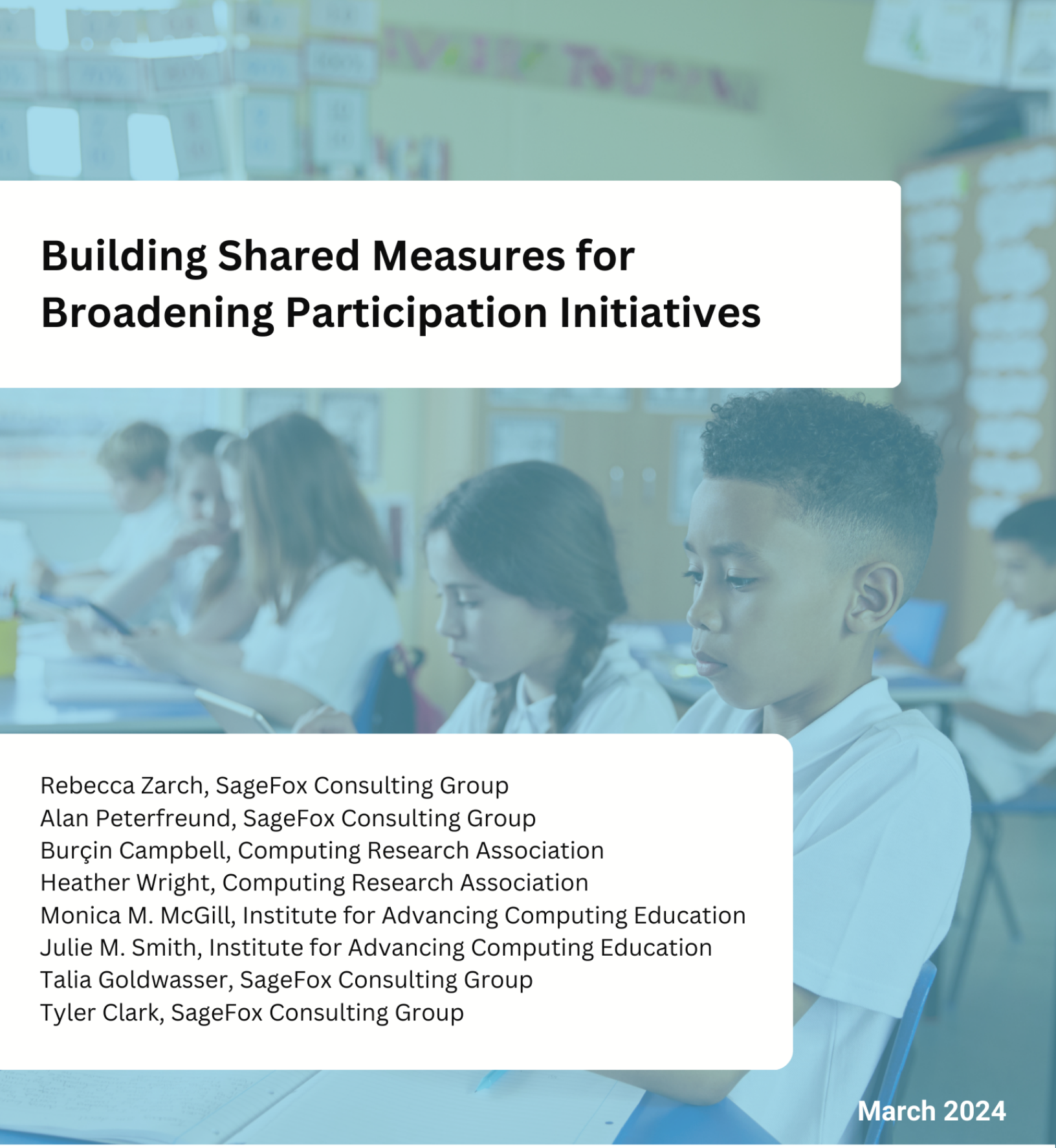
Building Shared Measures for Broadening Participation Initiatives



Rebecca Zarch, SageFox Consulting Group
Alan Peterfreund, SageFox Consulting Group
Burçin Campbell, Computing Research Association
Heather Wright, Computing Research Association
Monica M. McGill, Institute for Advancing Computing Education
Julie M. Smith, Institute for Advancing Computing Education
Talia Goldwasser, SageFox Consulting Group
Tyler Clark, SageFox Consulting Group

March 2024

This work was supported by National Science Foundation Award #2137842.

Contents

Recommended Citation: Zarch, R., Peterfreund, A., Campbell, B., Wright, H., McGill, M.M., Smith, J.M., Goldwasser, T., Clark, T. (2024). Building Shared Measures for Broadening Participation Initiatives. https://doi.org/10.62572/jZamU35

Executive Summary

The ability to aggregate and compare data across programs makes it possible to gain a deeper insight into the effectiveness of various interventions for improving equity in education, particularly for those interventions designed to create systemic change. This report describes the outcomes of participatory action research focused on an effort by the National Science Foundation (NSF) Broadening Participation in Computing (BPC) alliances to develop shared measures for data reporting.

Led by Computing Research Associates and SageFox Consulting and with support from the Institute for Advancing Computing Education, ten alliances participated in this research study. This project built on the shared measures developed by the NSF's INCLUDES Network Coordination Hub Shared Measures Initiative. The participants found these shared measures to be workable in their own context if some modifications were made. Alliance representatives then attended a two-day, in-person meeting to discuss and define the proposed shared measures in more depth. Building upon this workshop, the leadership team developed a template to be used across the different alliances to report measures that could then be aggregated. The leadership team then piloted the template, and seven alliances completed the template with their alliance's data. The leadership team aggregated the data to test the effectiveness of the process and of the measures.

Key Accomplishments

The major accomplishments of this project were:

- Engaging the community in a project that is critical for understanding their collectively aggregated impacts.
- Articulating a set of challenges related to developing methods for aggregating data across projects.
- Creating a template and instructions, with definitions, that fostered the ability to collect and aggregate data across alliances.
- Generating positive changes within different alliances – an unexpected

side effect of alliances reflecting on their internal processes for collecting and aggregating data.

Key Challenges

Developing shared measures across distinct broadening participation initiatives is inherently multidimensional and complex. This is due, in part, to the reach and depth of the alliances' partnerships, the variety of strategies and activities they engage in, and the types of resources and products they offer. Within the context of this project, we uncovered many such challenges, including:

- Defining key terms, such as defining the activities and outcomes of alliances; accurately capturing data related to diversity, equity, and inclusion; and defining partners and their level of involvement.
- Data collection, including identifying instrumentation, aggregating both qualitative and quantitative data, and handling labels that may shift over time.
- Data aggregation, including accounting for differences in year 1 versus year 5 of a project, avoiding over-counting of data when participants are benefiting and included in activities from more than one alliance, distinguishing between direct and indirect impacts, and addressing data privacy and misuse concerns.

A recurring theme that arose was how the data will be used and whether comparison measures might be put into place that would benefit some alliances (e.g., those that involve more participants in activities) more than others (e.g., those that focus on systemic change through policy initiatives and have few participants that are engaged directly with the alliance). Data privacy issues also arose, particularly for vulnerable and marginalized populations.

Key Lessons Learned

Lessons learned included that:

- Starting from the shared measures used by a similar group – as opposed to starting from scratch – aided the project in its efforts to proceed quickly.

- Considering data over time was critical, especially for alliances who do not work directly with students.
- Distinguishing between direct and indirect impact was critical for alliances that often struggle to demonstrate their value in a numeracy-based climate when, by design, they don't engage with students or educators directly.
- Distinguishing between outputs, outcomes, and impacts as central to metric development.

Summary

Despite the challenges, the project was able to develop a preliminary set of shared measures, which are likely to be useful in future data analysis.

Acknowledgements

We acknowledge and thank each of the members of the alliances that participated in this study. In particular, we recognize and thank the following individuals for their time and expertise that went beyond alliance participation: Christine Ong (UCLA CRESST), Tom McKlin (The Findings Group) and Wendy DuBow (National Center for Women & Information Technology).

We also thank Roohia Meer, Computing Research Association, for her assistance on this project.

1 Introduction

Many different groups are working to broaden participation in areas where representation is currently disproportionate. In some cases, several such groups may have the same goal (e.g., to increase the number of underrepresented students majoring in a STEM subject). One of the obstacles these groups can face is that, despite their shared goal, they use disparate methods and measurements to assess their collective outcomes, making it difficult to evaluate the broader results of their efforts. The adoption of ***shared measures*** can help gauge their collective efforts. However, adopting shared measures presents some formidable challenges; this report presents the results of one effort to establish shared measures – by organizations focused on broadening participation in computing – in order to share lessons learned with groups that may be considering the adoption of shared measures in other fields.

