Quantifying and Assessing Trends on the National Science Foundation's Broader Impact Criterion

Dina Verdín, Purdue University

Dina Verdín is a Ph.D. student in Engineering Education and M.S. student in Industrial Engineering at Purdue University. She completed her undergraduate degree in Industrial and Systems Engineering at San José State University. Dina is a 2016 recipient of the National Science Foundation's Graduate Research Fellowship (GRF). Her research interest focuses on first-generation college students, specifically around changing deficit base paradigms by providing asset base perspectives for understanding this community.

Quantifying and Assessing Trends on National Science Foundation's Broader Impact Criterion

The American Innovation and Competitiveness Act (S.3084) reapproved the National Science Foundation's (NSF) merit review criteria i.e. Intellectual Merit and Broader Impacts, called for an update of the policy guidelines for NSF staff members and merit review process participants, and emphasized the importance of transparency and accountability. Evaluating Project Summaries based on Intellectual Merit and Broader Impacts has been the standard of maintaining excellence and accountability since 1997. Intellectual merit consists of proposing activities that advance knowledge, while Broader Impacts statements, "describe the potential of the proposed activity to benefit society and contribute to the achievement of specific, desired societal outcomes." While Intellectual Merit has been widely understood since its inception, the Broader Impacts criterion has undergone continuous discussion among Principle Investigators (PI), panel reviewers, NSF Committees of Visitors, and the broader scientific community.

To better understand Broader Impacts and how they could be quantified, I conducted an exploratory study during a 10-week internship in the Engineering Education and Centers (EEC) division of the Engineering (ENG) Directorate at NSF. This study used a convergent parallel mixed method design, collecting both quantitative and qualitative data, simultaneously, to answer two research questions 1) What trends are Program Officers seeing in the Broader Impacts criterion and 2) Which Broader Impacts statements are being addressed in Project Summaries submitted to the National Science Foundation.

The quantitative approach consisted of examining 82 awarded Project Summaries in the EEC division to obtain a quantifiable assessment of the extent to which PIs who applied to EEC addressed the Broader Impacts suggestions outlined in NSF's *Proposal and Award Policies and Procedures Guide*. The qualitative approach involved interviews of four program officers from the EEC division regarding the trends in addressing the Broader Impacts criterion. The data were collected in parallel strands, independently from each other, and were brought together for comparison.

Results from the examined abstracts indicate that PIs were more likely to have Broader Impacts statements around increasing public scientific literacy, public engagement in science and engineering, and addressing issues of developing a diverse STEM workforce. While the interview data from Program Officers indicated that PIs were more inclined to write "pie in the sky" statements around Broader Impacts. Program Officers also indicated that reviewers tend to place more weight on Intellectual Merit than Broader Impacts even though they are informed to put equal weight in both criteria. There exists no quantitative metric for measuring or assessing Broader Impacts statements PIs propose in their NSF awarded grants. This study is an exploratory attempt to unpack what is currently being funded using awarded Project Summaries and outline tensions around addressing Broader Impacts by proposing a possible quantitative metric for measuring Broader Impacts activities.

Background

The National Science Foundation (NSF) from its inception in 1950[1], has remained, to date, the Federal government's principle steward for research and education in science and engineering (S&E) fields. NSF is recognized as "the Nation's premier agency supporting basic research and education in mathematics, science, engineering, and technology"[2] (p. 3). The agency helps provide the research infrastructure and educational opportunities to ensure an innovative and productive science and engineering enterprise, and supports the advancement of our nation's most pressing challenges, as reiterated in the most recent strategic plan "Empowering the Nation through Discovery and Innovation—NSF Strategic Plan for Fiscal Years 2011-2016." The strategic plan supports NSF's mission, providing programmatic and operational underpinnings, and setting the context for a broad, balanced NSF portfolio[2,3].

To maintain its high standards of excellence and accountability, in 1997, the National Science Foundation incorporated a policy that would evaluate proposals submitted to the agency under two Merit Review criteria: Intellectual Merit and Broader Impacts, replacing the four criterion system established in 1981[4,5]. As noted by former NSF Director Arden Bement, "The [Broader Impacts] criterion was established to get scientists out of their ivory towers and connect them to society"[6]. The 1997 criteria, publicized by the National Science Board via *Important Notice 121, New Criteria for NSF Proposals*[7], comprised of a set of contextual elements aimed at assisting the grant proposer and reviewer stating:

What are the Broader Impacts of the proposed activity?

- How well does the activity advance discovery and understanding while promoting teaching, training, and learning?
- How well does the proposed activity broaden the participation of underrepresented groups (e.g., gender, ethnicity, geographic, etc.)?
- To what extent will it enhance the infrastructure for research and education, such as facilities, instrumentation, networks, and partnerships?
- Will the results be disseminated broadly to enhance scientific and technological understanding? What may be the benefits of the proposed activity to society? (NSF, 1997)

Similarly, in the 2016 *Proposal and Award Policies and Procedures Guide*[8], NSF provided a discussion of what may constitute as Broader Impacts, which coincides with the description provided in the 1997 *Important Notice 121, New Criteria for NSF Proposals*[7].

The restructuring of the merit review criteria, to include Broader Impacts, was in response to the need to demonstrate the societal benefits accomplished by NSF-funded projects—"This [Broader Impacts] criterion can be traced to NSF's congressional mandate (the Science and Engineering Equal Opportunities Act of 1980, last amended in December 2002) to increase the participation of underrepresented groups (women, minorities, and persons with disabilities) in STEM (science, technology, engineering, and mathematics)" [9] (p.4). Years after the inception of the two merit review criteria, it became clear that proposals were not addressing both Intellectual Merit and Broader Impacts. Thus, in 2002, in a letter from the Office of the Director, the National Science Foundation announced proposals that did not explicitly address both merit review criteria would be returned without review[10]. This amendment to the proposal granting process emphasized NSF's

commitment to funding projects that connect science and engineering to society, via broader impacts.

In 2010, the America COMPETES Act[11] reauthorized the National Science Foundation, mandating the retention of the Broader Impact criterion, and requested that NSF issue a report to Congress regarding the effects of this criterion. One of the requirements requested in the report by Congress was to provide

> evaluations performed by the Foundation to assess the degree to which the Broader Impact aspects of research proposals were carried out and how effective they have been at meeting the goals described in the research proposals"[11] (p. 686).

