

Tracing the Spark that Lights a Flame: A Review of Methodologies to Measure the Outcomes of International Scholarships

Mirka Martel

14.1 INTRODUCTION

International scholarship programs in higher education serve an important purpose in shaping the personal and professional pathways of their recipients. An increase in the availability of international scholarships worldwide has furthered interest in learning how these programs measure the outcomes of their interventions (Creed et al. 2012; Mawer 2014). Donors and academic institutions that invest in international scholarships are interested in understanding the potential returns—financial, political, social, or otherwise—on their investments. Recipients of international scholarships, and those interested in applying, would like to know how the opportunity will enhance their lives. Finally, policymakers and researchers in the international education sphere are eager to study whether an investment in an individual scholarship could have a ripple effect that produces impacts beyond that individual.

This chapter provides an overview of evaluation methodologies to measure the effects of investments in international higher education

M. Martel (✉)

Institute of International Education (IIE), New York, NY, USA

scholarships. In writing this chapter, I reviewed a variety of methodologies used to measure scholarship outcomes. Just as international scholarship programs are diverse in their design, the conclusion of my research is that there is no “one size fits all” approach to scholarship evaluation. Methodologies are complex and need to be chosen in a deliberate manner, taking into account important factors, including the program being evaluated and the time and resources available for evaluation. I reviewed over 30 evaluations of existing and former programs worldwide completed in the past 15 years (2001–2016). While most evaluations referenced are of international higher education scholarships that have an academic degree as an outcome measure, several non-degree programs, and secondary education programs, are mentioned for their innovative techniques. I present an overview of methodologies using several analytic lenses: understanding the theory of change; choosing the unit of analysis; the timeline for evaluation; and approaches to data collection. The chapter concludes with the importance of relaying evaluation outcomes to key audiences to improve programs and influence research in the field. There is still much that can be done to publish findings about the added value of programs and how they benefit recipients and their surroundings. Evaluations that are rigorous and transparent in nature provide important evidence to improve policymaking in international higher education and access to innovative and effective scholarships.

14.2 MAPPING A THEORY OF CHANGE

What is the void that international scholarship programs seek to address in higher education and how is this need being fulfilled? And what is the hypothesized change within scholar recipients, as well as other potential beneficiaries, that are affected by these programs? These two questions serve the basis for defining the theory of change of international scholarship and fellowship programs, a necessary first step for determining program outcomes. The theory of change is a detailed narrative of a program’s intended change and how it takes place, a methodological tool to trace the desired outcome (Center for Theory of Change 2016). Each international scholarship program serves a purpose; whether it is to provide students with professional skills that will enhance their future careers, open their eyes to the surrounding world through mutual understanding of other cultures, or learn the value of volunteerism and giving back to their communities (Perna et al. 2014). These purposes can be mapped to a theory of change that details how the program will make a difference in the individual’s life. For

example, if we posit that an international scholarship program can develop one's professional skills, then the theory of change hypothesizes that the skills ascertained will contribute to one's professional growth and career choices.

Programs that have a well-defined theory of change understand their program's place in the field of international education, and the need that their scholarship addresses. From this point the evaluation methodology outlines *how* the program intends to address this need, whether through changes in the primary beneficiary, the recipient of the scholarship, or secondary beneficiaries, the individual's surroundings. The theory of change enables programs to see a variety of complex relationships between the student that pursues an international scholarship and his or her spheres of influence: the home and host institutions, peers, and the home and host community. Conventionally international scholarship programs that have an individual focus center the theory of change on the individual, as he or she is the primary beneficiary of the program intervention (Boeren et al. 2008); Dassin and Navarette posit alternatives to this approach in the next chapter.

14.2.1 Defining Outcome and Impact Measures

While a theory of change provides a bird's eye view of the transformation that a program sets to achieve, the evaluation methodology specifies the outputs, outcomes, and impacts of the program intervention. In 2012, the Institute of International Education (IIE) conducted a review of foundations and non-profit organizations in the United States that administer large-scale international fellowship programs.¹ We compiled and examined published reports and study methodologies conducted by a variety of social science research institutions. The consensus was that most programs do not have a coherent strategy for tracking and measuring outcomes or impacts over time.

To date, most evaluations of international scholarship programs focus on the scholarship process and short-term outputs, such as rates of completion and program satisfaction (Creed et al. 2012). Although this data is useful to understand the short-term effects of a program, evaluations lack a more focused examination of the medium- and long-term pathways of scholarship recipients. The timing of evaluations is key in determining what effects can be measured (See Fig. 14.1). Since most evaluations occur at one time, usually as a program is finishing or has finished, these cross-sectional

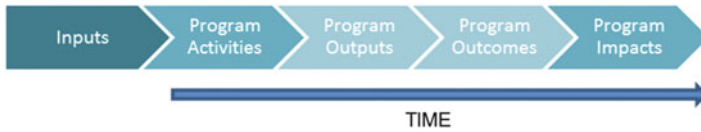


Fig. 14.1 A pipeline graph depicts the progression of program activities to outputs, outcomes, and impacts over time

assessments cannot estimate change over time and instead provide a very limited, one-dimensional view of program outcomes.

14.3 CAUSALITY AND CONTRIBUTION

With the theory of change as a first step in defining a program's goal, evaluators next face the methodological quandary of proving that the stated program intervention is causing the intended outcome or impact. Particularly in complex social systems, the difficulty of attribution is a real one: the extent to which a program is able to prove, with a level of certainty, that the intervention was the cause of change in a recipient's outcomes is often limited. Given that international scholarship programs at the tertiary level involve scholars who are young or mid-level professionals, the issue of external validity is pertinent, that is, the ability to show that external factors are not influencing the scholarship recipient in combination with, or in lieu of, the program itself (Bamberger et al. 2011). Two methodologies that attempt to address issues of causality in complex systems are counterfactual and contribution analyses.

