Evaluating Virtual Exchange

A Toolkit for Practitioners

The Stevens Initiative
September 2019

Evaluation is vital to the success of any educational program, especially when the program is as innovative and dynamic as virtual exchange. As part of its commitment to sharing best practices with the field, the Stevens Initiative worked with its external evaluators at RTI International to develop these suggestions for good practices in evaluating virtual exchange programs, as well as recommended items for surveys of virtual exchange participants, which can be accessed here.

The Stevens Initiative is sponsored by the U.S. Department of State, with funding provided by the U.S. Government, and is administered by the Aspen Institute. It is also supported by the Bezos Family Foundation and the governments of Morocco and the United Arab Emirates.

Why Evaluate a Virtual Exchange Program?

There are many reasons you might want to evaluate your virtual exchange program:

- Clarify your strategy: Designing an evaluation prompts you to identify the goals of your program and the specific outcomes you expect your program to effect.
- Monitor outcomes and impact: Evaluation will help illuminate the effects of your program on participants and other relevant stakeholders.
- Understand and improve program implementation: Through evaluation, you can take stock of program challenges and successes and examine if your program is being implemented as intended.
- Build an evidence base: Evaluation can provide concrete evidence of the benefits of your program. This can help make the case for funding and inform future program design.

General Guides to Evaluation

If you are new to evaluation, the following documents provide a helpful overview of different evaluation methods.

- The W.K. Kellogg Foundation (2017), *Step-by-Step Guide to Evaluation*: This comprehensive and well-organized guide has chapters on evaluation approaches and methodologies, developing evaluation questions and measures, quantitative and qualitative data collection and analysis, communicating evaluation findings, and many other relevant topics. It also includes an extensive list of other evaluation resources.
- The World Bank, Planning, Monitoring, and Evaluation: Methods and Tools for Poverty and Inequality Reduction Programs: This brief toolkit explains various monitoring and evaluation methodologies and includes examples of tools that governments and international organizations have developed to put them into practice. For a more in depth treatment please visit Ten Steps to a Results-Based Monitoring and Evaluation System.
- The Organisation for Economic Co-operation and Development (OECD), *Outline of Principles of Impact Evaluation*; The World Bank (2006), Conducting Quality Impact Evaluations Under Budget, Time, and Data Constraints; and Rogers, P. J. InterAction, (2012). *Introduction to Impact Evaluation*: These user-friendly guides offer advice for designing and implementing impact evaluations.

Decide What to Measure

Evaluation should be built in to the design and planning phase of any program creation and prioritized from the outset. Early in the evaluation process, you will want to specify what you want to measure. What are the key questions about your program that you and your key stakeholders –the program participants, facilitators, funders, or others – are most interested in answering? Identify and develop these questions before moving on to the collection method. Below are some examples of general guiding questions you might consider in evaluating a virtual exchange program:

- **Demographics:** Who participated in the program? What were the demographic characteristics of the participants? If the program was aimed at participants with certain characteristics, such as those from a specific socio-economic background, did it ultimately reach those participants?

- **Program Completion:** What was the target number of participants for the program? How many actually started the program? What percentage of people who started the program completed the program or reached a meaningful threshold of participation? When did participants drop out of the program, and what caused this attrition?
- **Facilitator Training:** Who were the educators and facilitators involved in the program? Did they have the intended qualifications? To what extent did they receive the intended training? Were they effective in their facilitation role?
- **Program Implementation:** Was the program delivered in the way it was intended to be delivered? If not, how did it deviate from the model and why? When did the change happen and did that affect the results of your program?
- **Challenges:** What challenges did educators, facilitators, and administrators face in implementing the program? (Examples of categories of potential challenges you might want to explore include technology issues; participant recruitment and engagement; educator recruitment and training; curriculum development; program scheduling, timing, and other logistics; and language barriers among participants.) How did they address those challenges?
- **Effective Practices**: What practices or strategies in implementing the virtual exchange program seemed to be most successful? How were they successful?
- **Participant Satisfaction:** To what degree were participants satisfied with the program overall? To what degree were they satisfied with specific aspects of the program (e.g., individual curricular modules; synchronous or asynchronous exchange components)? What are their recommendations for improvement?
- **Outcomes:** What outcomes were expected for participants and, if applicable, educators, facilitators, institutions, or other stakeholders? What outcomes did the program have?