This report was Congress' way of adding accountability to the funds dispersed to NSF, ensuring that awards were being given to proposals that were used to benefit society and meet national goals.

Challenges with NSF's Broader Impact Criteria

In a National Science Board (NSB) meeting held on February 2010, a Task Force was assembled to address the Merit Review Criteria—Intellectual Merit and Broader Impacts. The Task Force was "charged with examining the two Merit Review Criteria and their effectiveness in achieving the goals for NSF support for science and engineering research and education"[5] (p. 32). Preliminary findings revealed that the Intellectual Merit criterion was well understood among the community of scholars and NSF personnel, however, the Broader Impacts criterion was "not generally well understood." Thus in an attempt to shed light on what is understood around the NSF Merit Review criteria, a systematic review was conducted, to gather data through surveys and conduct an analysis from key stakeholder groups i.e. "Principal Investigators (PIs) and institutions that submit proposals for NSF research and education grants, reviewers of proposals, NSF staff—including Senior Leadership, Division Leadership, and Program Directors—and Advisory Committee Members" [2] (p. 6).

The findings taken from the survey of over 4,516 participants and a selected group of interviews revealed six major themes. Specifically, for the Broader Impacts criterion, the study found a lack of understanding of the criterion and calls for improved guidance were needed, "only 32% of NSF Officials and 19% of NSF Advisory Committee Members felt that reviewers had a very high level or high level of understanding of the Broader Impacts criterion, and only 22% and 15% of Officials and Members respectively felt principal investigators had a very high level or high level of understanding"[2] (p. 46). Among the stakeholders, there was a greater agreement that Intellectual Merit should be weighted greater than Broader Impacts—it should be noted that NSF requires that both Intellectual Merit and Broader Impacts be *equally* weighted in the panel review process. This highlights a tension between what stakeholders deem valuable in NSF proposals and what the Foundation instructs both PIs and Reviewers to focus on. When stakeholders were questioned about the role institutions should play in supporting Broader Impact activities, "70% of respondents to the *Survey of NSF Officials and the Advisory Committee Members* indicated that principal investigators' institutions should do "much more" or "somewhat more" to support Broader Impacts activities" [2] (p.52). Lastly, there was a high indication that the post-award assessment of Broader Impacts activities was weak and needed improvement.

In 2013, the EEC division Committee of Visitors (COV)—a committee of external experts who provide feedback to the division regarding the quality and integrity of the merit review process—stated in their report to EEC, "the Broader Impact statements in proposals were often "open-ended," highly qualitative and indefinite with little quantitative data"[12] (p. 10). Similar observations were made and brought to the attention of EEC in previous COV reports. A study by Nagy[13] investigating factors that influenced response to the Broader Impacts criterion surveyed 124 NSF grant recipients—50% full professors, 32.2% associate professors, 11.3% assistant professors, and 5.7% other. The study revealed that while the principal investigators self-reported having a "good deal of experience" conducting Broader Impact activities, there was essentially "little training in community engaged activities and [they] lacked confidence in their ability to evaluate such work"[13] (p.82). The study also found that while institutions publicly endorsed Broader Impact related activities (for example, community engagement), the tenure and promotion policies "did not recognize such activities" nor was there sufficient funding and infrastructure to support their Broader Impact activities provided[13] (p. 82).

Another study sought to examine the attitudes and understandings of Broader Impacts criterion of 31 faculty members by looking at NSF Faculty Early Career Development (CAREER) awardees, from the Engineering Directorate, at four high research institutions. The CAREER Program is a National Science Foundation-wide activity that offers the "most prestigious awards in support of junior faculty who exemplify the role of teacher-scholars through outstanding research, excellent education and the integration of education and research within the context of the mission of their organizations" [14] (p.2). Faculty in this study were asked how they felt about NSF's Broader Impact criterion, of which 80% viewed the criterion favorably. However, despite favorable feelings about the Broader Impact criterion, only half of these faculty members "would continue with the same level of involvement in socially relevant activities if it was not required by NSF"[15] (p. 69). Hallienn[15] also found there was a lack of feedback through NSF regarding the Broader Impact activities, suggesting through her findings that faculty required guidance, presumably from the agency, to help facilitate their Broader Impact strategies. Other concerns about the Broader Impacts criterion arose from Roberts[16] study highlighting that proposals overstate the potential societal benefits. The study points to the absence of a dissemination plan, suggesting that the research findings may only benefit the scientific community and by not going beyond this community the project may potentially fail to benefit society.

This document has highlighted a few studies on the scholarship of Broader Impacts, however, in conversations with those in the EEC Division it became clear that this was worthy of further exploration. The uncertainty around Broader Impacts ranges from an underlying misunderstanding of what the criterion means to a lack of understanding of its importance. While I do not claim to have expert knowledge on the Broader Impacts criterion, I attempted to investigate how closely NSF-funded Principal Investigators are aligning their Broader Impact statements with the suggested activities outlined in the 2016 *Proposal and Award Policies and Procedures Guide*[8].

Motivation and Framing of the Study

In the summer of 2016, through the Hispanic Association of Colleges and Universities (HACU), I was granted the opportunity to be a Summer Scholar at the National Science Foundation, specifically in the Division of Engineering Education and Centers (EEC). My internship at NSF gave me an opportunity to learn about the process of how proposals are awarded, the policy and

procedures behind granting awards, and how submitted proposals are reviewed. During my time in EEC, the Division was undergoing their three-year evaluation by an external Committee of Visitors. This process is intended to assess the quality and integrity of the program operations and management, and the division's contribution towards NSF's mission and strategic goals. I quickly became fascinated by the challenges in how to assess the Broader Impact Merit Review Criteria after speaking with several Program Officers and reading the 2013 report from the EEC COV. The EEC COV report echoed some of the concerns Program Officers were hearing from the PI community around the Broader Impact criterion. Thus, with the support of my summer mentors, I decided to conduct an exploratory study to gain insight into how the PI community was addressing and proposing to assess Broader Impacts in their awarded proposals. The study would be conducted within a 10-week period given that, that was the timeframe of my internship at the National Science Foundation. Given the accessibility limitations that I had, I decided to conduct a mixed method study to obtain information from Program Officers and from the one-page summary of awarded proposals. I understood this study was going to have several limitations since I was not able to review full proposals; however, my intention was to contribute to the conversation NSF and the PI community were having around Broader Impacts. Therefore, it should be noted that the analysis in this study is not intended to be a full representation of the EEC Division nor of NSF, but rather a glance into how a subsection of awarded proposals are addressing the Broader Impacts criterion. Additionally, while this was an independent research study I received continuous support and direction from my NSF EEC mentors. However, the discussion and conclusions stated in this report are mine alone.