14.3.1 Counterfactual Approaches

Discussion of anticipated change from an international scholarship program necessitates consideration of the counterfactual: What would have been the pathway of the scholar had he or she *not* received the scholarship? Evaluation methodologies that employ a counterfactual design, such as randomized control trials (RCTs), can definitively measure the change related to the program intervention (Gertler et al. 2011; Jadad and Enkin 2007). In international education, counterfactual studies have mostly been implemented in basic education programs, in which randomized education interventions among students at the primary level are compared to counterparts who do not participate in the treatment (see, e.g., Banerjee et al. 2005).

These interventions isolate the program intervention on a clearly determined quantitative outcome variable, such as subsequent test scores.

In international scholarship programs, these types of studies have been very rare due to two limitations. First, most international scholarship programs choose students because of various criteria that span academic, professional, and social distinctions, and as such very few recipients are randomly chosen. Thus, it is difficult to identify a cohort of students that can serve as a comparison group that would be identical to the recipients' characteristics. Further, at the tertiary level most students already have extensive years of education and social conditioning that compel them to consider an international scholarship; it is difficult to determine that students who are chosen for a comparison group would have been interested in, or chosen for, the same scholarship program.

While noting these limitations, it is my belief that investments in comparative studies could yield important evidence about the advantage of scholarship programs. Several examples can illustrate the possibilities. In 2013, the US Agency for International Development conducted a retrospective evaluation of the SEED higher education program that compared recipients of the scholarship to individuals who had applied to the program but were unsuccessful (Chesterfield and Dant 2013). The unsuccessful applicants chosen had the closest individual profile to scholarship recipients and thus were comparative to the trajectories of scholarship recipients. In 2008, AFS conducted a study to measure the long-term impacts of their study abroad programs by comparing program alumni to their peers 25 years following their study abroad experience (Hansel and Chen 2008). Even earlier, in 1993, IIE conducted a study of the Japan-US Fulbright program and compared differences between Fulbright recipients from Japan and the US and their non-program colleagues (Uyeki 1993). These evaluations demonstrated the positive outcomes of program participants *in comparison* to plausible counterparts.

All three of these evaluations have one major caveat: the comparison group was constructed retrospectively, meaning *after* the program finished. To improve the precision of the evaluation, the counterfactual should be constructed along with the program from its inception. IIE's Higher Education Readiness (HER) Program in Ethiopia, for example, created a quasi-experimental design to compare scholarship recipients to non-participants (Valuy and Martel 2016). The selection process included two phases. First, the scholarship applicants were screened for eligibility to ensure that all study participants met the criteria of the scholarship. Thereafter the top

candidates were chosen for the intervention, while all others were included in the comparison group. A baseline survey was conducted to capture academic differences between the two groups prior to intervention. IIE conducted a quantitative difference-in-difference analysis to reveal variation in academic outcomes between scholarship recipients and the comparison group. Counterfactual program designs are also being implemented by the MasterCard Foundation Scholars Program and the Commonwealth Scholarship Commission in the UK, though these studies are ongoing and have not published full findings.

Evaluations that employ a counterfactual analysis in international scholarship programs will likely increase in the future, given the focus on and interest in this type of research. These methodologies will need to grapple with the limitations of scholarship design and selection, as well as the external factors that may influence recipients over time.

14.3.2 Contribution Analysis

Debates over whether randomized control trials are a valid methodological option in complex social environments have led some researchers to consider alternative approaches to studying program outcomes (Cook et al. 2009; Mawer 2014). One such approach is contribution analysis, which focuses on the additive value of a program without discounting the effect that external factors may have on the beneficiaries. As a leading advocate of contribution analysis has noted: *In assessing attribution, contribution analysis does not use a counterfactual-based argument, but rather builds a case for reasonably inferring causality, recognizing that in many situations one cannot prove causality in the positivist tradition* (Mayne 2011, p. 6). Its use in international scholarship programs has been limited, as discussed by Mawer (2014). The United Nations handbook for evaluations mentions the methodology, though no examples of the methodology being used in practice were found (Rotem et al. 2010). Contribution analysis was also used in a hybrid evaluation approach to two international education programs in the Netherlands: the Netherlands Programme for Institutional Strengthening of Post-secondary Education and Training Capacity (NPT) and the Netherlands Initiative for Capacity development in Higher Education (NICHE) (Ramboll Management Consulting 2012).

The importance of clearly defining a methodology for measuring the program theory of change in any evaluation cannot be overstated. Contribution analysis is an approach that could be studied further in this regard.

While some evaluations I reviewed could be considered examples of contribution analysis, no evaluations save one (NPT and NICHE) mentioned the methodology outright. As a result, this type of analysis warrants more research and evaluations that employ the approach to measure their program outcomes.

14.4 SETTING A UNIT OF ANALYSIS

Evaluations that set out to measure program change must adequately trace the outcomes of an international scholarship opportunity to the program activities. The goal of most graduate scholarship programs is to enable the participant to increase his or her potential in the workforce, the community, and ultimately society. There is an added dimension of *international* programs: some programs posit that the intervention is meant to increase mutual understanding, while others have a broader goal to influence development impacts in the host or home communities (Boeren et al. 2008). Many program implementers share a desire to document impacts that go beyond the individual experience.

14.4.1 *Measuring Outcomes at the Individual Level*

A majority of international scholarship evaluations focus on individual recipient outputs and outcomes. The evaluations reviewed demonstrated a spectrum of rigor. Methodologies that study change in the individual usually focus on (1) scholarship completion and satisfaction, (2) change in academic and professional attributes, and (3) change in personal attitudes or beliefs. All these methodologies have in common that the primary purpose is to measure the benefits of the scholarship to its recipient.