Identify Participant Outcomes to Measure

Measuring participant outcomes will likely be a key feature of your evaluation. Therefore, you want to think early about what outcomes you realistically expect to effect. The following are some examples of types of outcomes that might be of interest in the context of virtual exchange:

Global Competencies: The Stevens Initiative has identified a list of key global competencies that commonly serve as intended outcomes for many virtual exchange programs. This list is not exhaustive but may be a useful starting point for thinking about the types of domains you might want to measure. It builds on and adapts to the virtual exchange context frameworks developed by the Asia Society, Harvard University Graduate School of Education, OECD, and U.S. Department of Education, among others. Visit the Stevens Initiative Resources page to learn more about these frameworks.

- Cross-cultural communication: The ability to communicate effectively (both orally and in writing) with individuals who come from different cultural backgrounds.
- Cross-cultural collaboration: The ability to work productively with individuals from different cultural backgrounds.
- Empathy: The ability to share and understand the feelings of others, particularly those with different cultural backgrounds.

- Perspective taking: The ability to understand and perceive a situation or concept from the point of view of others, particularly those with different cultural backgrounds.
- Knowledge/knowledge-seeking: Knowledge, or interest in gaining knowledge, about another culture or country.
- Critical thinking: The ability to objectively and skillfully analyze information and form a reasoned judgment.
- Problem solving: The ability to identify and understand problems and develop solutions by systematically applying appropriate skills.

Substantive Learning: Many virtual exchange programs have a substantive topic focus and associated learning goals (e.g. engineering, journalism, or public health). Consider what, specifically, you expect your participants to learn or what proficiencies you expect them to develop.

Behavior changes: Consider whether you expect participants to take certain actions as a result of participating in your program. For example, do you expect them to be more engaged in civic organizations? Do you expect them to speak out on a specific issue or topic in their communities?

Decide How to Collect Your Data

Once you have determined the domains and outcomes that interest you, the next step is to consider how to measure them. You do not need to use all of the quantitative and qualitative methods listed below, though it is appropriate and even desirable to use multiple methods in order to have the best possible understanding of your program and its outcomes. Each program should select evaluation methods that are appropriate for their domains and program structure. You should also consider how much can realistically be measured in one evaluation without overly burdening participants. For a discussion of the advantages and disadvantages of quantitative and qualitative methods, see chapter eight of The W.K. Kellogg Foundation (2017), *Step-by-Step Guide to Evaluation*.

Surveys

Surveys can be an effective way to measure participant demographics, program outcomes, stakeholder satisfaction, and recommendations for improvement, among other topics. Surveys allow you to efficiently obtain information from a large number of people (e.g. all participants). Surveys can contain closed-ended questions, open-ended questions, or both. Closed-ended questions present respondents with specific answer options while open-ended questions provide space for any response. When closed-ended questions are used, answers are quantifiable and can be analyzed using statistics. Open-ended survey items can provide useful qualitative data in a written format, although large numbers of survey responses can be time-consuming to analyze.

The following document contains detailed guidelines for designing survey items: Irwin, C.W. & Stafford, E.T. (2016). Survey Methods for Educators: Collaborative Survey Development (part 1 of 3) (REL 2016-163). U.S. Department of Education. You may also want to review the literature to locate existing survey measures for specific domains that may be appropriate for your program. A few examples of existing survey measures include the PISA Global Competence Questionnaire and Soliya Impact Assessment Tools.

The Stevens Initiative has adapted and devised several survey measures of global competencies for its awardee programs. Those measures are shared alongside this report.