Research Design

As outlined in the introduction section, there is an opportunity for improvement in the understanding of what constitutes Broader Impacts, how to assess Broader Impacts, the value placed on Broader Impacts, and the importance the community places on this criterion. Two research questions motivated this study,

> RQ1. What trends are a subset of Program Officers in EEC seeing in proposals submitted to their division pertaining to the Broader Impacts criterion?
>
> RQ2.Which Broader Impacts statements are being addressed in a sample of proposals submitted to and awarded by EEC?

To answer these questions, I conducted a small-scale exploratory study using a convergent parallel mixed method approach because it allowed me to quantitatively assess if the Broader Impact categories were being addressed, qualitatively understand overall trends in the proposals, and use both methods to outline potential avenues for measuring Broader Impacts. In research question one, the intention was to gather information about Broader Impact trends that Program Officers have seen in the programs they manage. While research question two was limited to Project Summaries. Therefore, in the analysis of research question one, the discussion will be around proposals, while research question two will only address appraised awarded Project Summaries.

Method

I employed a convergent parallel mixed method research design, collecting both quantitative and qualitative data simultaneously[17]. This method was selected because it provided a way to develop a complete understanding of the Broader Impact Criterion using different but complementary datasets. Figure 1 best outlines the use of a convergent parallel mixed method research design,

depicting the collection of two independent strands of data—quantitative and qualitative—simultaneously. The data were collected in parallel strands, independently from each other, and were be brought together to compare the results.

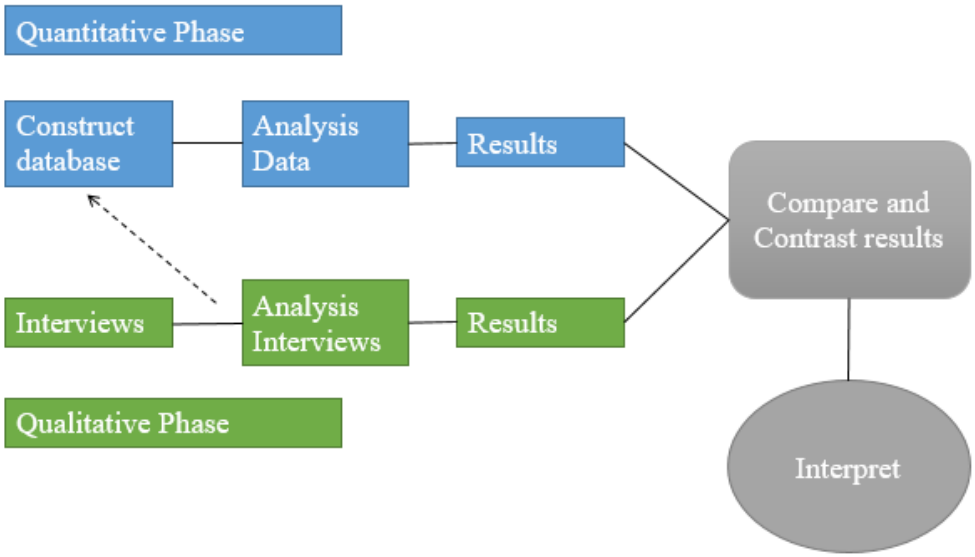

Figure 1 Convergent parallel Mixed Method approach

As the quantitative database was being constructed using Microsoft Excel, I was scheduling and conducting interviews with Program Officers. Quantitative and qualitative data was analyzed separately, "one does not depend on the results of the other"[17] (p.78). The rationale for selecting this approach was to "triangulate the methods by directly comparing and contrasting quantitative statistical results with qualitative findings for corroboration and validation purposes" [17] (p. 77). In some cases, data from the interviews help shape the inclusion and exclusion criteria for the quantitative phase. While both datasets were collected simultaneously, the qualitative phase will be presented first followed by the quantitative phase.

Qualitative Method

To assess *RQ1. What trends are a subset of Program Officers in EEC seeing in proposals submitted to their division pertaining to the Broader Impact criterion?*, the data came from four semi-structured interviews with Program Officers in EEC. The interview protocol served to provide structure, consistency among each participant, and to protect the overall arrangement and purpose of the interview (McCracken, 1988). A purposeful sampling technique was used, which involved "selecting certain units or cases based on a specific purpose rather than randomly"[18] (p.80), that is, participants were recruited from the same Division as the Project Summaries sampled in the quantitative phase. Four participants were selected to cover a range of programs across the EEC Division. Selecting participants from the same Division as the sampled Project Summaries allowed me to compare, contrast, and help inform some of the quantitative findings. The author would like to emphasize that throughout the interview Program Officers were only asked about trends within their program portfolio, interview questions regarding personal perceptions about Broader Impacts were not asked. Additionally, the responses from the four Program Officers is not a comprehensive representation of NSF's portfolio.

Analysis

Participants were asked to answer open-ended questions centered around three major topics, 1) Broader Impacts trends among EEC applicants, 2) Broader Impacts trends in evaluation methods, and 3) Broader Impacts knowledge of review panelist. These major themes were selected from prior literature review. The goal was to understand how the qualitative findings in the three general themes compared with the quantitative findings, thus an *a priori* coding approach was appropriate. The interview responses were analyzed based on the themes using thematic analysis. The three themes are presented as subsections below.

Broader Impact Trends among Applicants

It's been a long-standing issue within the PI community on how to address the Broader Impact criterion. Interviews with the four EEC Program Officers revealed several trends, first, it was evident that there were still misunderstandings on what constitutes Broader Impacts, noting "some people see Broader Impacts as being, they use it interchangeably with broadening participation." While certain portfolios within the EEC Division focus on broadening participation, outside of this specific program emphasis, "you can do Broader Impacts without even impacting communities." The idea of separating broadening participation, still considered an important task, with Broader Impact remains a challenge with Principle Investigators. PIs may be inclined to connect broadening participation with Broader Impacts because of the suggested outcomes highlighted in the *Proposal and Award Policies and Procedures Guide*.

When asked about the type of statements PIs employ in their proposals Program Officers spoke about vague claims or "cookie cutter language." Most specifically one Officer stated,

> … Broader Impacts come off as this pie in the sky, someday, somewhere down the road, the results of my project, because I'm talking about diversity, could lead to a more diverse workforce for the increased competitiveness for the United States.