- **Scholarship completion and satisfaction:** Most scholarship programs have ample statistics about the completion rates of their students, and this quantitative data is often used to justify program success. Unfortunately, program completion rates offer a very limited look into the outcomes of the scholarship. While all recipients may successfully complete their scholarship, the program may lack information about the recipients' reflections on the scholarship. For this reason, many organizations have also employed methodologies that survey students at the end of the scholarship to rate their satisfaction with their program. Further, these surveys are useful program

management tools, as programs can adjust and improve their implementation in consequent cycles (Kusek and Rist 2004).

- **Change in academic and professional attributes:** Methodologies that go beyond program satisfaction offer a more detailed look into the recipient's experience. However, these evaluations require more time and resources. Methodologies with the most advanced approach employ a pre-post design, surveying recipients at the beginning and end of their scholarship opportunity. For example, the USAID LOTUS Scholarship Program in Egypt requires scholarship recipients to take English proficiency tests at the beginning and end of their scholarship to measure language improvements (Institute of International Education 2016). The Boren Awards, sponsored by the National Security Education Program (NSEP), provide language learning opportunities for US students abroad. The program recently published a rigorous 15-year study of oral proficiency gains among its scholars (Mason et al. 2015).
- **Change in personal attitudes or beliefs:** Scholarship programs that have an international component expose the scholar to an environment other than his own. Some programs, particularly those sponsored by governments, are interested in increasing mutual understanding and measure impact related to citizen diplomacy (Bhandari and Belyavina 2011). Programs are interested in understanding whether scholarship recipients have changed their views of cultures other than their own. US Department of State programs (e.g., Visiting Fulbright Student Program, Benjamin A. Gilman International Scholarship Program) include survey outcomes related to scholars' cultural exposure and feelings of mutual understanding (SRI International 2005; Research Solutions International 2016). The Erasmus Mundus program, sponsored by the European Union, reports on changes in mutual understanding as well (PPMI 2012).

Methodologies that explore individual change may also focus on thematic elements. For example, many scholarship programs have a leadership component and focus on emerging leaders as their primary beneficiaries. These programs evaluate change in leadership through pre- and post-program assessments that use indices specified by the donor or implementing organization. Leadership assessments have been carried out in evaluations of IIE's Higher Education Readiness (HER) program (Valuy and Martel 2016), PACT International's Girls Youth Development in

Action (CARE 2009), and the Gates Foundation Millennium Scholars Program (Amos et al. 2009).

14.4.2 *Measuring Outcomes Beyond the Individual*

The importance of measuring outcomes beyond the individual is to understand how a scholarship program with an individual focus may contribute to effects in the scholar's spheres of influence. In the short term, programs may have evidence that a student will finish his or her degree after scholarship completion, return to his or her country, and obtain employment post-graduation. Beyond this, however, few programs delve into key questions: "What did beneficiaries do with their scholarship success?" or "How did scholarship recipients use their knowledge gained to bring about social change?" To answer these questions practitioners must explore methodologies that define and measure change from the individual to the communal.

Kirkpatrick Model

A useful methodology that maps individual to communal change comes from a revised model of Donald Kirkpatrick's (1979, 1994) Four Levels of Evaluation. Kirkpatrick's model describes the levels of impact that measure change resulting from an academic experience, ranging from a short-term training to a full-degree program. The model outlines levels of change starting from the individual and proceeds to measure change at the institutional level. Adaptations of the Kirkpatrick model have been used in evaluations of several international scholarship programs, including the USAID ATLAS/AFGRAD program (USAID 2004) and the Canadian Francophone Scholarship Program (CIDA 2005). The USAID program evaluation was the first to add a fifth level to measure *impact that may occur beyond institutional boundaries, for instance, in a sector, or at the national, regional or international level* (p. 87). A recent IIE publication of the International Fellowships Program (IFP) tracking study includes a graphic representation of the revised Kirkpatrick model with this fifth level of impact that measures broader societal impact (Martel and Bhandari 2016) (Fig. 14.2).

As noted earlier, the theory of change often stems from the individual experience and maps outcomes on the organization or community. Kirkpatrick's methodology is useful in that it de-emphasizes the individual as the only possible change outcome. Many studies end at Kirkpatrick's level one or two, assessing the impact solely at the individual level. Kirkpatrick, rather, focuses on the application and behavioral transfer of knowledge to

KIRKPATRICK'S LEVELS OF EVALUATION

LEVEL 1: REACTION

The participant's impression of the program. This includes the participant's level of satisfaction with their fellowship experience.

LEVEL 2: LEARNING

The acquisition of knowledge and skills from the intervention. Were the objectives of the program met? Did the participant indicate learning something from the intervention?

LEVEL 3: APPLICATION

The application of the participant's knowledge, skills, and attitudes (KSAs) in his or her home organization (whether work or volunteer). The ways in which behavioral changes are applied in his or her actions.

LEVEL 4: ORGANIZATIONAL RESULTS

The participant's behavioral changes lead to impacts at the organizational and community levels.

LEVEL 5: EXTERNAL RESULTS (added by IIE)

The participant's behavioral changes lead to impacts at the community/national/international levels, leading to changes in policy and social behavior.

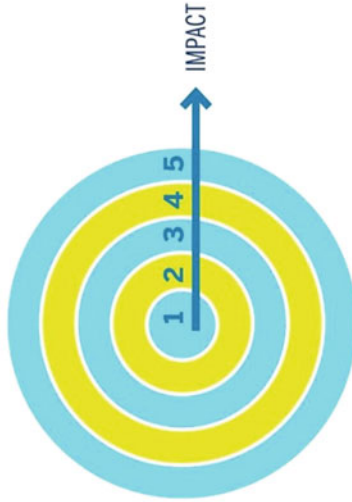


Fig. 14.2 A revised version of the Kirkpatrick model focuses on two levels of potential impact beyond the individual: organizational and external (societal)

one's environment or secondary beneficiaries. Further, Kirkpatrick's model is *goal based*, meaning that the model identifies the intended goals of the intervention, but does not necessarily evaluate the processes to achieve those goals. This is different from a system-based approach, where each goal is examined based on the process by which one achieves it. Therefore, the model allows the flexibility to approach program success based on outcomes.