The Broadening Participation in Computing (BPC) Program of the National Science Foundation (NSF) Directorate of Computer and Information Science and Engineering (CISE) seeks to increase the representation of various groups currently underrepresented in computing. One of the goals of NSF for this effort is "to encourage thoughtful engagement of and meaningful action by the community on this long-standing issue." (https://www.nsf.gov/cise/bpc/)

The BPC Alliances are organizations designed to provide expertise to the computing community in support of broadening participation. Each alliance has a unique focus (e.g. demographic group, phase of the education system). Alliances do more than promote diversity and participation; they are increasingly focusing more on systems change and capacity building. To achieve their goals, each BPC alliance works with multiple collaborators. Alliances are intended to collectively impact across the PreK to workforce ecosystem and for the populations included under the NSF BPC umbrella. Partner organizations can also implement multiple interventions designed to support the goals of the alliance. Thus, systems of shared measures must consider the individual partners that comprise the alliance, and the collection of alliances themselves need to be woven together in a

way that tells a story about the collective impact of each alliance.

Creating evidence at the program level is necessary for continued program support and to understand the collective impact of the BPC alliances. However, it has been challenging to demonstrate the impact of all alliances collectively in part because each project or alliance has relied on their own framework of proximal and longitudinal measures deemed most appropriate for that project. These measures address direct participation in alliance offerings and possibly indirect impact on individuals; the partnerships established to do the work; and the capacity being built to do BPC systems change work.

These measures are important for the growth and effectiveness of individual alliances. However, they create challenges when trying to demonstrate the collective impacts across the entire BPC program. There is no common framework or mechanism to collect, collate or aggregate these data into a shared story across the alliances. This disconnect formed the basis for the first evaluation of the BPC program. Since then, there has been continued collaboration among the alliances and, most recently, the work of BPCNet has championed a platform that provides a forum for continued dialogue, a wide range of resources, and tools (https://bpcnet.org/) that allow access and tailoring of national data sets.

Figure 1.1: An iAAMCS activity working with students.

To address this problem, the Shared Measures project aimed to produce common definitions and a framework to enable comparable measurement across the NSF BPC community of practice (including BPC alliances and demonstration projects) and potentially across the NSF's broader set of over a dozen broadening participation programs. These shared frameworks and definitions have been designed to facilitate learning across the broader BPC community about model practices for broadening participation. The primary goal of this project was to develop a shared understanding of capacity building and associated activities/outcomes that will then lead us to identifying a framework for shared measures.

All alliances for broadening computer science education pathways agreed to

participate in this project. Alliances included in this project were:

- AccessComputing (Alliance for Access to Computing Careers)
- AIICE (Alliance for Identity-Inclusive Computing Education)[1]
- CAHSI (Computing Alliance of Hispanic-Serving Institutions)
- CRA-WP (CRA's Committee on Widening Participation in Computing Research)
- ECEP (Expanding Computing Education Pathways)
- iAAMCS (Institute for African-American Mentoring in Computing Sciences)
- LEAP (Diversifying Leadership in the Professoriate)
- NCWIT (National Center for Women & Information Technology)
- REAL-CS (Researching Equity and Antiracist Learning in CS)
- STARS Computing Corps

This project had four phases:

- Phase 1: Framing (October 2021-March 2022). This phase set the foundation for collaborative work by developing an understanding of the history upon which this project was building; individual and collective goals of the alliances, current measurement activities and seeking areas of commonality. This group also explored the current opportunities and limitations for reporting the true value of their work given NSF reporting structures. During this phase the group identified ***capacity building*** as the priority for exploration through this project. The group also reviewed and agreed to a recommended set of ***participation measures*** for gathering data about program participants.
- Phase 2: Collaborative Work (April 2022-December 2022). At a two-day, face-to-face meeting the group did a deep dive into what capacity building looks like at the alliance level and they generated a set of constructs across the community to test through a shared reporting template.

[1] AIICE is not a BPC-A; however, they agreed to participate since they are a BPC focused Alliance (INCLUDES).

- Phase 3: NSF BPC Alliance Implementation Efforts (December 2022-June 2023). Alliances tested and revised the reporting template. During this time the alliances met regularly to give greater definitional clarity to the constructs.
- Phase 4: Aggregating Lessons Learned, Sharing Findings, and Defining a Sustainability Plan (March 2023-June 2024). Aggregating results confirmed that the capacity building constructs could be gathered and aggregated across alliances with some confidence. During this time the participation measures were also piloted with the larger BPC community as part of the CISE-EWF PI meeting. The group met with the NSF and produced technical reports and recommendations to be hosted on the BPCnet.org website.

Multiple priorities for the NSF and the alliances must be simultaneously considered when attempting to generate shared data. Some of these include:

- The NSF must provide evidence to the U.S. Office of Management and Budget that BPC alliances are being impactful, since millions of dollars have been committed to change. Evidence that shows that these funds are creating change is critical for continued support from the NSF and Congress.
- Honoring history and context is necessary for each alliance and its particular focus areas and groups. When presenting this data quantitatively, history and context must be preserved.
- Each individual alliance has a need to tell their own compelling stories of capacity building that are well suited for their alliance.
- Projects and leaders must collect and analyze data over time to measure the impact of their programs.
- Future projects must learn how to organize their evaluation plan and build on the lessons learned.

In this report, we provide a background of the BPC alliances and shared measures efforts, an overview of our project, and a summary of the existing challenges to data collection. We then present the toolkit created as part of this project and the lessons learned.

2 Background

2.1 Existing Efforts in Shared Measures

The broader BPC community has engaged (and continues to engage) in a number of efforts to achieve alignment in measurement systems. A decade ago, the BPC community worked collaboratively to develop a set of common metrics (Mcklin, 2012). That effort, which included members of this project team and current alliance partners, laid the groundwork for the subsequent program level evaluation (Education Development Center, 2017). At the NSF BPC alliance level, individual alliances have worked (and continue to work) on developing metrics that best reflect the focus of their communities. These efforts include both metrics among those within a single organization and those across multiple organizations. For instance, the ECEP alliance has been working with the state K-12 education systems to identify common metrics and definitions in order to make comparable measurement possible (Dunton et al., 2022; Zarch et al., 2019). AccessComputing has focused on how best to represent participation of students with disabilities (Blaser & Ladner, 2020). CAHSI, with a focus on Latinx students, has addressed metrics associated with their retention and progression (Zatz et al., 2017).

Figure 2.1: The CAHSI Allyship focuses on the success of female and all students in computing by pairing students in the fundamental courses with peers at their institution who have completed the data structures course.

Beginning in September 2012, the Education Development Center (EDC), Westat, and Kansas State University conducted a comprehensive evaluation of the

NSF BPC-A program (Education Development Center, 2017). This evaluation effort also focused on a set of metrics named the ***Common Core*** (no relation to the Common Core standards) to request reporting from the NSF BPC alliances. In its 2017 evaluation report on the NSF BPC-A program, the evaluation team made the following recommendation:

> *In addition to requiring the collection of participant data, we recommend that NSF actively support a culture of using project and program data to examine the efficacy and influence of individual alliances. This data-driven culture can be fostered through continuing to support the convening and collaboration of alliance evaluators. Based on our own observations, we recommend that NSF continue to support efforts to maintain this infrastructure and convene the evaluators on at least an annual basis both within the alliance program and across other programs with common goals, such as NSF INCLUDES (Education Development Center, 2017).*

Figure 2.2: Students participating in AccessComputing activities.

Given this context, the specific challenge addressed by this project has been the advancement of shared measurements and definitions within the context of the NSF CISE EWF BPC Program. The project was aimed at analyzing the current state of the organizations involved in broadening participation and identifying their needs. By design, the process was based on the existing efforts of each of the NSF BPC alliances and complementary work underway within the NSF INCLUDES community. The outcome is an alignment among the alliances in a shared framework of measures and definitions.

2.2 Shared Measures Project Process

The shared measures project was based on four phases of work. Our work was accelerated by NSF INCLUDES community receiving funding to investigate and publish their shared measures. We used these measures, which were developed in part by the same team that led the BPC-A evaluation, as the basis for early discussions, particularly around participation measures.

2.2.1 Phase 1: Framing

The first phase of the project set the foundation for the collaborative work and was designed to be flexible in response to opportunities with the alliances. There were four core meetings (in October 2021, November 2021, January 2022, and February 2022) designed to orient the team, provide a summary of the current landscape of data collection across the projects, review educational equity and national data resources, and define a working plan for the NSF BPC Shared Measures project.

Key to this initial phase was creating a shared value proposition associated with the project and creating a trusting and open environment. The values associated with this work, as discussed in the previous section, were relatively easy to establish due to prior working relationships and shared commitment to the overarching goals. Building trust was more challenging given concerns about the potential public representation of the resultant products and their reflection on the participating projects (many of which are discussed in more detail in this report). Orienting the PIs of the participating NSF BPC alliances was considered to be an important part of this process. To support this work, the core team was engaged in reviewing the current goals and objectives from each alliance, understanding the NSF INCLUDES Shared Measures project as a potential model for BPC-A, and preparing for the PI Orientation Meeting (March 2022). Emerging from this work and from engaging the BPC-A representatives came a clear call for focusing on how to describe the capacity building efforts of the alliances to meet the collective impact goals for BPC.