Considerations for designing and collecting surveys for virtual exchange programs:

- **Survey design and timing:** Determine how many surveys you will need to administer and when. Ideally, you would use a pre- and post-program survey design, where you administer surveys immediately before and immediately after the program to assess changes in participant responses over time. If a pre-program survey is not feasible, you may choose to only administer a post-program survey and ask participants to reflect and report on changes as a result of the program. You may also want to supplement conventional post-program survey questions with retrospective questions that ask participants to think back to the start of the program and assess where they were on a given dimension before the program started. If feasible, you may also consider doing a follow up survey after some time has passed following the end of the program to assess longer-term outcomes.
- **Matching de-identified responses:** A good practice in evaluation is de-identifying participant information to grant anonymity and encourage honest feedback. If you are conducting pre- and post-program surveys, it is nevertheless imperative that you be able to match an individual participant's responses on the pre-program and post-program surveys. When you analyze responses to matched questions on the pre- and post-program surveys, you should limit your analysis to participants who completed that question on both surveys, so you need to be able to track who responded and which post-program response to compare with which pre-program response. It is best to use a unique identifier that is not a participant's name or other Personally Identifiable Information (PII), such as birth date or identification number, to protect participants' information and anonymity. If conducting surveys through an online system, you may be able to generate a unique link for each participant, which serves as an effective way to track and match participants' responses.
- **Response rates:** Low survey response rates can undermine the validity of your findings. Some strategies to help improve the chances of achieving high response rates include sending multiple reminders to participants to complete the survey, granting time and space during the program (e.g. the last program session) for survey completion, embedding the survey in your virtual exchange platform and prompting participants to complete the survey after logging in, explaining clearly what the survey will be used for and that it is anonymous, providing incentives for survey completion (e.g. entry in a raffle, a gift card), and keeping the survey short so that it is not onerous to complete.
- **Age-appropriateness:** Consider the age level of your participants in designing survey questions. Questions that may be appropriate for university students may not be appropriate for middle school students.
- **Translation:** You may want to offer a version of your survey in the native language of each participating classroom or group to ensure respondents understand the nuance of the questions being asked.

Qualitative Methods

Qualitative methods, including interviews, focus groups, and program observations, can be valuable tools in evaluating virtual exchange programs. They can add context and detail to your quantitative findings, help you obtain critical feedback on your program, and allow you to probe deeply on important topics and ask "how" and "why" questions. They can also help you understand whether your program model is being implemented with fidelity, and what challenges practitioners are facing.

In its evaluation of the Stevens Initiative, RTI has developed interview and focus group protocols, which will soon be released on the Stevens Initiative website. RTI is also presenting a Stevens Initiative webinar on qualitative methods in conjunction with the release of this publication. (Visit the Initiative's YouTube channel to watch a recording of the webinar.)

Considerations for qualitative data collection:

- **Informed consent:** Explain the purpose of the focus group or interview and get explicit participant consent. If you are going to record the focus group or interview, be sure to explain what the recording will be used for and obtain explicit consent to record. Consent can be gained verbally or through a written consent form at the start of the focus group or interview.
- **Multiple perspectives:** You may want to consider conducting interviews and focus groups with different categories of program stakeholders (e.g. participants, educators, facilitators, school administrators) to get feedback from different perspectives.
- **Different regions:** If possible, you should conduct interviews and focus groups with participants and educators in the different places or regions that participate in your virtual exchange. Consider conducting interviews and focus groups with youth in their native language, even if your program is conducted in a different language. This may be particularly important for secondary school level programs, where foreign language proficiency may be lower, but may be appropriate at the postsecondary level, as well. Participants may feel freer and more comfortable giving nuanced and critical feedback in their native language.
- **Critical feedback:** Interview and focus group participants may be reluctant to provide feedback that's critical of the program. Take steps to reassure them that critical feedback is welcomed. To the extent possible, when conducting participant focus groups, have authority figures (e.g. educators, facilitators, or administrators) leave the room so participants can speak freely. Tell participants their feedback will be used to help improve the program, and feedback will be left anonymous: no names will be linked with any comments that are ultimately reported outside the focus group or interview.

Once you have collected qualitative data, you should analyze it systematically. Chapter 8 of The W.K. Kellogg Foundation (2017), *Step-by-Step Guide to Evaluation* provides some helpful tips on coding and summarizing qualitative data.

Reflection Essays

Asking participants, educators, or facilitators to write short essays in response to specific questions at different points during the program has the potential to provide rich and detailed information about their experience and what they learned in the program. Keep in mind that writing these essays can be burdensome – which can result in low response rates – and analyzing the data can be time-consuming, too. See this report from SUNY COIL about how they used participant reflection essays as part of their evaluation during their Stevens Initiative award.