Kirkpatrick's levels generally follow the progression of program outputs, outcomes, and impacts over time. The crux of the Kirkpatrick model for assessing wider impacts comes in level three: the *application* of the scholar's new knowledge in his or her surrounding environment. This level is a necessary channel for transfer from the individual to the communal. If the scholar does not apply what he or she has learned as a result of the scholarship, then program outcome ends at the "individual level" (levels one and two). If application takes place, we can analyze the scholar's pathways through various prisms: how the scholar's actions lead to change at the organizational, communal, or societal levels.

Social Network Analysis

Students who participate in a scholarship program are introduced to various networks during their program experience. These networks can be analyzed further to understand how the relationships of the scholar deepen over time and how they enable changes in his or her choices after the scholarship. Networks are a resource for scholars to apply and share their knowledge, and therefore networks map to the Kirkpatrick model. Further, the effect of scholarship networks has a causal relationship to the program intervention. In other words, the scholar would not have exposure to these networks were it not for the scholarship opportunity. As a result, methodologies that explore these networks can conclude that the impact of the network is solely attributed to the scholarship.

Social network analysis (SNA) is a methodology that can be used to study how program recipients are leveraging their networks. SNA is a tool in modern sociology to identify the links between individuals in various social systems (Scott and Carrington 2011). It can also be used in monitoring and evaluation to probe deeper into the power of social interactions. Using SNA, programs are able to measure and depict how well the program supports development of scholar networks (see Fig. 14.3). Through qualitative analysis, programs can also analyze how current scholars and alumni use networks for change (Tvaruzkova 2012).

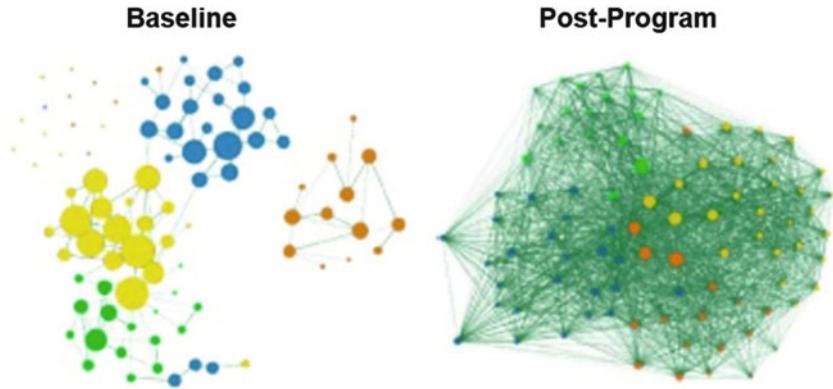


Fig. 14.3 A sociogram depicts the social networks among scholars before and after a scholarship program. Circle size is larger for scholars with more connections. Each line is a connection between two scholars

There are many different types of networks that are associated with an international scholarship. First, the participants of a scholarship program, and alumni of that program, constitute a primary program network. For example, the Fulbright Program has an extensive program network that is comprised of current scholarship recipients and thousands of alumni. All these students and scholars together share a common characteristic (receipt of a Fulbright scholarship) and can be analyzed based on their network's properties. Scholarship recipients that study internationally also have a new network in their host community, whether academic or social. Students visit international campuses and meet host students, faculty, and families, all of whom participate in a new network for the scholarship recipient. Finally, the various stakeholders of the scholarship program, including the donors, home academic institutions, and home communities can also include key networks.

The influence of networks can be measured in various stages. A program that focuses on networking, such as the Seattle International Foundation's Centroamerica Adelante Program (Valuy 2016) can measure how participants are interacting with each other, how many of the relationships are reciprocal, and which relationships are stronger than others. Information about networks can then be contextualized to understand how effective networks are, whether they lead to a potential for collaboration or joint projects. Several programs have been able to show that program networks lead to significant outcomes and new collaborations among its recipients

and alumni (Martel and Bhandari 2016; Marsh et al. 2016). Ideally, programs use both quantitative methods to measure the networks created and qualitative methods to analyze the significance of these networks, areas for collaboration, and network sustainability.

14.4.3 Outcomes for Multi-country Programs

International scholarship programs in higher education include recipients and scholars from many different countries and contexts; evaluations of these programs need to consider how any methodology may be implemented in the international context. Bamberger (1999) and Bamberger et al. (2011) have discussed the challenges of conducting cross-cultural evaluation. Two considerations addressed are the extent to which various stakeholders are involved in the evaluation methodology, and the close attention evaluators should be paying to local customs and values. Large-scale programs that employ a global evaluation methodology may find that practices and methods used in one context may be largely inappropriate in another. Evaluators must consider cultural sensitivity, and while the overall methodology has a “global” face, the “local context” must not be understated. Above all, multi-country evaluations must take into consideration cross-cultural sensitivity in data analysis (Chouinard and Cousins 2009). As a result, most evaluators advocate for a mixed methods approach in multi-country evaluations, in order to sequence and present outcomes that are cross-cutting without losing more in-depth information about local experiences and outcomes (Bamberger et al. 2011).