2.2.2 Phase 2: Collaborative Work

The NSF BPC Shared Measures team organized an in-person meeting in August 2022 in Denver, Colorado. This collaborative working meeting brought together each alliance's Metrics Coach (MC) and at least one other representative from each alliance (Usually the Principal Investigator and/or Evaluator or Project Co-ordinator) to participate in a deep dive discussion on defining the shared frameworks and key metrics. Prior to the gathering, each MC ensured that within their alliance (a) the direction of the project was meaningful to the alliance goals and objectives, and (b) the selected attendees were clear on the expectations for participation in the face-to-face meeting, including contributions during the meeting and upcoming Phases 3 and 4. MCs also engaged the relevant organizations and partners of the alliance to ensure many voices of the alliances were represented at the meeting. The meeting was designed to meet the following goals:

- Maintain trust built among key partners and the project team
- Develop a shared language and definitions which will be important for accessing reporting data consistently
- Agree upon framework(s) for understanding participant demographics in context
- Determine the phases to parse CS interventions by academic phase (i.e. K-8, 9-12, Community College, Undergraduate and Graduate programs)

The alliances left the in-person collaborative meeting prepared to engage in their regular work with a renewed focus on how they collect and report on impact with a shared lens of capacity building to broaden participation in computing with underrepresented group

Four of the alliances reported leaving this meeting and modifying their annual report style (for October 2023 submission) to include the new conceptions of capacity building.

2.2.3 Phase 3: NSF BPC Alliance Implementation efforts

During Phase 3 the leadership team translated the work from the face-to-face meeting into a set of templates and definitional guides to explore capacity build-

ing. Over two months the Metrics Coaches met to refine the template based on their alliance's feedback and perspectives until an agreed upon form was created. The group then worked independently to populate this form with their alliance teams. Regular meetings surfaced areas for greater definitional clarity. Draft aggregation of data surfaced challenges with reporting that led to another, final, substantive change to the template which was then populated with "best and final data" in March 2023.

2.2.4 Phase 4: Aggregating Lessons Learned, Dissemination, & Sustainability

In phase 4 the leadership team aggregated the data and an extended meeting allowed alliance participants to collaboratively interpret the collective impact of the program. Building on the lessons learned, data reporting guidance was developed for use within the NSF BPC community as an exemplar for reporting alliance and/or project impact on broadening participation in computing. Along with this tool will be a guidebook providing recommended data practices for the BPC community when collecting and reporting data related to the shared measures developed through this project. Finally, alliances will work to use shared language and frameworks in their annual reports, allowing for the potential aggregation of impact across alliances.

Under a different award, this project informed data collected across the CISE BPC community as part of the CISE PI meeting in 2023.

2.2.5 Leadership Team and Participating Alliances

The leadership team for this Shared Measures project consisted of Computing Research Associates (CRA), which hosts the bpcnet.org forum, and SageFox Consulting Group. Details of this project were shared with participating alliances ahead of time; many of the key participants had prior collaborative relationships. This was to help ensure that there was a co-ownership of this Shared Measures work among all participating alliances and representatives from each alliance and intentionally done to build trust and cohesion that went beyond co-design.

The project team has consistently told the alliances that a) the goal of this project is to develop measures that work across alliances to understand the col-

lective impact and that b) the leadership team is not going to serve as a central data hub; rather the way data is collected and reported will ultimately be determined by each alliance, through annual reporting. The leadership team consistently emphasized that the purpose of data collection was ***not*** for reporting or judging but rather for a research project focused on determining the best ways of collecting data as an aggregate unit.

2.2.6 Evaluation

A CRA evaluator who was external to the project team conducted an outcome evaluation of the project. Given the scope and duration of the project, an outcome evaluation was the most appropriate evaluation method. The primary outcomes that were used to evaluate the success of this project included (a) the engagement and satisfaction of the alliance representatives, (b) the production of deliverables with shared measures, and (c) alliance use of those metrics. The team leads and alliance representatives specified measurements for each outcome at the start of the project following the first orientation meeting.

The evaluator attended the virtual and in-person meetings, surveyed the representatives, and monitored the creation, use, and dissemination of the shared measures. The evaluator provided an interim report near the end of Year 1 focusing on representative participation in and satisfaction with the process. In Year 2, the evaluator investigated the use of metrics, discussions on lessons learned throughout the measurement implementation process, creation and use of resources, and continued alliance interaction. The evaluator generated a final report based on these data near the end of the grant period. Both reports were shared with the participating alliances and included in the annual reports to NSF.

3 Challenges to Shared Measures Approaches

Efforts by the community and scholars to achieve a national understanding of broadening participation in computing efforts have been challenged by issues such as uncoordinated and decentralized data systems, lack of clear and common definitions, sampling problems, and lack of mandates, requirements, and funding. For example, some efforts have used self-reported data, including trying to understand where CS courses are offered (Code.org et al., 2022), but these efforts were challenged by sampling issues and definitions of ***computer science*** (Google & Gallup, 2020). In ECEP's work with 23 states and territories participating in the alliance, definitional issues came to the forefront, including defining what counts as a CS course or what is considered a high school (Fletcher & Warner, 2021).

3.1 Inherent Challenges of Measuring Combined Impact

Collective measurement is difficult to achieve given the variance built into the very fabric of each alliance's activities. There are multiple challenges in such a project, including the problems with measuring and comparing projects that collect different forms of data, the context for that data, and how different tools are used to measure the same data. The loss of context can be very problematic, for example, and any effort designed to build a process for integrating evidence from various projects will face common challenges.

We recognize that the challenges faced in developing shared measurements across broadening participation efforts are multidimensional and complex. As part of their evaluation of the NSF BPC alliances, Goodyear et al. identified some of the sources of the complexity of documenting and assessing outcomes across

Figure 3.1: CAHSI students, faculty, and staff attended the 2023 Great Minds in STEM (GMiS) conference in Pasadena, CA.

NSF BPC alliances. These sources included the reach and depth of the alliances' partnerships, the variety of strategies and activities they engage in, and the types of resources and products they offer (Goodyear et al., 2017). This variation of strategies and activities across alliances is a direct result of the complex issue that the broader BPC community is facing and is therefore unavoidable.

3.2 Challenges Defined by Previous Research

3.2.1 General Challenges

Synthesizing and aggregating sets of disparate data has critical challenges and limitations. Looking at studies collectively is compelling, as it tells a different story of what the collective evidence shows (Clarke et al., 2014). Further, the creation of large datasets has been a goal of multiple organizations; however, there is also a need to protect the integrity of the data when it is merged.

Kadadi et al. (2014) recognize seven key challenges of big data collection and integration, including:

- **Scope of data.** Categories for race/ethnicity and gender continue to shift. How will this changing scope impact interpretation of the data?

- **Data inconsistency.** How will inconsistent data be integrated and interpreted?
- **Query optimization.** How can the resultant data be integrated into a database so that the resultant queries can be operationalized meaningfully and remain respectful of the various contexts from which the original data was derived?
- **Inadequate resources.** How will adequate resources be provided to ensure ongoing training for collecting data?
- **Scalability.** How will the data systems be developed in a way that ensures scalability?
- **Implementing support systems.** How will support systems be established and maintained both through data collection and data aggregation? This may include the creation and support of tools that enable data collection as well as querying of data in ways that ensure data integrity.
- **Extract Load Transform (ELT) process in big data.** How will the data collection be initiated, be transformed in a manner that adheres to the data system paradigm, and maintain data integrity?

3.2.2 The Role of Derived Data

Searls (2005) explores the usage of "primary data, which is stored in operational or 'working' databases, to derived data, which is refined and presented at a higher level, where it is aggregated, visualized, statistically characterized, interpreted and used to drive decision-making" (Searls, 2005, p. 45). Searls (2005) states that "the integration of derived data...bears more on issues of 'organizational memory,' institutional communication and the proper juxtaposition of related information than on data analysis proper" (Searls, 2005, p. 45). The Shared Measures project sought to engage in the process of collecting derived data across multiple projects rather than connecting or using each alliance's working datasets. We maintain that each alliance is a study of a set of interventions to make collective change, and similar scrutiny should be given to aggregated data that is reported for an alliance.

3.2.3 Data Cleaning and Bias Mitigation

Certain aspects of data aggregation and analysis require care about assumptions about how the data will be used–with a clear definition of its use required ***a priori***. With this definition, careful planning can take into account methods of data cleaning, including how to handle data that is missing or that is not needed (and therefore may not be collected in the first place) (Osborne, 2012). Clifton et al. (2004) also note the need for the identification of duplicated data and how this data will be treated. Related, the data aggregation and analysis process also considers whether statistical analysis will be planned at some point in the future and how to build the resultant, aggregated data set in such a way that minimizes the potential for errors. The process also raises issues related to how basic assumptions about the data can bias the results. Finally, even in the case of non-statistical analysis that relies wholly on descriptive data, biased results can occur, and others have developed methods of weighting data so that more reliable data can have a greater impact on the results. The need for careful data analysis becomes a greater issue when the data includes experimental data such as pre- and post-tests that provide evidence on the outcomes of the alliance's activities.

3.2.4 Data Privacy and Security

Data privacy and security is important for protecting the identity of individuals as well as potentially protecting the identity of the alliances themselves. Clifton et al. (2004) call for creating a privacy framework for integration that is both flexible and clear to the end users, so they have confidence in it (Clifton et al., 2004). Further, even with derived data, there must be a schema to "establish semantic correspondences between schemas" (Clifton et al., 2004, p. 22) in such a way that doesn't expose the source data and schemas. For example, if an alliance has a unique project that no other alliance has, the single set of data from that project can be identified, potentially exposing the partner organization and the individuals involved.