Exams or Other Assessments

To assess participant learning on the substantive topic of the exchange, you may want to incorporate traditional exams or other assessments, such as portfolio evaluations or evaluations of final projects. As with other methods, think ahead about how you will analyze and report on the results of the assessment and final projects. For example, reporting on the substantive knowledge and skills participants gained from the program is difficult to do without also assessing participants' knowledge and skills before the program starts (which might not be feasible) unless you are confident the topic and skills are new to all or most students. Training and providing a rubric to those who will review these assessments will help you make sense of the findings from those evaluations.

Administrative Data Collection

Consider whether and how you might be able to use participant data you collect in the course of conducting your program. Do participants apply to your program, and if so, what information do you request on your application form? For example, you may already know the gender breakdown of your participants, and therefore do not also need to collect that on a survey, unless understanding survey responses by gender is important. Be thoughtful, however, about asking for information that is sensitive or that youth might not be able to reliably report. Young people may not be a reliable source about their family income, and this is a sensitive topic, as well. If one of your program goals is to serve typically underserved youth, think about how you might be able to track and report this in other ways. Can you collect data about the institutions or communities where your program is offered to shed light on family income or other relevant indicators?

Control or Comparison Group

Consider incorporating a control or comparison group into your evaluation design. This would be a group of youth who are similar to the youth in your program but who do not participate in the program. It can be hard to know if the changes reflected in evaluation are due to outside events or participation in the virtual exchange. To help isolate the effects of the exchange, tracking the perceptions of youth who are exposed to the same outside events, but don't participate in the virtual exchange, helps contextualize findings.

Identifying and collecting data from a control or comparison group can be challenging. Randomized control trials in particular need to meet a very high standard in order to produce valid results. Consider whether aiming for this standard is necessary to improve your understanding the effect of your program. And consider whether there are conditions that would make it practical to form control or comparison groups. Is there a classroom at the same institution or a nearby institution that serves demographically similar students to the classroom of students participating in the virtual exchange who could also take your surveys? If more young people want to participate in the program than can be accommodated, can you randomly assign some of them to the program and the others to a control group? If you can offer control or comparison group members a spot in a future cohort of the program, this might allay concerns about fairness.

For more information about control and comparison groups, including randomization, see The World Bank (2006), Conducting Quality Impact Evaluations Under Budget, Time, and Data Constraints. See also Dunning, Thad. *Natural experiments in the social sciences: a design-based approach*. Cambridge University Press, 2012 and Cook, Thomas D., Donald Thomas Campbell, and Arles Day. *Quasi-experimentation: Design & analysis issues for field settings*. Vol. 351. Boston: Houghton Mifflin, 1979.

Analyzing and Reporting Data by Country/Region

Consider segmenting your data by country or region when you analyze and report it. This will enable you to examine whether and how the results differ across these groups. For example, knowing that youth in one country experience larger gains in an outcome of interest or report higher levels of satisfaction with the program can help you to think about how to refine the program so that youth from all countries benefit.

The Stevens Initiative and RTI believe robust evaluation, designed in advance, is vital for virtual exchange programs. Thoughtfully deciding what and how to measure will help you demonstrate the impact of your program and will help the broader virtual exchange community continue to make the case for this emerging field.

The Stevens Initiative is an international effort to build global competence and career readiness skills for young people in the United States and the Middle East and North Africa by growing and enhancing the field of virtual exchange. It is a lasting tribute to Ambassador J. Christopher Stevens, a public servant who dedicated himself to building understanding between people from different countries. The Stevens Initiative is sponsored by the U.S. Department of State, with funding provided by the U.S. Government, and is administered by the Aspen Institute. It is also supported by the Bezos Family Foundation and the governments of Morocco and the United Arab Emirates.

The Aspen Institute is an educational and policy studies organization based in Washington, DC. Its mission is to foster leadership based on enduring values and to provide a nonpartisan venue for dealing with critical issues. The Institute is based in Washington, DC; Aspen, Colorado; and on the Wye River on Maryland's Eastern Shore. It also has offices in New York City and an international network of partners. For more information, visit www.aspeninstitute.org.

Stevens Initiative

www.stevensinitiative.org stevensinitiative@aspeninst.org @StevensInit /StevensInitiative