14.5 A TIMELINE FOR EVALUATION

Evaluations of international scholarship programs take place at different times, based on when they are commissioned, the financial resources that are at their disposal, and the extent to which monitoring and evaluation is considered in the program design. Many current evaluations are one-time assessments at the end of the program funding cycle. This is understandable, as these types of evaluations require the least time and resources. Further, donors are often eager to learn early the short-term outcomes of the program. However, as already mentioned, these evaluations are frequently limited in their design and rigor. Most importantly, they are not able to measure adequately the outcomes and impacts of international scholarship programs over time. I advocate for a more nuanced approach to

international scholarship evaluation, specifically in considering the timeline for evaluation.

14.5.1 Retrospective Alumni Study

An augmentation to the conventional, post-program evaluation is to conduct an impact study several months, or years, following the participant's program experience. Most program outcomes and impacts are linked to change that will take place over time after the program has completed, and thus evaluations that take place immediately at the end of the program cycle simply cannot measure these outcomes. At most, evaluators can capture data on what alumni intend to do with their scholarship experiences; however, there is no opportunity to learn whether these intentions come to fruition. Retrospective alumni studies allow programs to study the potential impact of the program intervention beyond the individual. The Schlumberger Foundation, for instance, conducted a 10-year retrospective evaluation of its Faculty for the Future Program (Institute of International Education 2015). The program dispersed scholarship opportunities to women in science and technology in developing countries. An evaluation ten years later provided evidence about the impact of the program participants in teaching and publishing in their home countries, mentoring students in science and technology, and inspiring the next generation of women scientists. The value of both assessments in conducting data collection after program intervention was that the evaluators were able to trace program alumni pathways home and discuss potential outcomes and impacts on the home communities of scholars.

While the scope of this type of evaluation is comparable to a post-program evaluation, the timing and resources may be more complex given that the evaluation takes place when alumni are no longer associated with the program. Additional resources may be required to track alumni. This is further exacerbated when the program concludes and no one maintains contact with the alumni. These types of evaluations should account for ample time and resources to track alumni prior to data collection and should consider incentives for alumni to participate in data collection.

14.5.2 Longitudinal Study

A further level of rigor is to collect data on international scholarship outcomes and impacts at various points in time; in this chapter I will focus on

longitudinal studies that take place after a program has finished. Longitudinal studies aim to systematically analyze the lasting or significant changes—positive or negative—in people’s lives brought about by a given action or series of actions. A longitudinal study of international scholarships focuses on the program participants and their pathways over time. This type of study offers a detailed view of the changes in the scholars’ lives during and after the program, and the extent to which the program intervention may have contributed to these changes. Further, longitudinal studies allow programs to anchor outcome measures over time. Evaluators can gather data on program outcomes at various points in time in a comparable way, with the potential to have baseline data prior to the program intervention.

As Creed et al. (2012) have indicated, longitudinal tracking studies are not common in international scholarship programs in higher education. The financial burden of conducting a longitudinal study often makes this an unwelcome choice for donors. Since longitudinal studies take several years before data is available, the high costs and time investment required can make them unattractive to funders and program implementers alike. As a result, most programs resort to one-time assessments and evaluations. The second difficulty of measuring program impact longitudinally is related to the question of causality, that is, the ability to attribute long-term impacts to the initial scholarship experience. If a student volunteers at an organization 10 years after their initial scholarship experience, can one attribute this decision back to the scholarship opportunity? In the 10 years, it is undoubtedly true that other external factors may have influenced the participant and his or her opportunities and choices.

Despite these challenges several studies have implemented longitudinal designs on a large scale. IIE’s work conducting a 10-year tracking study of the Ford Foundation International Fellowships Program (IFP) has given us the opportunity to study long-term impacts on over 4300 alumni worldwide (Martel and Bhandari 2016). The first findings of the study collate survey outcomes from all 22 countries where the program was implemented. The next phase of the evaluation includes a qualitative approach in which local researchers are conducting case studies in select countries. This mixed approach allows for the combination of global perspectives (quantitative) and local applications (qualitative). The MasterCard Foundation is also conducting a rigorous, 10-year study of the MasterCard Scholars Program (Cosentino et al. 2015). This longitudinal study is taking place during the program’s implementation and uses a counterfactual design to measure differences over time between program participants and

non-participants. Early results point to significant differences between the two groups, though these results are based on very small sample sizes (MasterCard Foundation 2016).

Both IIE and the MasterCard Foundation are using innovative approaches in longitudinal studies to collect mixed methods data over long periods of time. The two studies focus on the participants of the program and their personal trajectories, and the extent to which the program interventions may contribute to life choices and opportunities. Both programs have also integrated opportunities for participant and alumni engagement, whether through networking events or alumni awards, as effective ways to maintain interest among alumni beyond the fellowship or scholarship. These techniques address the limitations of engaging with large samples of beneficiaries over time in data collection.

14.6 DATA COLLECTION METHODS

Data collection methods are the means by which evaluators collect information from various evaluation participants, including scholarship participants, their peers, institutional partners, and various other stakeholders. Each evaluation should use the data collection methods that are most suitable for exploring the study's methodology. To date, most evaluations simply use data collection methods that are readily available and used often, such as surveys and interviews. While there is no doubt these methods are useful, ideally researchers should have the time to consider the evaluation methodology and which methods will best explore the outcomes and impacts of a program. This will ensure that each program evaluation identifies the best methods suitable for its purpose.

It cannot be underestimated that the choice of data collection methods is also strongly influenced by the time and resources available for evaluation. For example, online surveys are a timely and relatively inexpensive method to collect large amounts of data. On the other end of the spectrum, oral histories or narratives, or ethnographic research, can take years and significant resources to collect. Ideally, each program should take into consideration the time and resources available for evaluation, and based on this determine most appropriate methods.