Specifically, Program Officers noted that many proposals assert bold and often unmeasurable statements about how one project can advance the entire nation. Claiming national impact or advancement would require Principle Investigators to conduct longitudinal impact studies spanning, for example, 5 or more years and in multiple locations, however, proposals seldom have this level of detail. There is currently a lack of accountability towards longitudinally measuring and tracking how a study has impacted a community or group of students. Program Officers acknowledge the difficulty establishing a one-to-one comparison of how awarding X number of dollars will translate to an equitable outcome.

There is currently no policy explicitly requiring a measurable, "one-to-one correspondence," however, Program Officers have been advising the community of PIs on moving towards measurable and scalable statements of Broader Impact outcomes. There is a general understanding from both PIs and Program Officers of the ambiguity around assessing or measuring *benefits to society.*

Lastly, Program Officers spoke about Broader Impacts as reaching non-academic spaces, in addition to the academic spaces. For example, "putting a briefing together, a policy brief that could be posted on one of the learning societies websites … that sometimes have literally thousands of people…" In some cases, as described by one Program Officer, these alternative forms of

dissemination could have greater circulation than a journal publication. The Program Officer goes on to suggest Principal Investigators should go the "extra mile to make sure that the information is disseminated to the spaces that can most benefit from it." In general, Program Officers agreed that the extent to and manner which PIs address Broader Impacts can vary greatly. Trends in the different programs revealed that Principal Investigators often use far reaching claims rather than measurable outcomes.

Broader Impact Evaluation Methods

Program Officers were asked if the Broader Impact criterion could be evaluated similar to Intellectual Merit, there was a general agreement that it should, stating,

> Yeah, they should be. Again, the idea that you're doing actual activities … If you have the model of "Someday, maybe, U.S. will be more competitive because of my little project." No. That's not the kind of Broader Impacts we want. We want those specific activities so you can say, even simple, "I reached this many people," or whatever. That's probably too simple a venture, but at least it's something.

Another Program Officer added, with respect to assessing activities, "you could do some pre/post test. Some quasi-experimental designs. 'Prior to participation, they [students] had this belief, after participation, they [students] have this' …" These are examples of Broader Impact evaluation methods that are not typically seen in proposals, it's not all-encompassing, though it demonstrates a shift in how Principal Investigators can start evaluating their Broader Impact statements.

Program Officers were asked if there were any trends in how Principal Investigators were proposing to measure Broader Impact, the response was as follows,

> No, that's a very rare thing and probably something that we should do more of. There's very little of that … How do we actually determine the impact of the projects? We don't have a good way of doing that.

While Program Officers are aware of the importance of measurable outcomes, in most programs they have not seen trends in Broader Impact evaluation methods. An exception to this is the Research Experiences for Undergraduates (REU) and the Research Experience for Teachers (RET), housed under the Workforce Development Cluster. After the completion of the award, PIs are asked to submit specific information regarding broader impacts. Principal Investigators who are awarded an REU are required to longitudinally track how many of their participants went off to enroll in a graduate program or published scholarly work. Similarly, regarding the RET program, one Program Officer stated,

> I've seen data on impacts that the program has had on the teachers, on their career, on the way they teach ... A lot of this is … about how their students are now looking at engineering and science… some of it is reflection …

This method of evaluation can serve as a sort of post-test of how the teacher's research experience impacted his/her teaching and the students' engagement in the material. This is not to say REUs and RETs are models for how Broader Impacts could be measured as they also come with flaws. That is, Principal Investigators need multiple forms of evidence to suggest that REUs lead to students enrolling in graduate studies. Additionally, assessing the impact an RET and/or REU had

on its participants would require the Principal Investigator to embed this as a research question in the overall project scope.

Panel of Reviewers' Knowledge of Broader Impacts

It should be first acknowledged that there is a tremendous amount of guidance regarding the merit review criteria given to panelists in the form of briefing materials, slides, webinars, etc. During the interviews, Program Officers were asked how the panel of reviewers' knowledge on Broader Impact was assessed prior to their selection. It was agreed upon that the reviewers were selected

> …based on their technical expertise, in terms of what topics they are researching, what methodologies they are associated with, things like that… which is really thought of in terms of their Intellectual Merit expertise.

This approach highlights the National Science Foundation's mission "to promote the progress of science ..."[19], however, with the 1997 move to emphasize more on Broader Impacts there needs to also be a shift in the reviewer selection process.

Findings have revealed an explicit effort from the Program Officers to shift from an emphasis solely on high Intellectual Merit to an equal emphasis on both Intellectual Merit and Broader Impact. Although reviewers are encouraged to make the shift in their assessment of proposals, one Program Officer points out that reviewers tend to think about Broader Impact as

> … a pass or fail… as long as you really didn't flunk the Broader Impact section then you passed it … differences are made based on Intellectual Merit, that's where the PIs are much more grip and much more solid, evidence-based statements they can make…

Thus, I observed that Program Officers attempt to shift the review process to have a stronger emphasis on Broader Impacts, may prove challenging to the community of reviewers who are most knowledgeable in assessing Intellectual Merit. Program officers remind the panel of reviewers that the Division places equivalent weight on Broader Impact in addition to Intellectual Merit in a briefing prior to the start of review discussions. However, panelists are not always clear on how to review both criteria equally. To ensure proposals are reviewed with a high emphasis on Broader Impact, one Program Officer states,

> … it's mostly educating them, both P.I. community and the reviewers, about what Broader Impacts are and the kinds of things we're looking for … It's just being sure people are aware.

The shift in mindset, placing emphasis on both high-quality Broader Impact as well as Intellectual Merit, will be a challenge for most reviewers for several reasons: 1) reviewers will need to be aware of how to identify a Broader Impact statement that is not a 'pie in the sky' claim, 2) reviewers may require information about what measurable Broader Impact outcomes entail, and 3) Program Officers may be required to scaffold reviewers in assessing measurable Broader Impact outcomes. These are a few of the possible challenges Program Officers may face. The biggest challenge and the one that should be well understood first is how to assess and potentially measure Broader Impacts.

