Evaluation and Program Planning 55 (2016) 103–111

Contents lists available at ScienceDirect

Evaluation and Program Planning

journal homepage: www.elsevier.com/locate/evalprogplan

An evaluation capacity building toolkit for principal investigators of undergraduate research experiences: A demonstration of transforming theory into practice

Audrey S. Rorrer*

University of North Carolina at Charlotte, 9201 University City Blvd., Charlotte, NC 28223, USA

ARTICLE INFO

Article history:
Received 25 October 2012
Received in revised form 16 December 2015
Accepted 17 December 2015

Keywords:
Evaluation capacity building
Toolkits
Participatory evaluation
Undergraduate research experiences

ABSTRACT

This paper describes the approach and process undertaken to develop evaluation capacity among the leaders of a federally funded undergraduate research program. An evaluation toolkit was developed for Computer and Information Sciences and Engineering[1] Research Experiences for Undergraduates[2] (CISE REU) programs to address the ongoing need for evaluation capacity among principal investigators who manage program evaluation. The toolkit was the result of collaboration within the CISE REU community with the purpose being to provide targeted instructional resources and tools for quality program evaluation. Challenges were to balance the desire for standardized assessment with the responsibility to account for individual program contexts. Toolkit contents included instructional materials about evaluation practice, a standardized applicant management tool, and a modulated outcomes measure. Resulting benefits from toolkit deployment were having cost effective, sustainable evaluation tools, a community evaluation forum, and aggregate measurement of key program outcomes for the national program. Lessons learned included the imperative of understanding the evaluation context, engaging stakeholders, and building stakeholder trust. Results from project measures are presented along with a discussion of guidelines for facilitating evaluation capacity building that will serve a variety of contexts.

1. Introduction

The current zeitgeist in the field of evaluation purports that building evaluation capacity is important, necessary, and valued (Cousins & Earl, 1992; Patton, 2008; Johnson & Weiss, 2011; Bourgeois, Whynot, & Theriault, 2015). It is important to develop the capacity of professionals who are leading programs because many have no training in evaluation research and lack the requisite skills to conduct an effective and meaningful assessment of outcomes. Not only are evaluation skills necessary, but building capacity empowers principal investigators to conduct consistent and valid assessments, provides understanding from the unique perspectives and contexts of each project, and strengthens generalizability of collective findings across programs with similar goals (McDavid & Hawthorn, 2006; Patton, 2008).

Evaluation capacity building, ECB, is an activity separate from conducting evaluation that is done at both the organizational and individual level (Labin, Duffy, Meyers, Wandersman, & Lesesne, 2012), and by definition must be sustainable with appropriate support tools (Preskill & Boyle, 2008; Stockdill, Baizerman, & Compton, 2002). A common thread across the ECB literature is that in order for capacity building efforts to be successful, they must be collaborative, participatory, and supportive of an evaluative culture (Labin, Duffy, Meyers, Wandersman, & Lesesne, 2012). Labin, et al. (2012) note that while there is broad agreement within the evaluation community about the definition of ECB, there is an ironic lack of discussion regarding the process of evaluation capacity building.

Common barriers to ECB have been identified as “timing, lack of alignment, and information overload” (p. 21, Johnson & Weiss, 2011). These barriers are particularly pertinent within the context of federally funded undergraduate research programs. Federally funded programs have been criticized for their dearth effective evaluation practice and dissemination of evaluation (Clifford & Millar, 2008; Shavelson & Towne, 2002), yet there is an expectation of credible evaluations across such programs (Frechtling et al., 2010).

To address the barriers to and need for evaluation capacity within a subset of National Science Foundation (NSF) Research

* Fax: +1 704 687 3516.
E-mail address: Audrey.rorrer@uncc.edu (A.S. Rorrer).
[1] Computer and Information Sciences and Engineering, CISE.
[2] Research Experiences for Undergraduates, REU.

http://dx.doi.org/10.1016/j.evalprogplan.2015.12.006
0149-7189/

Experiences for Undergraduates program in computer and information sciences and engineering fields, an online evaluation toolkit was developed to provide targeted instructional resources and tools for quality evaluation at the project level and aggregate program level. The development and use of digital evaluation support tools is not unique, and is quite common among national multi-site evaluations (Campbell, Townsend, Shaw, Karim, Markowitz, 2015). For example, the National Science Foundation's ADVANCE: Increasing the Participation and Advancement of Women in Academic Science and Engineering Careers grant program provides a website instructing institutional awardees on specific data to be collected, reported, along with reporting standards and timelines (http://www.advance.vt.edu/measuring_progress/tag_toolkits.html).

The CISE REU evaluation toolkit (the toolkit) is unique in its scope, feasibility, and adaptability to other REU programs and domain content. The toolkit aligns with the Joint Committee on Standards for Educational Program Evaluation Standards (Yarbrough, Shulha, Hopson, & Caruthers, (2011) in the incorporation of core principles of utility, feasibility, propriety, accuracy, and accountability. The rationale and approach to the development of the toolkit is presented in this paper along with the implications for evaluation theory and practice.