14.6.1 *Quantitative Methods*

Web-based and paper surveys continue to be the most prevalent method used to collect data about international scholarship participants. Most programs conduct surveys because they are relatively low cost and can be collected across different countries, cohort years, and various mediums. Surveys are most suitable when evaluators are interested in collecting cross-sectional data from large numbers of evaluation participants. As international scholarship programs often include scholars from various countries, surveys are an ideal tool to collect information across multiple locations. Most surveys include specific questions that probe for outcome and impact measures. Some program evaluators also construct matrices to measure outcomes for character attributes or opinions.

Surveys have two limitations that evaluators should anticipate: selection bias and self-reported findings. The first is an issue of survey response. Most surveys conducted to measure social outcomes and impact take place after a program has ended and the survey participants are no longer associated with the program. Even if a sampling scheme is in place, survey participants still participate in surveys on a voluntary basis. Self-selection bias stems from the possibility that some scholarship alumni could be disproportionately more likely to respond to the survey. A particularly common concern is that participants who had a particularly positive or negative experience may be more likely to respond than those who had less “extreme” experiences. In analyzing program outcomes, evaluators must keep this limitation in mind. A further limitation of surveys is that respondents self-report their answers. It is difficult to verify the information provided by respondents, especially when they are the sole source of data collection. When possible, it is preferable to collect data from more than one source, and from more sources than just the program scholars, so as to triangulate and confirm program outcomes and impacts from various different stakeholders.

Impact beyond the individual is difficult to measure without information from secondary sources, other than the program participant. One method being used to mitigate self-reporting bias in the case of alumni working in research and academic roles is bibliometric analysis. This type of analysis aims to quantitatively measure the impact of published academic writing. A recent evaluation of the National Science Foundation’s Partnerships for International Research and Education (PIRE) Program, for instance, conducted a bibliometric analysis to compare the relative impact of PIRE projects and other similar projects (Martinez et al. 2015). The evaluators

were able to compare field- and journal-specific impact of academic writing without relying only on self-reported data. However, this method has its limitations. Bibliometric analysis is limited to the reference library used in the citation search; reference libraries may not readily include all published sources worldwide, particularly journal articles written in regions outside the Western Hemisphere, and work completed in languages other than English.

14.6.2 *Qualitative Methods*

Interviews, focus groups, and case studies provide deeper understanding of the change that take place resulting from an international scholarship learning opportunity. While surveys can help practitioners understand comparable outcome measures, qualitative methods allow evaluators to delve deeper to understand the context of the anticipated and actual change among program participants and their spheres of influence. Qualitative research is time and labor intensive, and as such is less prevalent. However, several qualitative evaluations have highlighted the outcomes and impact of scholarship programs by exploring in depth the trajectories of program participants.

Innovative techniques are being used in qualitative methods to expand on outcomes and impacts beyond the individual scholarship experience. Participatory action research has been used to understand how the individual change among scholarship recipients can lead to collective change (Hofmann-Pinilla and Kallick Russell 2009; Chen et al. 2010). Scholarship recipients participate in a series of workshops to express the change they have undergone as a result of their opportunity and in groups (or otherwise) discuss the larger impacts of the program intervention. This method allows evaluators to collect data not only from individual participants, but also from groups of scholarship recipients who participate in the workshop together. Participatory action research advocates for alternatives to traditional question-answer methods, using modes such as drawing to allow participants to express their transformation. This powerful tool can guide scholarship recipients through the transformative process of documenting change from the individual to the communal.

Many program evaluations have combined quantitative and qualitative methods in some way, collecting data from surveys and interviews to satisfy specific purposes. The IFP alumni tracking study is deliberately sequencing quantitative and qualitative collection to allow the research team to collect

more broad, cross-cutting data first, followed by more in-depth qualitative data collection (Martel and Bhandari 2016). Mixed methods are a worthwhile tool in scholarship evaluation, but these methods should be used in a deliberate manner: simply using quantitative and qualitative methods in one evaluation does not constitute a rigorous mixed methods study. Evaluators should consider how to employ quantitative and qualitative methods, in what order, and how one set of data will inform the other. The most rigorous mixed methods studies clearly articulate how quantitative and qualitative data is integrated (Hesse-Biber 2010).

14.7 CONCLUSIONS: ANALYZING DATA FOR MEANINGFUL OUTCOMES

The outcomes of an evaluation are not only important to the accountability and transparency of a program. They are also valid tools for learning, both for the program itself and other practitioners in the field. The last section of this chapter advocates for the necessary exposure of evaluations to key audiences for the purposes of learning. Evaluations can enable donors, practitioners, programs, and their participants to improve policy and practice over time. While measuring outcomes and impact is but one domain of inquiry, the field remains severely limited due to the inability of evaluations in international scholarship programs to contribute to learning and ongoing improvement of practices. As such, evaluators and practitioners must continue to advocate for methodologies and evaluations that are public and widely available, and that are used for furthering the field.

Too often evaluations of international scholarship programs are completed and remain for internal use of donors only. In conducting research for this chapter, I came across several evaluations of major international scholarship programs that remained internal documents. It is important to emphasize that donors are not the *only* viable audiences for evaluation outcomes. As desire for measuring outcomes of programs increases among researchers, practitioners, and scholar participants, these audiences are equally eager to learn the outcomes and impacts of evaluations.

- **International higher education scholarship practitioners:** Evaluations can allow programs to learn ways that they can improve the program experience: the overall program design and its implementation. It can also provide recommendations for how to improve the

measurement of outcomes and impacts on scholars. Finally, evaluations have a valuable function for validating the hard work of program staff, the dedication of a program to make changes that contribute to scholars and their home and host communities.