Quantitative Method

Keeping true to the convergent parallel mixed method approach described in Figure 1, I simultaneously scheduled interviews with Program Officers and examined Project Summaries to construct a database of Broader Impact statements. Gathering Project Summaries from the COV 2016 portfolio would allow me to answer RQ2. *Which Broader Impacts statements are being addressed in a sample of proposals submitted to and awarded by EEC?*. To answer this research question required a narrow scope of what constitutes as Broader Impact. Thus, I used the Broader Impacts criterion as outlined in the 2016 *Proposal and Award Policies and Procedures Guide*—Chapter II: Proposal Preparation Instructions, this chapter suggests ways Broader Impact activities may be accomplished. To quantify Broader Impact statements, I created six categories, taken from the NSF Procedures Guide suggestive list and coded Project Summaries accordingly. The seventh category resulted from discussions with Program Officers. The categories were reviewed by my NSF mentors and approved as an appropriate starting point. The seven categories selected are highlighted in Figure 2. In Figure 2, category seven is off to the side and not branched together with the rest of the six categories because it was not listed in the 2016 *Proposal and Award Policies and Procedures Guide*. Therefore, I grouped together the activities suggested by NSF and placed category seven to the side indicating it was additional to what NSF was suggesting. Further discussion about the seven categories is in the subsection that follows.

I examined 82 Project Summaries to quantify if certain Broader Impact statements, shown in Figure 2, had been addressed. While I acknowledge Figure 2 provides a limited representation of what constitutes Broader Impacts, it serves as a starting point for trying to understand the multidimensional complexities behind Broader Impacts. A complete list of proposals, between FY 2013-215, were randomly selected from the full list, categorized into their respective programs i.e. Centers, Workforce Development, Broadening Participation in Engineering, and Engineering Education, and examined for this study.

The 82 Project Summaries selected for this study were the same Project Summaries given to the Committee of Visitors (COV) for review and evaluation of the Engineering Education and Centers Division. I did not evaluate the entire portfolio of proposals that were given to the COV, which included declined and returned without review proposals. I used the Project Summaries of awarded proposals assigned to the COV because they were randomly selected to include a distribution of awards representative of the EEC portfolio across all programs over a three-year period (FY2013-FY2015)[20].

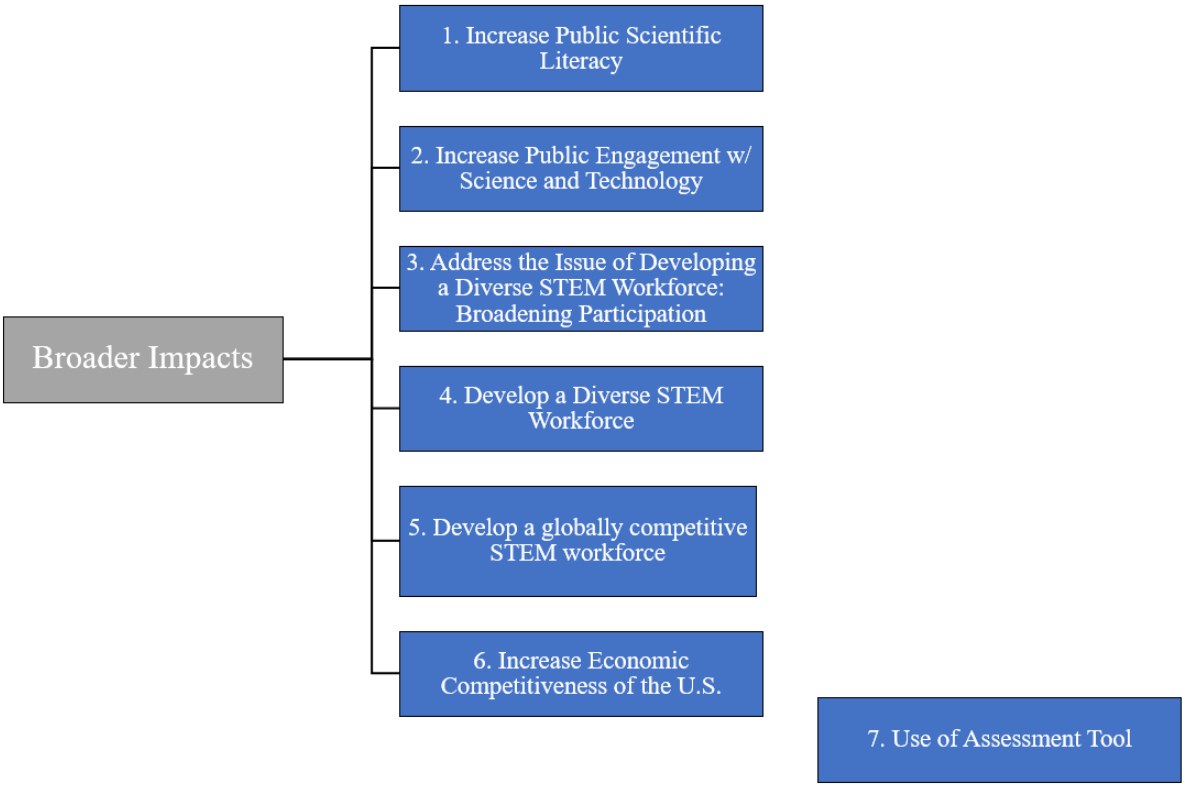

Figure 2 Broader Impact Categories

Before examining the 82 selected awards, a list of inclusion and exclusion conditions were outlined. In the section that follows, I summarize the Broader Impact categories that were measured in the quantitative phase of the project along with their respective inclusion and exclusion conditions. It should be noted that the inclusion criterion was intentionally stringent because the goal was to quantify the amount of awarded Project Summaries that talked about the seven criteria.

Constructing the Inclusion and Exclusion Criteria

The inclusion criterion, in part, was created from interviews of Engineering Education and Centers Program Officers—refer to the qualitative section. One of the common themes that arose through the interviews was the issue of only publishing research findings in journal articles and conference proceedings. In the interviews with Program Officers, it was emphasized that there was a need to reach both targeted and broad audiences, including practitioners as well as those who do not have access to prestigious journal publications. Thus, I decided to only count Project Summaries that stated alternative forms of dissemination for category 1—*Increase public scientific literacy*. Table 2 lists examples of alternative forms of dissemination, the list is not all encompassing but it provides examples of what was seen in the Project Summaries. While I acknowledge the value and utility of journal articles and conference proceedings, the aim was to quantify if alternative forms of dissemination were attempted in addition to the more standard methods.