3.2.5 Challenges in Scalability and Human Involvement

Similarly, Stonebraker, Ilyas, et al. (2018) note several defined challenges related to data collection and aggregation, including scalability, human involvement, and data cleaning. They note that scalable data integration is fundamentally challenging and requires schemas (and definitions) that map consistently in a way that preserves the original data. Similar to Osborne, they note that data cleaning processes must be built into the data workflow. Human domain experts are needed to ensure that the domains in which the data is collected consistently and accurately support the collection of heterogeneous data. Lastly, they note that "most data integration projects at scale have a political component, which must be dealt with." Similarly, political components might be shaped by various needs and goals of the alliances and/or the NSF which must be considered if a broader system of collecting and aggregating data across multiple projects is designed and implemented.

3.2.6 Methodological Triangulation

Finally, methodological triangulation is a research design methodology that focuses on the "confirmation of findings, more comprehensive data, increased validity, and enhanced understanding of studied phenomenon" (Bekhet & Zauszniewski, 2012, p. 40) through the synthesis of data produced across or within various studies. Arguably, the concept of merging data from various alliances and their projects is a form of data synthesis that can provide, to some degree, a level of understanding of the impact of broadening participation efforts. Denzin (2017) investigated using multiple methods to study a specific phenomenon with both qualitative or quantitative data, and the focus is on decreasing "deficiencies and biases that stem from any single method," (Mitchell, 1986, p. 19) creating "the potential for counterbalancing the flaws or the weaknesses of one method with the strengths of another" (Mitchell, 1986, p. 21).

Methodological triangulation also requires an understanding of how data is collected. Care must be taken when comparing data from disparate projects and sources, and even descriptive data must be contextualized with the projects and sources from which it was derived.

3.3 NSF BPC Shared Measures Group Challenges and Caveats

In addition to the general challenges outlined above, there are also some challenges specific to this project. Prior to piloting a shared measures collection process, the NSF BPC Shared Measures group raised a number of concerns. In this section, we explore these challenges in the context of our goal of creating a framework for collecting data that mitigates as many of them as possible in order to strengthen the process and improve the integrity of the resulting aggregated data.

3.3.1 Process and Project Context Related Challenges

Alliance members mentioned several challenges regarding data collection:

- **Defining alliances.** Defining what counts as an alliance is important because alliances and major projects that operate in similar ways to an alliance can collect and reflect the NSF's goals in ways that go beyond the funding division or program. Limiting this definition to only certain labeled projects may mean that the evidence of impact of the NSF-funded projects that have a goal of broadening participation may not be reflected in aggregated reports. It is critical to understand the distinction between a ***project*** and an ***alliance*** in terms of breadth of funding over time, the importance of relationship building, and scaffolding activities and capacity building over time.
- **Motivating partner organizations.** This is a process-related issue that must be addressed to ensure the data is collected in a timely manner and that alliances and partners can spend the time to ensure the accuracy of the data.
- **Capturing diversity, equity, and inclusion.** Are alliances operationally thinking about DEI in similar ways? If not, how can these differences be accounted for and mitigated?
- **Focusing on systemic change and capacity building.** This is important since NSF alliances have evolved over time to be more focused on systemic change. How do we ensure that the shared measures and resultant aggre-

gated dataset focus on systemic change and capacity building and not just the outputs?

- **Defining a common time frame.** How can a common time frame (e.g., academic year, calendar year, semesters, terms, etc.) for data collection be defined?
- **Defining partners and their level of involvement.** What is the threshold of involvement for alliance partners? Who qualifies for having their data collected?
- **Crafting universally-applicable definitions.** How are direct and indirect participants and outcomes defined? Are these definitions applicable across projects? What about the differentiation between teachers who have direct daily contact with students and others who may not have direct contact with students? How do we account for a changing definition of "direct" for participants across the intervention (e.g., participants that were indirect in one year, then direct in another)?
- **Considering power imbalances.** Who holds and does not hold power within the systems that the alliances are trying to influence? How are these power imbalances reflected in the data?
- **Separating individuals from organizations.** How do we measure the consequences of individuals participating in alliance-organized activities as distinct from their role as representatives of organizations? Or, if we decide to count an individual in both capacities, how do we handle double-counting?
- **Collecting data about resources and support.** Many alliances produce resources and provide support to other programs and individuals. If an alliance produces a book, for example, what meaningful data can be captured about it?
- **Preserving data context.** How can we preserve the historical and situational context of the data?
- **Capturing the impact of capacity building efforts.** Capacity building efforts are often too early in the systemic change process to have visible impacts. For example, policy changes may only impact one or two regions, but that impact can be a precursor to further impacts in other regions. This chal-

lenge also includes advocacy efforts, especially in the early stages where their impact is unclear.

- **Using data to rate projects.** Questions arose from the alliance members about how data might be used and whether the alliances might be assessed, rated, or compared across different projects. For example, if one alliance is reaching more learners than others, comparing the alliance against another alliance with vastly differing goals and outcomes could prove detrimental to the overall goals of the funding organizations.
- **Accounting for alliance members.** There is a challenge in accounting for people fulfilling various alliance team member roles (e.g., students, partners) versus people who are engaged in activities of the alliance as participants. This includes "train the trainer" programs, which may not collect teacher or student participant data.
- **Assessing practical costs of data collection.** How do these data collection processes impact an alliance's and their partners' financial resources and time? How does this impact their project work?
- **Honoring the impact of multiple organizations.** How can NSF account for an alliance's impact if the alliance and its partners receive funding from outside organizations? For example, if NSF only funded 10% of the alliance's budget, how is this allocation reflected so that NSF's impacts are accurately shared with other funders?
- **Exploring exposure to intervention.** How can we collect and integrate the aspects of various interventions like dosage and engagement? How do we account for the differing depth of engagement among different types of participants?
- **Capturing data about project derivatives.** Alliances can produce spin-off projects that may, for example, be sustained by funding from another organization. Do we want or need to collect data from projects that are derivatives of an alliance?

3.3.2 Data Collection Challenges

Additional data collection challenges are presented below.

- **Identifying existing instruments.** Are there existing instruments that can be used for data collection so that we do not have to recreate them?
- **Collecting data for projects that do not collect participant data.** Many alliances never collect participant data, since the alliance itself may only collect data at the capacity level (e.g., number of policies adopted). Alternatively, alliances may collect data about different participants in different ways (e.g., primary means or secondary means).
- **Collecting average numbers for extrapolation.** Data collection is costly, time consuming, and may be prohibitive based on the population being reached, including those who are most marginalized. Collecting the average number of people reached could be extrapolated from other projects.
- **Capturing quantitative and qualitative data.** How can both quantitative and qualitative data be captured (and synthesized) in a manner that respects various research processes? Aggregating impact across alliances when the measures are qualitative is very challenging.
- **Handling labels that shift over time.** It has historically been the case that labels and their meanings shift over time, particularly as they apply to individuals. How can labels (e.g., race/ethnicity, gender identity) be constructed in a way that accommodates these shifting labels?

3.3.3 Data Aggregation Challenges

Aggregating data across alliances has its own challenges. These are presented below.

- **Accounting for differences in Year 1 versus Year 5 data.** Start up years in alliances may have a considerably different level of participation and engagement, and their data may reflect this. When aggregated across various projects, how might this difference be reflected?
- **Avoiding over-counting of data.** Multiple partners are involved in multiple alliances. How can the identification and elimination of double, triple, or quadruple counting best occur?
- **Handling double-counting across years.** Many interventions require multi-year involvement or have multiple years of impact (for example, a teacher

may use a curriculum over several years). At what point(s) is their involvement and impact collected?

- **Addressing cumulative versus single entry data.** Reporting may occur in a project year but be cumulative or dependent on prior efforts.
- **Aggregating direct versus indirect impacts.** Even with the identification of direct versus indirect impacts, in what ways can direct and indirect data be compared?
- **Managing how measures are mapped to each other.** Some data may be self-reported versus assessed. Data may be observational or collected via interviews. It may also be institutional data (e.g., data collected by an institutions) or may be output-oriented (e.g., citations and publications).
- **Handling missing participant data.** How might missing participant data be handled? How might participants and partners who do not complete the program be handled?
- **Accounting for longitudinal work.** What are the specific issues related to longitudinal data collected by alliances, and how are those issues reflected in the data aggregation process?
- **Addressing data privacy and misuse concerns.** Data protection is a serious concern. How can measures be put into place that protect the identity of participants and partner organizations? Will the alliances receive some level of assurance that their data will be protected from potential (mis)use? How might the data make its way into machine learning or other artificial intelligence datasets and applications?

3.4 Mitigation Measures

Through the work of the BPC Shared Measures group, many of these issues (but not all) were considered when constructing a method for collecting data. For example, the group reflected on categories for race through a broad and intense discussion (including of the national census that collects similar data), and at the end this included an approach amenable for all alliances. This collective approach ensured that decisions were made through consensus. Likewise, issues

with maintaining the same definitions over time due to shifting definitions were also considered and addressed in a manner that provided for some mitigation of these issues.