- **International higher education donors:** Another key audience for evaluations are other program donors in the field. Evaluations provide concrete evidence that allows practitioners in the field to learn about the advantages of certain program interventions, and challenges and pitfalls of others. This can allow programs to work together to learn from each other, and avoids programs repeating mistakes. Evaluations can build a community around the shared value and importance of international higher education scholarship programs.
- **Researchers and evaluators:** At the onset of the chapter I discussed the potential of evaluation to expand the research field of inquiry around the value of international scholarships. The value of public evaluations means that credible research is developed on international scholarship practices. While the methodologies mentioned in this chapter vary in their rigor, one continuous thread throughout is that they rarely build on each other, meaning there is a lack of collaborative learning in the evaluation community. As a result, a key audience for expanding evaluation methodology is a community of learning among evaluators of these types of scholarship programs. There is also a key desire to connect the research done on these types of programs to international education more broadly.
- **Scholarship and evaluation participants:** Scholars participating in scholarship programs or interested in a scholarship program can benefit from learning the findings of evaluations to understand the potential benefits and challenges they may face in pursuing an international scholarship program. Scholars currently enrolled in a scholarship program are the participants surveyed or interviewed for data collection. It is important to consider the ethical responsibility that evaluators have to these participants, not only in having their voice heard in an accurate, responsible way but also in allowing them to learn the results of the evaluation once it is finished. More evaluators should make this commitment both with sponsors and participants, as it allows evaluation participants to read the outcomes of the evaluation and confirm that their opinions were adequately captured. This holds sponsors and evaluators accountable to pursuing an evaluation that is reflective of respondents, and not biased to client-driven outcomes.

Evaluators of international scholarship programs in higher education have an ethical responsibility to their respondents to accurately portray the evaluation findings. As such, these researchers must continuously ask themselves what purpose the evaluation serves and whose voice is being included or discounted. This is particularly important in scholarship programs that are aiming to increase equity and access to higher education. Equity-focused scholarships must be matched with evaluations that underscore the importance of transparency and accountability in evaluation. This will allow practitioners and evaluators to learn from findings and find appropriate solutions and policies to increase opportunities for international scholars in the future.

NOTE

1. While the report conducted by IIE was internal, the topics that emerged were discussed in a roundtable held at the Institute in December 2012 and were outlined on IIE's blog: <http://www.iie.org/Blog/2012/December/Alumni-Tracking#.V-CODc6cHIU>

REFERENCES

- Amos, L.B., Windham, A, de los Reyes, I.B., Jones, W., and Baran, V. (2009). *Delivering on the promise: An impact evaluation of the Gates Millennium Scholars Program*. Washington, DC: American Institutes for Research.
- Banerjee, A., Cole, S., Duflo, E. and Linden, L. (2005). *Remedying education: Evidence from two randomized experiments in India*, No. w11904, National Bureau of Economic Research. Available at: <http://www.nber.org/papers/w11904.pdf> (Accessed 23 Jan 2017).
- Bamberger, M. (1999). Ethical Issues in Conducting Evaluation in International Settings. *New Directions for Evaluation*, 82, pp. 89–97.
- Bamberger, M., Rugh, J. and Mabry, L. (2011). *RealWorld evaluation: Working under budget, time, data, and political constraints*. Thousand Oaks, CA: SAGE Publications, Inc.
- Bhandari, R. and Belyavina, R. (2011). *Evaluating and measuring the impact of citizen diplomacy: Current status and future directions*. New York: Institute of International Education.
- Boeren, A., Bakhuisen, K., Christian-Mak, A.M., Musch, V. and Pettersen, K. (2008). *Donor policies and implementation modalities with regard to international postgraduate programmes targeting scholars from developing countries*. Brussels: Vlaamse Interuniversitaire Raad.

- CARE. (2009). *Girls' leadership development in action: CARE's experience from the field*. Available from: <http://www.care.org/sites/default/files/documents/EDU-GE-2012-Girls-Leadership-Development-in-Action.pdf> (Accessed 1 August 2016).
- Center for Theory of Change. (2016). *Center for Theory of Change home page*. Available from: www.theoryofchange.org (Accessed 1 September 2016).
- Chen, P., Weiss, F. L., Nicholson, H. J. and Girls Incorporated (2010). Girls Study Girls Inc.: Engaging Girls in Evaluation through Participatory Action Research. *American Journal of Community Psychology*, 46(1–2), pp. 228–237.
- Chesterfield, R., and Dant, W. (2013). *Evaluation of LAC higher education scholarships program: Final report*. Washington, DC: U.S. Agency for International Development.
- Chouinard, J.A. and Cousins, J.B. (2009). A Review and Synthesis of Current Research on Cross-Cultural Evaluation. *American Journal of Evaluation*, 30(4), pp. 457–494.
- CIDA. (2005). *Evaluation of the Canadian Francophone Scholarship Program (CFSP), 1987–2005*. Quebec: CIDA.
- Cook, T.D., Scriven, M., Coryn, C.L. and Evergreen, S.D. (2009). Contemporary thinking about causation in evaluation: A dialogue with Tom Cook and Michael Scriven. *American Journal of Evaluation*, 31(1), pp. 105–117.
- Cosentino, C., Rangarajan, A., Sloan, M., Fortson, J., Moorthy, A., Humpage-Liuzzi, S. and Thomas, C. (2015). *Monitoring, Evaluation, and Learning Design for The MasterCard Foundation Scholars Program*. Washington, DC: Mathematica Policy Research.
- Creed, C., Perraton, H and Waage, J. (2012). *Examining development evaluation in higher education interventions: a preliminary study*. London: London International Development Centre.
- Gertler, P.J., Martinez, S., Premand, P., Rawlings, L.B. and Vermeersch, C.M. (2011). *Impact evaluation in practice*. Washington, DC: The World Bank.
- Hansel, B. and Chen, Z. (2008). *AFS long term impact study, 20 to 25 years after the exchange experience, AFS alumni are compared with their peers, Report 1*. New York: AFS International.
- Hesse-Biber, S.N. (2010). *Mixed methods research: Merging theory with practice*. New York: Guilford Press.
- Hofmann-Pinilla, A. and Kallick Russell, J. (2009). *Evaluation of the Leadership Development for Mobilizing Reproductive Health Program, Final Report Executive Summary*. New York: Research Center for Leadership Action.
- Institute of International Education. (2015). *Schlumberger Faculty for the Future: Program impact study*. New York: Institute of International Education.
- Institute of International Education. (2016). *LOTUS Annual Progress Report FY2016*. Cairo, Egypt: IIE Cairo.