For Broader Impact category *2. Increase public engagement with science and technology*, I applied the same exclusion criterion as category one. However, the inclusion criterion for category two is

different from category one in that I was looking for Project Summaries that spoke about creating programs, activities, events, etc. that would physically engage students with science and technology. Whereas category 1 was focused on material that was created not the actual event or activity itself. It should be noted that Project Summaries could potentially cover both category one and two. Since the NSF *Proposal and Award Policies and Procedures Guide* does not explicitly indicate how to increase scientific literacy or increase engagement in science and engineering, these inclusion and exclusion criteria were vital for establishing quantifiable measures.

The third category, *Address the issue of developing a diverse STEM workforce* and the fourth category, *Develop a diverse STEM workforce*, may appear to address similar issues, however, these two categories were intentionally kept separate. For category three, Project Summaries that were counted as meeting this criterion would have had a discussion on increasing participation and representation in STEM, thus to mitigate confusion between category three and four the name was changed to Broadening Participation. Category four was focused on developing a diverse STEM workforce *beyond* the scope of increasing participation and representation of underrepresented groups. If the Project Summary discussed the two categories in conjunction, they were counted for both criteria. The exclusion criterion for both categories was similar. If the Project Summary did not center the discussion around category three and/or four, then this led me to believe the focus was not increasing participation of underrepresented groups or developing a diverse STEM workforce. As described in the interview findings with the Program Officers, what is often seen in proposals submitted to the division are far reaching claims that somewhere in the distant future the outcome of the project will tangentially develop a diverse STEM workforce or increase participation.

In category *5. Develop a globally competitive STEM workforce* and *6. Increase economic competitiveness of the U.S.*, I was focused on whether Principal Investigators emphasize their study on global and economic competitiveness. For example, Project Summaries would have had to explicitly state how their project could or would develop a *globally competitive* STEM workforce, with emphasis on global competitiveness. Additionally, in categories *3-6* (see Table 2) I wanted to exclude Project Summaries that only address these areas at a surface level. That is, Project Summaries that have not provided additional detail in categories 3-6 were counted as not having addressed them. An additional detail that would have met the requirements in categories 3-6 includes—but is not limited to—literature on the importance and/or need for the topic area, ways in which the proposal will address the topic areas, and/or general discussion around the topic area.

Table 2 Broader Impact Inclusion and Exclusion Criteria

Broader Impact Categories	Inclusion Criterion	Exclusion Criterion
1. Increase public scientific literacy	Includes statements about broadly educating the general public, instructors, practitioners, parents, and students (outside of journal articles/conference proceedings) • Creating material for classroom or out-of-classroom settings • K-12 learning material • Open access sources • Going outside the scope of their own classroom and institution • Workshops • Printed articles in practitioner's magazines • Directly targeting key stakeholders This list is not all encompassing but provides a basis	Journal publications Conference proceedings
2. Increase public engagement with science and technology	Includes statements about in and out of school programming, activity, event, or learning module to engage K-12 and/or college students in STEM	Journal publications Conference proceedings
3. Broadening Participation	Includes statements about developing a diverse workforce through the participation of, but are not limited to: • full participation of women • persons with disabilities • underrepresented minorities • Veterans	Excludes one to two sentence statements as this does not focus the project increasing representation of underrepresented groups.
4. Develop a diverse STEM workforce	Includes statements about how the project would explicitly develop a diverse STEM workforce, outside the scope of participation of underrepresented students, includes but no limited to: • cognitive diversity • non-cognitive diversity • diversity in mindsets	Excludes one to two sentence statements as this does not focus the project on developing a diverse STEM workforce. This category *does not* refer to representation, but rather a discussion of workforce broadly.
5. Develop a globally competitive STEM workforce	Includes statements about how the project would explicitly develop a globally competitive STEM workforce.	Excludes one to two sentence statements as this does not focus the project on globally competitive STEM workforce.
6. Increase economic competitiveness of the U.S.	Includes statements about how the project would directly and explicitly lead to an increase in economic competitiveness.	Excludes one to two sentence statements as this does not focus the project on increasing economic competitiveness.
7. Use of Assessment Tool	Includes explicit mention of preexisting assessment tools or tools in the process of being developed. Also includes general mention of assessment regardless of using a specific 'tool'	

Coding

The coding of the Project Summaries was kept to a binary assessment, that is, Project Summaries received a 1 for each criterion addressed and a 0 for each criterion not addressed. Data was coded using Microsoft Excel. Assigning a 0 or 1 was determined by the inclusion and exclusion criteria outlined in Table 2. Since there is currently no quantifiable measure for Broader Impact statements submitted to the National Science Foundation, this approach was an exploratory step in that direction. The goal of this study was to 1) understand the complexity behind quantifying Broader Impacts and 2) create a dialogue with the community of engineering educators on how to support the National Science Foundation's challenge of quantifiably measuring Broader Impacts from awarded proposals.

Using the binary assessment, data were broken down into a comparison matrix in Table 3. Table 3 illustrates the percentage of Project Summaries that addressed each of the six categories. Using Microsoft Excel, I produced Table 3 by determining if categories were being addressed using the column filtering feature. I first assessed how many Project Summaries addressed the category individually by looking at columns that contained a 1, these individual evaluations are highlighted in gray. Then, for each category, I assessed how many of the sampled Project Summaries simultaneously addressed two categories. The percentages were taken from the total Project Summary sample size of 82. For example, the percentage of Project Summaries that addressed category 1. Increase public scientific literacy AND category 2. Increase public engagement w/ S&E were 32%.

Quantitative Findings

Findings from the Project Summaries examined revealed that 37 percent of Principal Investigators were proposing to increase scientific literacy (category 1) through alternative avenues, which included workshops, K-12 course material, open source material, practitioner magazines or intended to add new material to their curriculum on a subject not traditionally taught at their respective institutions, to name a few. While all examined Project Summaries stated that they would disseminate their work via journal publications and conference proceedings, the majority, 63 percent, only indicated disseminating their work via traditional methods (i.e. journal publications and conference proceedings). It should be noted that all Project Summaries had statements about publishing their work in journals or conference proceedings, but those that were counted as having met the Broader Impact category 1 also had alternative forms of dissemination. Findings also reveal that 32 percent of Project Summaries had combined Broader Impact outcomes, proposing to increase public scientific literacy and simultaneously increase public engagement with science and engineering (S&T).