Figure 3.2 illustrates a consensus that the Shared Measures group reached on how to disaggregate data and report it. There are three types of major actors that can be impacted and whose data should be collected: individuals, institutions and systems. There are three types of individuals: students, educators, and advocates, and these categories are further subdivided.

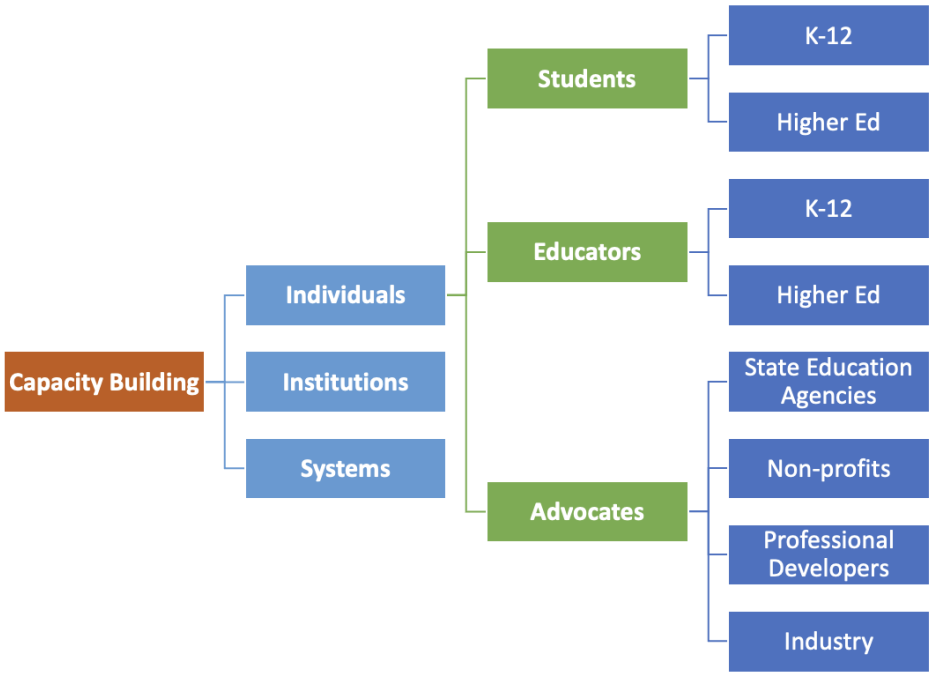


Figure 3.2: Individuals, Institutions, and Systems.

The language around broadening participation in computing, diversity, and systemic change must be developed, potentially as a framework. These can all vary across projects, and shared language is needed (including a broad definition of terms) to ensure data aligns to the framework. As part of this process, the Shared Measures team created a summary of the types of data categories, then shared that summary with the alliances so that they could provide feedback on whether the measures would work in their individual alliances and projects.

3.4.1 Measures Taken to Address Challenges

More specifically, we identified the following mitigation measures for future exploration and implementation of data collection of the alliances:

- Define the data collection and data reporting methods for various types of projects.
- For each definition, provide clear and concrete examples.
- Allow for flexibility between alliances, as long as there is consistency within an alliance.
- Use institutional data as much as possible.
- Develop a process based on best practices for cleaning data to mitigate data bias (including how to handle missing data).
- Ensure context is preserved, such as student disability status or locale.
- Ensure that inclusive terms are created and shared, such as for gender identity measures.
- Rather than using the term "other," ensure that options are comprehensive. This includes different options for grade bands, race and ethnicity, gender, roles, and more.

3.4.2 Acknowledging Ongoing Challenges

We recognize that many of the challenges raised have yet to be addressed. While some may be addressed through this Shared Measures project, we also recognize that this project is only a step in exploring the entire set of challenges and strategies for mitigation to provide clean data that leads to authentic storytelling about the collective impact of alliances' efforts.

4 The Toolkit

The toolkit developed and used by the alliances for this project consisted of three parts:

- A template which allowed alliances to describe their capacity-building efforts, specifically allowing for:
  - Alliances that may include the education-workforce CS ecosystem (e.g., teacher professional development)
  - Strategies to address education, such as providing professional development
  - Activities focused on the education system; goal of BPC is increasing degrees in CS
- A set of participation measures that leveraged the work of the NSF INCLUDES National Network Shared Measures group. These participation measures cover gender, race, products, and partners. The participation and partnership data helps complement this capacity building information to better understand specific gaps, particularly in terms of equity.
- Working definitions that guide both the participation measures and the capacity-building framework.

The toolkit was used with an eye toward discussion and refinement. The toolkit was co-developed and then revised twice with the alliances before data was collected. Its structure was intended to help the community understand the components of capacity building that could be articulated and reported for cross-alliance purposes, but it was not intended to replace the local-level evaluations so critical to each alliance for formative purposes and in-depth reporting.

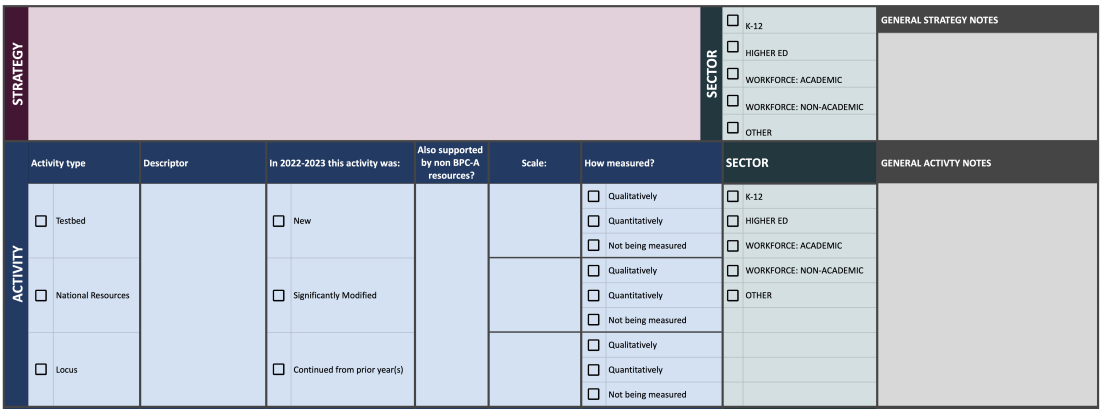


Figure 4.1: Template used for collecting strategy and activity data.

4.1 Capacity-Building Framework

The toolkit consisted of a basic form to collect data consistently across alliances. Figure 4.1 shows the two main components of the template, a strategy block and an activity block. There were five activity blocks initially included on each worksheet, and the alliances were asked to complete one worksheet for each of their strategies. Alliances could add or remove activity blocks to accommodate their activities for each strategy.

We expected that each alliance would have between 1 and 6 strategy areas for the purpose of this exercise. Alliances were given instructions, definitions, and coaching for completing the template. The alliances were asked to use their best judgment to describe how their alliance met each strategy and/or activity. When considering the data added to the templates, we collectively decided through discussion that consistency within an alliance matters more than precision across alliances. There were two key constructs alliances had to consider when using the reporting template:

- **Strategy.** Strategies are the long-term approaches to meeting the project goals. Strategies are supported by a set of activities. Early in this work, each team shared their overarching goals and measurable objectives. Strategies are the approaches used to support the goals. Each strategy can be supported by multiple activities.
- **Activity.** Individual activities are the components of a strategy. For example,

a workshop or a conference is usually part of a broader strategy of bringing together a community for a specific purpose. NSF annual reports have been excellent at gathering an alliance's activities; these are often more easily measured because they have participants who can be counted more easily than a strategy could be counted.

For each activity, the community agreed it would be useful to know the following:

- **Activity Type.** The BPC-A solicitation calls for alliances to engage in activities that are testbeds, national resources, and/or a locus, which are defined as follows:
  - **Testbed.** Activities that develop, test, and deploy interventions aimed at supporting students and faculty. These efforts should also focus on sustaining institutional transformation and promoting inclusive practices at the departmental and organizational levels. Examples include new/emerging professional development, new resource development, and conducting research.
  - **National Resource.** Activities that generate, vet, collect, curate, and disseminate best and promising practices for addressing underrepresentation to inform, educate, and connect the broader computing community. These should actively engage in motivating the community to help drive the changes needed at the federal, state, local, and institutional levels to transform computing education for all students. Examples include promoting best practices through workshops, developing toolkits, disseminating publications, and coaching, mentoring or consulting. When describing resources, the alliance may wish to reference the resource type list from the BPC-A participation measures adapted from NSF INCLUDES.
  - **Locus.** Activities that serve the academic computing community, facilitate formation of public/private partnerships, act as a distribution point for educational reforms, and/or provide a foundation by which demonstration and other projects with organizations and others may build upon. Examples include creating a professional network, hosting research conferences, and knowledge brokering.