- Jadad, A.R. and Enkin, M. (2007). *Randomized controlled trials: questions, answers, and musings* (2nd ed). Malden, MA: Blackwell Publishing.
- Kirkpatrick, D.L. (1979). Techniques for evaluating training programs. *Training and Development Journal*, 33 (6), pp. 78–92.
- Kirkpatrick, D. L. (1994). *Evaluating training programs: the four levels*. San Francisco: Berrett-Koehler.
- Kusek, J.Z. and Rist, R.C. (2004). *Ten steps to a results-based monitoring and evaluation system: A handbook for development practitioners*. Washington, DC: The World Bank.
- Marsh, R., Baxter, A., Cliff, R., Di Genova, L., Jamison, A. and Madden, M. (2016). *Career choices, return paths and social contributions: The African alumni project, Abridged report*. Toronto: The MasterCard Foundation.
- Martel, M. and Bhandari, R. (2016). *Social Justice and Sustainable Change: The Impacts of Higher Education, Ford Foundation International Fellowships Program Alumni Tracking Study Report No. 1*. New York: Institute of International Education.
- Martinez, A., Epstein, C. and Parsad, A. (2015). *Evaluation of the National Science Foundation's Partnerships for International Research and Education (PIRE) Program, Volume 1: Final report*. Cambridge, MA: Abt Associates.
- Mason, L., Powers, C. and Donnelly, S. (2015). *The Boren Awards: A Report of Oral Language Proficiency Gains during Academic Study Abroad*. New York, NY: Institute of International Education.
- Mawer, M. (2014). *A study of research methodology used in evaluations of international scholarship schemes for higher education*. London: Commonwealth Scholarship Commission in the United Kingdom.
- Mayne, J. (2011). Addressing cause and effect in simple and complex settings through contribution analysis, in Forss, K., Mara, M., and Schwartz, R. (eds), *Evaluating the complex: Attribution, contribution and beyond*. New York: Transaction Publishers.
- Perna, L., Orosz, K., Gopaul, B., Jumakulov, Z., Ashirbekov, A., and Kishkentayeva, M. (2014). Promoting human capital development: A typology of international scholarship programs in higher education. *Educational Researcher*, 43(2), pp. 66–73.
- PPMI. (2012). *Interim evaluation of Erasmus Mundus II (2009–2013)*, Brussels: European Commission Directorate-General of Education and Culture.
- Ramboll Management Consulting. (2012). *Final report: Evaluation of NPT and NICHE*. Berlin: Ramboll Management Consulting.
- Research Solutions International. (2016). *The Benjamin A. Gilman International Scholarship Program Evaluation Report*. Washington, DC: U.S. Department of State, Bureau of Educational and Cultural Affairs.

- Rotem, A., Zinovieff, M., and Goubarev, A. (2010). A framework for evaluating the impact of the United Nations fellowship programmes. *Human Resources for Health*, 8(7). Available at: <https://human-resources-health.biomedcentral.com/articles/10.1186/1478-4491-8-7> (Accessed 23 Jan 2017).
- Scott, J., and Carrington, P. J. (2011). *The SAGE handbook of social network analysis*. Thousand Oaks, CA: SAGE publications.
- SRI International. (2005). *Outcome assessment of the visiting Fulbright Student Program*. Washington, DC: U.S. Department of State, Bureau of Educational and Cultural Affairs.
- The MasterCard Foundation. (2016). *Education in Support of Social Transformation: Learning from the First Five Years of The MasterCard Foundation Scholars Program*. Toronto: The MasterCard Foundation.
- Tvaruzkova, M. (2012). *Evaluation: Advances in measuring the effectiveness of networks*. Paper presented at the Global Leadership Consortium, December 11, 2012, Washington, DC.
- USAID. (2004). *Generations of quiet progress: the development impact of US long-term university training on in Africa from 1963–2003, Report prepared by Aguirre International*. Washington, DC: USAID.
- Uyeki, E. (1993). *As Others See Us: A Comparison of Japanese and American Fulbrighters*. New York: Institute of International Education.
- Valuy, E. and Martel, M. (2016). *HER initiative to lead change: The power of education*. New York: Institute of International Education.
- Valuy, E. (2016). *Centroamerica Adelante Final Report*. New York: Institute of International Education.

Mirka Martel is the Head of Monitoring, Evaluation and Learning at the Institute of International Education (IIE). She manages performance and impact evaluations for IIE programs and external clients, including the Ford Foundation, Carnegie Corporation of New York, US Department of State, and the US Agency for International Development. Ms. Martel specializes in quantitative and qualitative evaluation of international education programs, with a specific interest in higher education exchange and leadership. She is the Project Director of a 10-year alumni tracking study of the Ford Foundation International Fellowships Program (IFP), which provided scholarship opportunities to over 4,300 social justice leaders in 22 countries. Prior to joining IIE, Ms. Martel worked for the Aguirre Division of JBS International. She provided technical assistance and evaluation support to education programs in Eastern Europe and Latin America funded by the US Agency for International Development and the US Department of State. Ms. Martel holds a Master's degree in International Affairs from the School of International and Public Affairs, Columbia University.