Table 3 Comparison Matrix of Broader Impact Categories

	1.Increase public scientific literacy	2. Increase public engagement w/ S&E	3.Broaden Participation	4.Develop a diverse STEM workforce	5. Develop a globally competitive STEM workforce	6.Increase economic competitiveness of the U.S.	7. Use of Assessment tool(s)
1. Increase public scientific literacy	**37%**	–	–	–	–	–	–
2. Increase public engagement w/ S&E	32%	**67%**	–	–	–	–	–
3. Broadening Participation	20%	41%	**62%**	–	–	–	–
4. Develop a diverse STEM workforce	11%	17%	26%	**33%**	–	–	–
5. Develop a globally competitive STEM workforce	10%	18%	18%	15%	**29%**	–	–
6. Increase economic competitiveness of the U.S.	7%	11%	13%	10%	11%	**20%**	–
7. Use of Assessment tool(s)	15%	28%	23%	12%	7%	6%	**37%**

A large portion of Project Summaries examined, 67 percent, were categorized as meeting the inclusion criterion (category 2) for increasing public engagement with science and engineering, which included statements about executing programs, activities, events, and/or learning modules to engage K-12 and/or college students in STEM. Of the 67 percent of Project Summaries that proposed increasing public engagement with S&T, 41 percent *also* addressed the issue of developing a diverse STEM workforce i.e. increasing participation of underrepresented groups. I also looked to see how many Project Summaries included a discussion about increasing public scientific literacy (category 1) *and* increasing public engagement in S&E (category 2), results suggest that only 32 percent of the sampled proposals were addressing both criteria. This result indicated that Project Summaries that talked about increasing public scientific literacy were also more likely to have a component that is in line with increasing public engagement. Similarly, 41 percent of examined Project Summaries spoke about increasing public engagement w/ S&E in conjunction (category 2) with addressing the issue of developing a diverse STEM workforce (category 3).

From the sampled Project Summaries, 62 percent spoke about addressing broadening participation (category 3). This category includes discussion about how the respective proposed project will address the issue or support the efforts of including women, persons with disabilities,

underrepresented minorities and/or Veterans. Project Summaries that did not explicitly make a connection to increasing representation and the STEM workforce were still included in the count. While only 33% spoke about developing a diverse STEM workforce, which is outside of the scope of increasing participation. That is, more than half of the Project Summaries intend to address issues of diversity in STEM through means of representation and less than half used alternative language around diversity i.e. mindsets, attitudes, etc. The categories that were least addressed or possibly difficult to assess were *5. Develop a globally competitive STEM workforce* and *6. Increase economic competitiveness of the U.S.* Respectively, these two categories were addressed only 29 percent and 20 percent out of the Project Summaries examined.

The category that was addressed the most in the sample was increasing public engagement w/ S&E (category 2), at 67 percent. The goal for this category was to quantify how many Project Summaries discussed the issue of increasing participation/representation in STEM. The intention of category *4. Develop a diverse STEM workforce,* was to quantify the amount of Project Summaries that discussed creating a diverse workforce outside the scope of physical representation. In this category, a diverse workforce did not include increasing participation/representation in STEM, however, it does include other forms of creating a diverse workforce e.g. mindsets, creativity, new perspectives etc. Of the Project Summaries sampled, 31 percent had detailed discussion of how they would propose to develop a diverse STEM workforce. In the sampled Project Summaries, PIs were inclined to discuss developing a globally competitive STEM workforce slightly more frequently than increasing U.S. economic competitiveness, 29 percent versus 20 percent respectively. However, overall categories five and six were the least talked about and least well defined among the sample.

Lastly, the research group sought to quantify how many Project Summaries explicitly spoke about using a method of assessment or assessment tool for their Broader Impact statements. Findings revealed that 37 percent of the Project Summaries examined proposed the use of some form of the assessment tool. The categories that used assessment tools more often were 2. Increase public engagement w/ S&E at 28 percent and 3. Broadening Participation at 23 percent. I postulate that because these two categories tend to coincide with developing in-school or out-of-school activities, it's natural for PIs to use assessment tools to validate the effectiveness of the activity.

Discussion

Since the 1997 decision by the National Science Board to evaluate NSF proposals based on a two Merit Review criteria, Intellectual Merit and Broader Impact, the PI community has faced continuous challenges on addressing the Broader Impact criterion. Findings from the 2011 Task Force revealed that the Broader Impact criterion is not well understood among many in the community of applicants. This issue remains true today, as noted in the interviews with Program Officers. These interviews revealed trends on how Broader Impact was being addressed one being that Principal Investigators were substituting broadening participation for Broader Impacts. While broadening participation may encompass one aspect of Broader Impact there are more aspects of Broader Impact that can be leveraged. Quantitative findings, of the sample of Project Summaries, revealed 62 percent addressed the issue of increasing participation of underrepresented groups. Data also uncovered that Project Summaries tended to have more overlap with the category 3. Address the issue of developing a diverse STEM workforce i.e. broadening participation. This was evident in that 41 percent of Project Summaries jointly talked about increasing public engagement

with S&T and broadening participation and 26 percent jointly talked about broadening participation and developing a diverse STEM workforce beyond physical representation. That is, Program Officers observed that proposals were not providing evaluation metrics for their Broader Impact statements. Broader Impact statements were generally presented as 'someday in the distant future' rather than designed into the proposed study with a measurable metric.

Program Officers were cognizant of the difficulty of assessing benefits to society, the idea of having a one-to-one correspondence, that is, assigning a dollar amount that would translate to solving *X* societal problems was also unreasonable. Program Officers were quick to propose simple first step solutions towards addressing the Broader Impact criterion, however. It was proposed that PIs could target alternative forms of dissemination that have a greater potential to reach a wide group of audiences. Additionally, by having a Broader Impact research question built into the research design, PIs can assess its outcome through pre/post testing. These first step propositions do not encompass every possible method of assessing Broader Impacts, rather they were provided as stepping stones towards moving in that direction. Another major component was the Panel of Reviewers' understanding and knowledge of how to assess Broader Impacts. It was noticed that while efforts were made to increase knowledge of potential Broader Impact outcomes, Reviewers tended to treat the criterion as a pass or fail. One Program Officer observed a type of reproducing cycle, that is, the PIs who submit lofty and unmeasurable Broader Impact claims may also be the PIs who are serving in panel reviews. Thus, by providing educational resources on how to address Broader Impact criterion to the PI community, for the benefit of their own proposal submissions, could also benefit those PIs serving as panel reviewers. Based on my observation, I believe a training workshop assisting Principal Investigators on how to incorporate Broader Impact activities into their research proposals would be the logical next step. Within this training workshop, the PI community should be made aware of current issues around addressing Broader Impacts. That is, PIs should be made aware about lofty, 'pie in the sky' statements that lack measurable assessment tools or metrics. While NSF does provide guidance on both review criteria in outreach events, it would be helpful to have additional training from other organizations, such as National Alliance for Broader Impacts (NABI), and others that may be supporting Broader Impact efforts. Principal Investigators should be given support, either by their home institutions or departments, on how the Broader Impact criterion could be addressed. A straightforward way is to be made aware of organizations such as NABI that aims to "create a community of practice that fosters the development of sustainable and scalable institutional capacity and engagement in broader impacts activity"[21]. Engineering education researchers can capitalize on this existing resource to support their broader impact efforts.