- **Sectors.** The sector of an intervention can focus on either the capacity building strategy or activity. In many cases, the participants in an activity may include folks from multiple sectors, and the alliance should consider the intention of the strategy when identifying where the effort is designed to impact change. Sector options for our shared reporting purposes include:

    - K-12
    - Higher Ed
    - Workforce: Academic
    - Workforce: Non-academic
    - Other

  Breaking out the K-12 and Higher Ed into different sectors ensures that the nuances between these groups is recognized and accounted for. Each alliance may wish to do their own reporting at a more granular level, for example looking at K-5, 6-8 and 9-12 as distinct parts of the K-12 system; or Community Colleges as distinct from undergraduate or graduate programs.

- **Timeframe.** Alliances have a long arc to their work, with many activities experiencing multiple iterations. We measured whether the activity was new, significantly modified, or continued from prior year(s). While new activities and activities continued from prior year(s) are simple to identify, whether an activity was slightly or significantly modified is more difficult to determine. In addition, having only one modification category (significantly modified) may not sufficiently tell the story of these arcs. Sometimes a slight tweak makes a huge impact, and sometimes an activity needs to be revamped more significantly. Finally, "continued from prior year(s)" may also not clearly tell the story. We have to be clear that this category includes evolution and innovation, not just "continuing' status, which could imply stagnation, although that is not its intent.

- **Resource support.** Alliances make use of multiple funding sources including additional NSF grants, other federal grants, state funding, and private donations. Knowing if an activity was solely supported by NSF BPC-A funds or if it also makes use of other funding helps determine the impact of NSF investments.

- **Scale.** Scale is measured across three levels: individual, organizational, and system. Understanding the scale of impact each activity has helps determine how an intervention might be contributing to shifting the education ecosystem and how it might be measured. This determination could lead to further work as the alliances explore how to share specific instrumentation. These levels are defined as:
  - **Individuals** These are changes that occur in an individual person as a result of participation. Often these are the easiest to measure and may include changes in attitudes, behaviors, knowledge, and/or skills. When describing individual participants, an alliance may wish to reference the participant type list from the BPC-A participation measures adapted from NSF INCLUDES.
  - **Organizations** Organizations are relatively closed systems in which the work is effecting change. For example, an organization could be a unit or department of a university, or it could be a university that is part of a larger network. In this case, context will determine what is an organization versus a system for the purposes of reporting. When describing an organization type, an alliance may wish to reference the partner type list from the BPC-A participation measures adapted from NSF INCLUDES.
  - **Systems** Systems are the complex interactions between organizational units; they may be explicit or implicit. Systems change can be hard to recognize but may be reflected through new patterns of operating or new policies, practices, relationships, approaches, and/or mindsets that lead to a different set of outcomes.

5 Shared Measures Results

In this section, we present the results of alliances' use of the toolkit, including a composite of strategies and activities across various sectors (K-12, Higher Ed, etc.) as well as a look at the forms of data collected by the alliances. A shared process of co-creation of the toolkit and collecting data was prioritized in this report and in the project. The results from the shared measures data collection are not definitive, as the project was focused on the process of co-creation rather than on results. That said, the results suggest that the alliances are, in aggregate, meeting the goals of the BPC-A program by providing important infrastructure and support across the education-workforce CS ecosystem. The infrastructure built by the alliances has created a space for novel activities and for an exploration of new ideas and approaches for BPC. This infrastructure also makes it easier to leverage contributions from companies, foundations, and additional grant funding.

Figure 5.1: A student participating in activities provided by the AccessComputing alliance.

5.1 Capacity Building Results

Across the alliances, there were 143 activities in 32 strategies, for a total of 124 distinct activities. Table 5.1 shows the number of strategies ($n = 32$) and activities ($n = 124$). The maximum and minimum number per alliance as well as the average (mean) are presented. We note here a wide range of activities per alliance.

When we examined the composite of shared measures from alliances, strate-

Table 5.1: Strategies and activities per alliance. *While some activities appeared under multiple strategies and thus were listed multiple times, the activity counts in this document include each activity only once. There were 19 duplicate activities, for a total of 143.

		Per Alliance		
	Total	**Max**	**Min**	**Average**
Strategies	32	6	2	4.0
Activities*	124	24	10	15.5

gies, and activities (Table 5.2), we found that activities were unique by sector but may be included in multiple strategies.

Table 5.2: Strategies and activities by sector.

Sector	# Strategies	# Activities
K–12	20	60
Higher Ed	26	89
Workforce: Academic	16	46
Workforce: Non-Academic	16	38
Other	9	14

When examining activities, we collected data across three three types, Testbed, National Resource, and Locus. Table 5.3 shows the activity count by sector. We see that the largest number of activities are focused on higher education.

Activities are measured across three scales: individual, organization, or system. Table 5.4 shows the scale across sectors. Here, there are more activities focused on higher education at the individual and organization level than at the systemic level or in other sectors. However, note that one activity at the system or organization level may lead to multiple activities at the individual level.

Table 5.5 shows activities by sector for new activities, significantly modified activities and activities continued from prior year(s), with 64 new, 18 significantly modified, and 174 continued from previous years. The high number of continued

Table 5.3: Activities by Sector and Type

	Testbed	National Resource	Locus
K–12	13	35	16
Higher Ed	23	42	35
Workforce: Academic	7	28	20
Workforce: Non-Academic	8	19	15
Other	0	10	5
Total # of Non-duplicated Activities	33	62	40

Table 5.4: Activities by Sector and Scale

	Individual	Organizational	System
K–12	42	36	21
Higher Ed	72	41	22
Workforce: Academic	38	17	15
Workforce: Non-Academic	31	16	9
Other	9	3	8

activities supports the notion that activities normally operate over multiple reporting periods. As with most of the areas on the form, it is difficult to cleanly parse modified from continued activities, so we urge caution on how these numbers are interpreted.

Table 5.5: Activities by Sector and Status

	New	Significantly Modified	Continued from prior year(s)
K–12	10	8	45
Higher Ed	22	10	61
Workforce: Academic	15	0	32
Workforce: Non-Academic	11	0	27
Other	6	0	9
Total # of Non-duplicated Activities	64	18	174

Table 5.6 shows the activities according to the presence of other resources. Other support isn't qualified by amount or type, making these values informative but also limited. We note that types ***Testbed***, ***National resource***, and ***Locus*** as well as ***New***, ***Significantly Modified***, and ***Continued*** align closely across ***BPC-A resources only*** and ***Other resources***.

Table 5.6: Activities by Presence of Other Resources

	BPC-A Resources Only	BPC-A + Other Resources
Testbed	18	15
National Resource	30	32
Locus	18	22
New	17	14
Significantly Modified	5	9
Continued	40	45
Total # of Non-duplicated Activities	62	62

One area that may be of interest for future work is how alliances are measuring their efforts. We examined the forms of data collected by the alliances. We sought to identify which forms of data for evidence of activities' impact were collected quantitatively, qualitatively, or not measured at all. The question was kept highly general, knowing that this project did not have the resources for a deeper dive on this important topic. Table 5.7 shows the measurement types used by activities across the three scales of impact. Here, we see that most data is measured quantitatively. We note that 61% of system-level activities are not measured, probably because measuring activities at the system level can be challenging.

5.2 Participation Measures

The NSF INCLUDES program is similar to the NSF's BPC alliance work, so a comparison between the shared measures used by INCLUDES and those developed by the alliances may be instructive, especially since the alliances used the shared measures developed by INCLUDES as a starting point for their own measure development. Some measures were adopted in the alliances work with virtually no

Table 5.7: Total Measured and Types of Measurements by Scale. Some scales are measured both qualitatively and quantitatively, so these percentages should not be added.

	N	**Measured**	**Measured Quantitatively**	**Measured Qualitatively**
Individual	90	91%	87%	46%
Organization	62	79%	66%	35%
System	36	39%	31%	25%

changes, such as the listing of resource categories. In other cases, the same measures were adopted but definitions were added: the alliances defined the terms 'disability' (for K12: students served under IDEA or Section 504; for higher education: students registered with the disabilities office) and 'low income' (for K12: students receiving free or reduced price lunch; for higher education: students receiving Pell Grants). Table 5.8 shows instances where the alliances created more precise categories for participants than were used in the INCLUDES shared measures.

Table 5.8: Participant Categories.

INCLUDES	**BPC alliances**
PreK-12 teachers and administrators	–PreK-12 teachers –PreK-12 administrators and educational leaders
Graduate students	–Master's students at IHEs –Doctoral students at IHEs –Postdoctoral fellows at IHEs
Faculty and administrators at 4-year IHEs	–Faculty and administrators at 4-year IHEs –Faculty and administrators of graduate programs
Other professionals	–Other non-PreK—12/IHE professionals –Non-academic researchers –Other alliance

Figure 5.2: CAHSI student, Alejandro Chavez, presenting his research at the GMiS conference.

The alliances also combined the categories of race and ethnicity. These changes show that, even when programs are quite similar in structure (both NSF-funded, for example) and goals (both focused on improving equity in STEM education), wholesale adoption of shared measures across different groups may not best meet the needs of all groups, although it can be productive to use another group's shared measures as a starting point instead of beginning de novo.

6 Discussion and Lessons Learned

The process of developing shared metrics across a complex NSF program allowed for deep discussions about the nature of the program that includes objectives, strategies, activities, outcomes and potential impacts. This process provided a venue that allowed broad scope thinking about what was common across the alliances as well as prompted some reflection at the project level that was self-reflective. The attempts at specific definitions that allow for tracking participants and describe capacity-building, as reflected in the attached templates, were useful in driving the conversation. They were also useful in providing shared language across the alliances in terms of what worked and what did not as descriptors of multiple-years effort to address broadening participation in computing.

