309. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The authors would also like to thank Mariana G. Martinez for her assistance in reviewing the data.

ASQ Advancing the STEM Agenda in Education, the Workplace and Society
Session 2-3

REFERENCES

- Alkin, M.C., & Christie, C.A. 2004. An Evaluation Theory Tree. In Alkin, M.C. (Ed.) *Evaluation Roots: Tracing Theorists' Views and Influences*. Thousand Oaks, CA Sage Publications, 12-65.
- Hannan, M.T., & Freeman, J. 1989. *Organizational Ecology*, Cambridge, MA: Harvard University Press.
- Hrabowski, F.A., & Maton, K.I. 1995. Enhancing the Success of African-American Students in the Sciences: Freshmen Year Outcomes. *School Science and Mathematics*, 95(1), 18—27.
- Lounsbury, M., & Glynn, M.A. 2001. Cultural Entrepreneurship: Stories, Legitimacy, and the Value of Acquisition of Resources. *Strategic Management Journal*, 22, 545-564.
- Mark, M.M., Henry, G.T., & Julnes, G. 1999. Toward an Integrated Framework for Evaluation Practice. *American Journal of Evaluation*, 20(2), 177-198.
- Maton, K., Hrabowski, F., & Schmitt, C. 2000. African-American Students Excelling in the Sciences: College and Postcollege Outcomes of the Meyerhoff Scholars Program. *Journal of Research in Science Teaching*, 37(7), 629-654.
- National Center for Women & Information Technology. 2011. *Resources: Assessment*. <http://www.ncwit.org/resources.assessment.html>
- Parsons, T. 1960. *Structure and Process in Modern Societies*. Glencoe, IL: The Free Press.
- Strauss., A. and Corbin, J. 1990. *Basics of Qualitative Research*. Newbury Park, CA: Sage Publications.
- Weiss, C. 2004. Rooting for Evaluation: A Cliff Notes Version of My Work. In Alkin, M.C. (Ed.). *Evaluation Roots: Tracing Theorists' Views and Influences*. Thousand Oaks, CA: Sage Publications, 153-168.
- Yin, R.K. 2003. *Case Study Research: Design and Methods*. Thousand Oaks, CA: Sage Publications.

ASQ Advancing the STEM Agenda in Education, the Workplace and Society
Session 2-3

AUTHORS' INFORMATION

Casey E. George-Jackson, Ph.D., currently manages Project STEP-UP at the University of Illinois at Urbana-Champaign, which investigates underrepresented undergraduate students in the sciences. She also serves as an Adjunct Assistant Professor in the Department of Education Policy, Organization, and Leadership at the University of Illinois. Her current research investigates issues of underrepresented students' access to and retention in higher education, including, but not limited, to the STEM fields. Casey can be contacted at cgeo@illinois.edu.

Blanca Rincon is currently a second year doctoral student in the Department of Education Policy, Organization, and Leadership at the University of Illinois Urbana-Champaign. Blanca is also a graduate research assistant for Project STEP-UP. Her research interests include issues of access and retention for underrepresented students in higher education, including the impact of participation in student organizations and departmental climate. Blanca can be contacted at rincon1@illinois.edu.