Limitations & Future Work

This study was conducted by an NSF Summer Scholar, therefore the findings, conclusions, and observations expressed in the qualitative phase of the study are those of the author and do not necessarily reflect the views of the National Science Foundation. The quantitative portion of the study comes with many limitations, limitations in the form of assumptions and biases. As there is no formal method of evaluating Broader Impacts, I created inclusion and exclusion criteria for the six Broader Impact categories based on interviews with Program Officers and my own interpretation. This poses biases as researchers tend to pull from their research area of interest. To mitigate this issue, I could have created a panel of engineering education researchers with a wide spectrum of research interest to assist in creating the inclusion and exclusion criteria.

By using a coding method for inter-rater reliability amongst a diverse team of engineering educators would enhance the reliability and trustworthiness in the analysis. This option was not available to me due to limited resources and time, however, I acknowledge the value and utility of this approach. This work is also limited to Project Summaries and Program Officers within the EEC division, future work including to more divisions within the Engineering Directorate is needed. It should be noted that engineering education researchers tend to submit proposals to divisions outside the Engineering Directorate. Therefore, to obtain a holistic understanding of how engineering educators are addressing the Broader Impact criterion, data from outside the Engineering Directorate would need to be gathered. Additionally, perspectives from the PI community were not collected thus this study excludes insight from the community of grant writers. I present these limitations also as future work, understanding how to address the Broader Impact criterion in a measurable and scalable way will require an extensive understanding of the criterion.

Acknowledgement
This study was made possible through a summer internship experience at the National Science Foundation by the Hispanic Association of Colleges and Universities (HACU). Special thank you to my mentors Don Millard Deputy Division Director, Amelia Greer Science Analyst, and all the Program Directors in the Engineering Education and Centers Division. An additional thank you to those Program Officers who participated in the interviews.

References

1. *National Science Foundation Act of 1950, Public Law No. 112–166*. https://www.nsf.gov/about/history/legislation.pdf.
2. National Science Board. *NSB/NSF Seeks Input on Proposed Merit Review Criteria Revision and Principles (NSB-11-42)*. Washington, D.C.; 2011. https://www.nsf.gov/nsb/publications/2011/meritreviewcriteria.pdf.
3. Foundation NS. *Empowering the Nation through Discovery and Innovation NSF Strategic Plan for Fiscal Years (FY) 2011-2016 (NSF 11-047)*. Washington, D.C.; 2011.
4. National Science Board. *National Action Plan for Addressing the Critical Needs of the U.S. Science, Technology, and Mathematics Education System*.; 2007.
5. National Science Board National Science Foundation. *Task Force on Merit Review Discussion Report (NSB/MR-96-15)*. Washington, D.C. https://www.nsf.gov/nsb/documents/1996/nsbmr9615/nsbmr9615.htm?org=NSF.
6. Lok C. Science for the masses: the US national science foundation's insistence that every research project addresses' broader impacts' leaves many researchers baffled. Corie Lok takes a looks at the system. *Nature*. 2010;465(7297):416-419.
7. Foundation NS. *Important Notice to Presidents of Universities and Colleges and Heads of Other National Science Foundation Grantee Organizations*. Washington, D.C.; 1997. http://www.nsf.gov/pubs/1997/iin121/.
8. Foundation NS. *Proposal and Award Policies and Procedures Guide*. Washington, D.C.; 2016. https://www.nsf.gov/pubs/policydocs/pappguide/nsf16001/nsf16_1.pdf.
9. Holbrook JB. Assessing the science-society relation: The case of the US National Science Foundation's second merit review criterion. *Technol Soc*. 2005;27(4):437-451.
10. Colwell RR. *Important Notice 125 Implementation of New Grant Proposal Guide Requirements Related to the Broader Impacts Criterion*. Washington, D.C.; 2002.
11. Act AC. Public Law 110-69. *Washington, DC GPO Sect*. 2007;1105.
12. National Science Foundation. *The National Science Foundation Division of Education and Engineering Centers (EEC) 2013 Committee of Visitors (COV) Report*. Washington, D.C.; 2013. https://www.nsf.gov/od/oia/activities/cov/eng/2013/EEC 2013 COV Final Report_June 4-2014.pdf.
13. Nagy D. *Evaluating the Broader Impacts of Sponsored Research through the Lens of Engaged Scholarship.* ERIC; 2013.
14. Foundation NS. *Faculty Early Career Development Program (CAREER) Includes the Description of NSF Presidential Early Career Awards for Scientists and Engineers (PECASE)*. Washington, D.C.; 2015. http://www.nsf.gov/pubs/2015/nsf15555/nsf15555.pdf.
15. Hallinen JR. The many quiet tensions: Perceptions of the broader impacts criterion held by NSF career award holders at very high research institutions of higher education. 2014.
16. Roberts MR. Realizing Societal Benefit from Academic Research: Analysis of the National Science Foundation's Broader Impacts Criterion. *Soc Epistemology*. 2009;23(3-4):199-219.
17. Creswell JW, Clark VLP. Designing and conducting mixed methods research. 2007.

18. Teddlie C, Yu F. Mixed Methods Sampling: A Typology With Examples. *J Mix Methods Res*. 2007;1(1):77-100.
19. L AATP, August E. National science foundation act of 1950. 2013:80-89.
20. National Science Foundation. *Core Questions and Report Template for FY 2016 NSF Committee of Visitors (COV) Reviews*. Washington, D.C.; 2016. https://www.nsf.gov/od/oia/activities/cov/eng/2016/EEC_2016_CoV_Report_FINAL.pdf.
21. National Alliance for Broader Impacts. About NABI. https://broaderimpacts.net/about/. Accessed December 16, 2016.