There were a set of key breakthroughs in thinking that are worthy of being called out as "lessons learned". For the participation data, building on the work of others allowed for more quickly agreeing on where new work and thinking could be brought forward. One key concept that emerged from these discussions was the importance of allowing for cumulative data over time. The attention to direct and indirect impact was also critical for alliances that often struggle to demonstrate their value in a numeracy-based climate when, by design, they don't engage with students or educators directly.

Bringing forward NSF's concepts of test bed, locus and national resource as a framework for describing capacity building, accelerated the work on mapping potential metrics associated with alliance activities and strategies. This aligned with the language of the solicitation that funds all alliances. However, these terms are meant to be applied to a whole project or grant. But as an alliance, participants found that they were applying these terms at the ***activity*** level, not the alliance level, indicating a bit of a gap in the alignment with NSF.

The discussions associated with appreciating the distinctions between out-

puts[1], outcomes[2] and impacts[3] were important in casting the position of the metrics. The group was very mindful that this phase of the work was primarily focused on outputs and outcomes and was able to move forward in this stage aware that addressing long term impacts would need to be part of future work.

The valuable knowledge that resulted from this project also indicated areas that can be further developed. In phase one, when the alliances were asked to comment on whether the metrics used by the NSF INCLUDES project would be workable, not workable, or workable if modified, there was no option for not applicable. As a result, when a metric was not applicable (e.g., number of elementary school students who are female for an alliance that only work with college students), their responses may have been workable (suggesting that this category would work even if it was not relevant) or not workable (since the alliance did not have participants in this demographic). Additionally, some respondents appeared to answer the question based on whether they did or did not currently collect a particular data point – not whether the definition and phrasing would be workable if they collected that data in the future. Thus, revision of the survey to

Figure 6.1: Students participating in activities presented by the AccessComputing alliance.

[1] Outputs are directly produced by the project and are often tangible and easy to measure; however, they are not the reason why the project was necessary. An example may be five workshops.

[2] Outcomes are produced by the outputs. For example, the five workshops may have changed participants' knowledge of a particular topic.

[3] Impacts are the result of outcomes being put into place and reflect a a more significant, broader change. For example, participants who incorporated their knowledge into practice would impact the students that they support.

provide more clarity and direction may have improved the data.

When reducing multi-million dollar projects to numbers, there is a shared concern that the context will be lost and the results will not accurately portray the collective work and impact of the alliances. Many valid concerns have also been raised that center around accountability, utility, and meaning.

Other limitations of the project include:

- Only eight alliances ultimately participated in the data collection process.
- Every activity was counted the same. We did not scale based on number of participants, budget, or scope.
- Alliances may have applied the definitions (for example, of a 'strategy' or a 'locus') differently. This was expected and is part of the challenge inherent to shared measures.
- Alliance may not have included all of the activities that they engage in related to the non-academic workforce since this data is not usually reported by specific activity to the NSF.
- Alliance members might have data entry errors.
- The status designations of new, significantly modified, or continued are very broad and may not adequately describe activities.
- We wanted to examine both strategies and activities, but the analysis ended up focusing on activities due to their more concrete nature.

7 Conclusion and Future Work

The sentiment, often attributed to Albert Einstein, that "not everything that can be counted counts and not everything that counts can be counted" encapsulates the conundrum inherent in the effort to establish shared measures. At the same time, data that can be aggregated has a stronger likelihood of becoming an impact magnifier, especially in situations where the number of participants with (multiple) marginalized identities is relatively small. Shared measures may enable greater analysis of the CS experiences of these and other students served by the BPC alliances. One participant in phase two noted that future funding for BPC could be imperiled without evidence that recipients of previous rounds of funding have been successful at actually broadening participation in computing. To the extent that the adoption of shared measures makes it easier to convey the success of research efforts, those measures may be foundational to ensuring continued support for the program. This applies doubly to systems change efforts.

Future shared measures work may involve implementing the shared measures and then analyzing lessons learned and future directions based on that implementation. But as one participant mentioned, the goal was to "get to good enough, and try it out" with the expectation that more refinements will be necessary over time. Even so, based on the experience, another participant noted that the the notion of a shared NSF database for certain kinds of data could also hold potential.

Future work can also include the exploration of the commonalities of activity-specific measures across the alliances, linking participation data to capacity building efforts, and further understanding partnerships within and across alliances. Future work will also require an active facilitator and resources to refine the toolkit.

References

Bekhet, A. K., & Zauszniewski, J. A. (2012). Methodological triangulation: An approach to understanding data. *Nurse Researcher*, *20*(2).

Blaser, B., & Ladner, R. E. (2020). Why is data on disability so hard to collect and understand? *Proceedings of the 2020 Research on Equity and Sustained Participation in Engineering, Computing, and Technology (RESPECT)*, *1*, 1–8.

Clarke, M., Brice, A., & Chalmers, I. (2014). Accumulating research: A systematic account of how cumulative meta-analyses would have provided knowledge, improved health, reduced harm and saved resources. *PloS one*, *9*(7), e102670.

Clifton, C., Kantarcioglu, M., Doan, A., Schadow, G., Vaidya, J., Elmagarmid, A., & Suciu, D. (2004). Privacy-preserving data integration and sharing. *Proceedings of the 9th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, 19–26.

Code.org, CSTA, & Alliance, E. (2022). *2022 State of Computer Science Education Understanding Our National Imperative* (tech. rep.).

Denzin, N. K. (2017). *The research act: A theoretical introduction to sociological methods*. Transaction Publishers.

Dunton, S., Zarch, R., Xavier, J., Warner, J., & Peterfreund, A. (2022). Determining metrics for broadening participation in computing: Connecting data to multi-state computer science education policy efforts. *Policy Futures in Education*.

Education Development Center. (2017). Evaluation of the National Science Foundation's Broadening Participation in Computing—Alliances Program: Year 5 Report.

Fletcher, C. L., & Warner, J. R. (2021). CAPE: A framework for assessing equity throughout the computer science education ecosystem. *Communications of the ACM*, *64*(2), 23–25.

Goodyear, L., Silverstein, G., & Thurston, L. P. (2017). The influence and promise of alliances. *Communications of the ACM*, *60*(6), 29–30.

Google & Gallup. (2020). Current perspectives and continuing challenges in computer science education in US K-12 schools.

Kadadi, A., Agrawal, R., Nyamful, C., & Atiq, R. (2014). Challenges of data integration and interoperability in big data. *2014 IEEE international Conference on Big Data*, 38–40.

Mcklin, T. Common core indicators for describing alliance programs. In: American Evaluation Association 2012 Annual Meeting. 2012.

Mitchell, E. S. (1986). Multiple triangulation: A methodology for nursing science. ***Advances in Nursing Science***, ***8***(3), 18–26.

Osborne, J. W. (2012). ***Best practices in data cleaning: A complete guide to everything you need to do before and after collecting your data***. Sage Publications.

Searls, D. B. (2005). Data integration: Challenges for drug discovery. ***Nature Reviews Drug Discovery***, ***4***(1), 45–58.

Stonebraker, M., Ilyas, I. F., et al. (2018). Data integration: The current status and the way forward. ***IEEE Data Eng. Bull.***, ***41***(2), 3–9.

Zarch, R., Xavier, J., & Peterfreund, A. (2019). Using state-based data systems to support broadening participation in computing. ***Proceedings of the 2019 Research on Equity and Sustained Participation in Engineering, Computing, and Technology (RESPECT)***, 1–1.

Zatz, M. S., Gates, A. Q., & Santiago, D. (2017). Advancing the Collective Impact of Retention and Continuation Strategies for Hispanics and Other Underrepresented Minorities in STEM Fields. ***Proceedings of the NSF INCLUDES Conference***.

Appendix A. Template and Instructions

A.1 Basic Capacity Building Template

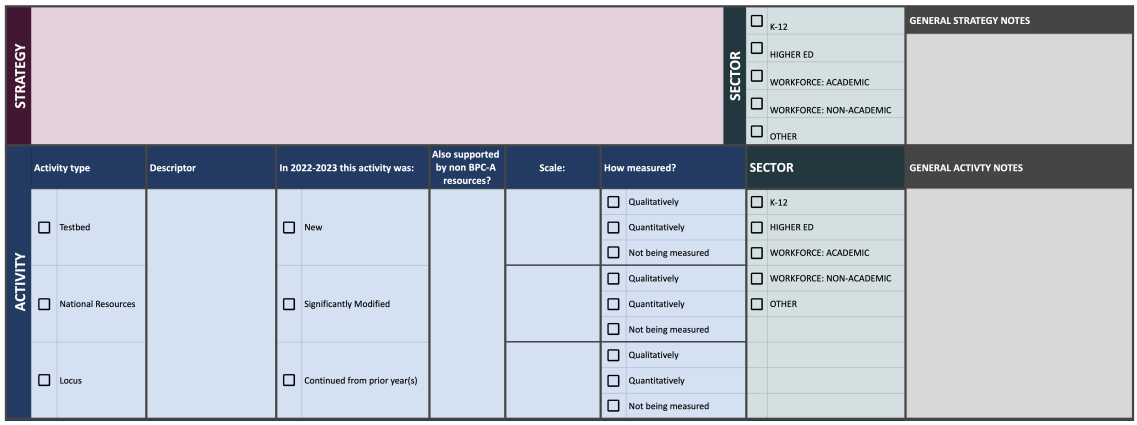


Figure 7.1: Template used for collecting data.