2. Background

2.1. Overview of CISE research experiences for undergraduate program

The National Science Foundation (NSF) sponsors roughly 700 programs for undergraduate research project sites with 7000 college student participants, commonly referred to as Research Experiences for Undergraduates, or REUs, in scientific disciplines at universities each year. The NSF REU program is organized into domains where identified REU sites are funded to manage iterative cohorts of undergraduate students who engage in research projects, most typically during the summer. Within the domain of Computer and Information Sciences and Engineering (CISE), there are between 50 and 70 REU sites with up to 700 undergraduate participants each year. The lifecycle of these site projects is typically a three-year period with an annual iterative cycle of recruiting undergraduate student participants during the fall and spring semesters for each upcoming summer research experience. While CISE REU sites share common goals, each site has unique research project objectives, for example, one site may focus its research projects on artificial intelligence and another on developing educational computer games. A clear challenge in building evaluation capacity in this domain is in balancing the need for individual site context with the need for broader, generalizable program impact (Hardre, Slater, & Nanny, 2010; McDonald and Hawthorn, 2006; Patton, 2008).

2.2. Evaluation challenges for CISE REU project sites

The NSF set accountability standards and provides general evaluation guidelines via a program evaluation handbook (Frechtling et al., 2010) available on the agency website. However, financial support for REU sites is directed toward supporting the undergraduate students and their research, with little support funding for evaluation implementation. Most REU sites are led by a single Principal Investigator, who manages student participant recruitment and selection, research laboratory preparations, and general site administration, which includes student housing arrangements and stipend payment procedures. Site project evaluation is not only limited in financial support, but also in expertise, both of which are commonly noted challenges in evaluation literature (Degroff, Schooley, Chapel, Poister, 2010; Worthen & Schmitz, 1997).

Prior to the development of the CISE REU evaluation toolkit, the focus of most site evaluations consisted of reporting the number of student participants, descriptions of their research projects and any research project dissemination. Student outcomes and overall program goals were not evaluated with consistency. Instead, each site project was left to internally develop evaluation strategies and techniques that were employed by project staff with little to no background in evaluation theory and practice, as indicated in the group discussions held prior to and during the development of the toolkit. Thus, CISE REU sites across the nation were reinventing methods and instruments for assessing project outcomes to meet the evaluation demands of the funding agency.

These circumstances present the challenges of and need for implementing effective, efficient, and valid project evaluations that not only address the individual site projects, but that also enable the aggregate program evaluation across CISE REU sites. In a study of computer science faculty about undergraduate research experiences, Matzen and Alrifai (2012) found that there was great variation in pedagogy, objectives, and project implementation despite the similar objectives of developing student research skills, preparation for graduate school and professional development. These findings suggest variation in project evaluation, especially given that no specific CISE REU site evaluation guidelines existed prior to the development of the toolkit discussed herein. The needs assessment that initiated the toolkit development, discussed below, confirmed that many variations of project evaluations were occurring across the CISE REU program.

Several key challenges are inherent to the CISE REU community, and are indicative of cluster evaluations. Cluster evaluation means that the evaluation "involves multiple sites in which project-level evaluation occurring at each site is conjoined to a greater or lesser degree with an over-arching, cross-site evaluation of some 'cluster' of these projects at the national, state, or initiative level" (p. 300, Worthen & Schmitz, 1997). Noted challenges indicative of cluster evaluations that the CISE program face include conceptual clarity, separation of evaluation authority and responsibility, alleviating project level anxiety, tracking longitudinal results, and sufficiency of evidence (Worthen & Schmitz, 1997). Within the CISE REU community, there was considerable lack of consistency of impact and outcomes measures, low evaluation expertise, and a level of concern and trepidation at the project level evaluation requirements. These issues together compound the problem of satisfactory evidence. The intent of the Toolkit is to identify common indicators across the program sites that when examined across the sites in a collective fashion would hold more significance, more fully engage the stakeholders, and alleviate the burden of evaluation for those PIs without access to evaluation expertise.

To address these challenges, an ECB process, as prescribed by Volkov and King's checklist (2007), was executed. The process for establishing a climate open and ripe for evaluation capacity development, the plan, the infrastructure of tools, support for and use of the tools are described herein. The overall logic model is presented in Fig. 1.

2.3. Assessment work group for CISE REU project sites

The CISE REU community convenes annually in a meeting of the Principal Investigators (PIs) to accomplish two primary goals. The first goal is to welcome new PIs to the community and orient them to the practice of managing and assessing their projects. The second goal is to engage in a collective conversation about the merits and impacts of the overall program. In 2009, the community was organized into several topical interest groups referred to as Working Groups. Meeting attendees elected to join the group of