A.2 Guiding Terms and Definitions

All definitions are intended as guidance and starting points. They will evolve and have greater clarity as we empirically define them through our work. Each alliance should use their best judgment for their context when thinking through how to apply the terms in this document.

Capacity Building The NSF INCLUDES defines Enabling Sustainable Changes in Systems as "projects that take actionable steps to transform policies, practices, relationships, approaches, and/or mindsets, with the goal of making STEM cultures more inclusive, advancing equity, and broadening participation in STEM." The Shared Measures participants have been discussing Capacity Building as the variety of strategies and activities that support making STEM cultures more inclusive through the transformation of policies, practices, relationships, approaches,

and/or mindsets. As alliances work with the reporting templates, this definition can be empirically refined and expanded.

Strategy Strategies are the long-term approaches to meeting the project goals. Strategies are supported by a set of activities that may build upon each other over time. Early in this work, each team shared their overarching goals and measurable objectives. Strategies are the approaches used to support the goals.

Activity Individual activities are the components of a strategy. For example, a workshop or a conference is usually part of a broader approach to bringing together a community. NSF annual reports have been excellent at gathering the activities that an alliance engages in each year. These are often, but not always, more easily measured and have "participants" who can more easily be counted than a strategy can be.

A.2.1 Types of Activities

(These definitions are slightly modified from those in the NSF BPC Solicitation based on group feedback.)

Testbed BPC Alliances should develop, test, and deploy interventions aimed at supporting students and faculty. These efforts should also focus on sustaining institutional transformation and inclusive practices at the departmental and organizational levels. Examples may include new/emerging professional development, new resource development and conducting.

National Resource: BPC Alliances should generate, curate, collect, vet, and disseminate best and promising practices for addressing underrepresentation to inform, educate, and connect the broader computing community. They should actively engage in motivating the community to help drive the changes needed at the federal, state, local, and institutional levels to transform computing education for all students. Examples may include promoting best practices through workshops, developing toolkits, publications, and coaching, mentoring or consulting. When describing resources you may wish to reference the resource type list from the BPC-A participation measures adapted from the NSF INCLUDES.

Locus BPC Alliances are expected to serve the academic computing community. They should facilitate formation of public/private partnerships, act as a distribution point for educational reforms, and provide a foundation by which demonstration and other projects with organizations and stakeholders may build upon.

Examples may include creating a professional network, hosting research conferences, and knowledge brokering.

When trying to decide how to classify an activity it may help to consider the intention of the effort and the currency of the activity. The BPC-A workgroup recommended national resources be considered static; these are resources that someone can participate in, or access with minimal guidance. A locus, however, requires much more significant facilitation and active engagement between participants. In some cases this may be a blurry line such as a professional development program. If a program has developed a robust training that is delivered to participants, it might be a national resource; however, if it requires negotiation and evolves due to the sustained interaction of participants and facilitators, it could be a locus.

Many activities will move between types; for example, they might start start as a testbed, which, once established will become a national resource. Reports should be for the status of the activity in the current reporting year. Alliances will determine which classification makes sense for their activities.

Development of activity:

- New: 2022-2023 was the first time it was offered
- Significantly Modified: The activity may have been piloted or offered in the past but looked substantially different in the 2022-2023 academic year
- Continued from prior year(s): the activity may have evolved based on participant feedback but fundamentally has similar goals and structure to prior years.

A.2.2 Scale of Intervention

The participants identified three types of domains in which the strategies are affecting change, with activities often addressing more than one, and often all three, domains. These definitions will need to be defined more empirically over the course of the Shared Measures implementation phase by the community, and what is offered below should help guide initial applications.

IndividualThese are changes that occur in an individual person as a result of participation. Often these are the easiest to measure and may include changes in attitudes, behaviors, knowledge or skills. When describing individual participants

you may wish to reference the participant type list from the BPC-A participation measures adapted from the NSF INCLUDES.

OrganizationsOrganizations are relatively closed systems in which the work is affecting change. For example, an organization could be a unit or department of a university, or it could be a university that is part of a larger network. In this case, context will matter when determining what is an organization versus a system. When describing an organization type you may wish to reference the Partner Type list from the BPC-A participation measures adapted from the NSF INCLUDES.

SystemsSystems are the complex interactions between organizational units that may be explicit or implicit. Systems change can be hard to recognize but may be reflected through new patterns of operating emerge, or new policies, practices, relationships, approaches, and/or mindsets that lead to a different set of outcomes.

A.2.3 Sector of Intervention

The sector of intervention can focus on either the capacity building strategy or activity. For the purpose of this reporting mechanism, we are focusing on the strategy. In many cases, the participants in an activity may include folks from multiple sectors, and the alliance should consider the intention of the strategy when identifying where the effort is designed to impact change. Options for our shared reporting purposes include:

- K-12
- Higher Ed
- Workforce- Academic
- Workforce- Non-academic
- Other

Each alliance may wish to do their own reporting at a more granular level, such as defining K5, 6-8 and 9-12 as distinct parts of the K-12 system or at community colleges as distinct from undergraduate or graduate programs.

A.2.4 Additional Resources

Many strategies and activities are supported by resources outside of the BPC-A grant. Additional resources may include financial support, either from other grants, foundations, corporations, or sponsorships. They may also include in-kind contributions such as physical space for meetings or human resources such as experts, consultants, or volunteers.

A.3 Capacity Building Measures Template Instructions

The purpose of this template is to prompt each alliance to think through how capacity building can be described to reflect their BPC-A program. The template is a spreadsheet that currently has one strategy field supported by five activity fields. To add additional strategies, copy and paste this template. Similarly, if you have more than five activities, copy and paste an activity block. If you have fewer than five activities in a strategy area, then you can cut the extra activity blocks.

Refer to Section A.2.2 for definitions.

In all cases, please use your best judgment for how to describe your alliance; consistency within an alliance matters more than precision across alliances.

A.3.1 The workbook

Each workbook has six tabs. Tab one is a blank template. You may copy this onto new tabs if necessary. Subsequent tabs are designed for one strategy per tab (five tabs to start). You may rename tabs if it is easier to navigate. Complete one tab for each strategy.

A.3.2 How to complete the template

1. Describe the strategy.

2. Indicate the sector(s) this capacity building strategy is designed to influence.

3. Each strategy is supported by a set of activities. For each activity, indicate if it is a testbed, national resource, and/or locus. Although there are "all that apply" options, each activity will likely only fit into one category.
4. If you have more activities, copy the activity rows.
5. Describe the activity.
6. Indicate if this activity was new, significantly modified, or continued from prior years.
7. **Also supported by non-BPC-A resources:** indicate if there is a significant investment of non-BPC-A resources including other grants from foundations, companies, or state/local agencies; significant in-kind donations; significant volunteer support or other resources outside of the BPC-A funding.
8. **Scale:** There is a set of dropdown menus for indicating the scale. Think of these as "all that apply" for the activity-they do NOT need to align with the preceding rows (activity type, if the project was new). A project may reach individuals AND organizations.
9. **How Measured:** This question is geared towards future work and aligns with the scale. For each scale indicated, identify if the activity is being measured at that level of scale. If so, is it being measured qualitatively or quantitatively? For example a professional development activity may be affecting individuals AND organizations AND systems. There might be the following measures:
10. **Individual:** Surveys examining participant satisfaction (could be quantitative and/or qualitative)
11. **Organization:** May be quantitatively measured by tracking percent of faculty with the training.
12. **System:** May not be measured yet.
13. **General Notes:** Use this space liberally to document any important information that does not otherwise get included in the template or to raise any questions, concerns, or feedback about the template.

A.2 Participant Measures Template

Table 7.1 is used to collect general information about who the participants are in the alliance. Table 7.2 is used to identify the artifacts produced by the alliance to share knowledge about the alliance and its programs.

Table 7.1: Partners of the alliance.

Organization type	**Created (new) in 2022-2023**	**Sustained in 2022-2023**
Community-based organization / foundation / other non-profit		
Company		
Federal agency or lab		
Institution of HIgher Education (IHE)		
IHE-Affiliated center or program		
Museum or science center		
PreK-12 school		
State or local education agency (SEA /LEA)		
State or local agency (Non-SEA /LEA)		
BPC Alliances		
Other (Specify)		

Table 7.2: Products of the alliance.

Products	**2022-2023 (count)**
Book/Chapter	
Brief (resource, practice or research, white paper, prospectus)	
Conference/Workshop paper or presentation	
Journal or juried conference paper	
Other conference presentation / paper	
Newspaper/newsletter article or blog	
Project website	
Webinar / video	
Conference	
Co-sponsorship/funding distributed	
Evaluation report	
Knowledge base	
Other (Specify)	