Measuring Youth Development: How Out-of-School Time Programs Collect and Use Data

Hannah Lantos, Zakia Redd, Jessica Warren, Michael Bradley, and Sham Habteselasse

ChildTrends  Wallace  February 2024
Contents

Acknowledgements........................................................................................................................................................................... ii
Executive Summary .................................................................................................................................................................................. 1
Introduction........................................................................................................................................................................................... 5
Chapter 1: Methods .............................................................................................................................................................................. 7
Chapter 2: Outcome Measurements.................................................................................................................................................. 10
Chapter 3: Factors Programs Consider When Deciding What Outcomes To Measure............................................................... 16
Chapter 4: Measurement Challenges ........................................................................................................................................... 20
Chapter 5: Key Findings And Recommendations..................................................................................................................... 24
Appendices .............................................................................................................................................................................................. 24
  Appendix I. Additional References........................................................................................................................................... 29
  Appendix II. Bibliography For Ost Measurement Literature Scan........................................................................................... 31
  Appendix III. Expert Informants ........................................................................................................................................... 49
  Appendix IV. Programs That Participated In The Study................................................................................................................ 50
  Appendix V. Data Collection Protocols................................................................................................................................... 52
  Appendix VI. Survey................................................................................................................................................................... 58
References ............................................................................................................................................................................................... 62
Acknowledgements

Numerous people helped develop this report. We would first like to recognize and thank the individuals from the 38 programs who participated in the study and shared their experiences, insights, and data collection tools through interviews and surveys. We also thank the 23 experts for providing recommendations for programs to include in this review.

In addition, we thank Mark Waits and Brent Franklin of Child Trends for their careful reading and editing of this report and helpful suggestions. We also thank our senior reviewers, Kristin Anderson Moore and Emilie Phillips Smith, for their thoughtful critique and extensive comments.

Finally, we express our gratitude to The Wallace Foundation for commissioning and funding the research on which this report is based. We especially thank Bronwyn Bevan for her ongoing guidance, substantive feedback, and support. We also thank Gigi Antoni, other members of the Youth Development team, and Pam Mendels and Daniel Browne of Wallace’s editorial team for their insights, recommendations, and review.

Executive Summary

Out-of-school-time (OST) programs and their funders rely on sound data to make decisions about everything from professional development and student recruitment to the selection of activities to offer students. Programs operate at a range of times (before and after school, weekends, summer) and in a variety of locations (e.g., schools, community-based organizations, city parks and recreation centers), are run by a variety of entities (e.g., government agencies, private community organizations), and receive funding from a variety of sources (e.g., government, philanthropy)—each of which may be interested in a different set of data and come with its own reporting requirements. This means there is a great deal of variation in the types of data programs collect.

In 2019, The Wallace Foundation (Wallace) commissioned Child Trends to conduct a study of the kinds of youth outcomes OST programs are interested in measuring, the tools they use to measure those outcomes, and the challenges they experience in doing so. The study included a literature scan and interviews with leaders and staff members at 28 OST programs. Twelve of the 28 also completed surveys; a separate group of 10 provided information by survey only. The study expands on past research by a) focusing on programs that work in specific content areas (e.g., the arts, civic engagement and social justice, career and workforce development) and b) covering both quantitative approaches (i.e., tracking numerical data) and qualitative approaches (gathering descriptive information through surveys, interviews, etc.) to data collection.

While its findings apply to OST programs in general, the study focused on particular types of programs (i.e., afterschool, summer, online) and particular content areas, as well as programs that serve school-age children and adolescents from marginalized communities, those that support students’ social and emotional learning (SEL), those that serve systems-involved youth, and those that focus on promoting equity—for example, by training staff to recognize and overcome personal biases or by recruiting and retaining leaders and staff who reflect the diversity of the participants served.

Key findings

What outcomes do programs measure?

- Programs in the study measured outcomes that were closely related to the content they delivered (e.g., the arts, career and workforce readiness, civic engagement and social justice). They also generally measured SEL outcomes and other outcomes required by funders.

- Programs thought critically about how to measure progress toward achieving equitable outcomes for the youth they serve. Approaches included disaggregating data by race and other demographic variables and tracking the development of equity-related knowledge and skills.

- Programs used a number of criteria to decide what outcomes to track, including whether a given outcome was consistent with the program’s logic model or theory of change, usability and shareability of the data, the effort and capacity required to collect the data, availability of valid and reliable measures, youth interest, and whether measuring a given outcome would promote equity.

- In addition to outcomes, programs consistently measured outputs (i.e., steps that lead to desired outcomes), particularly program participation and quality. This finding aligns with our literature scan that revealed that public and private funders have invested heavily in quality assessment tools and participation tracking systems. Comparatively few programs used qualitative methods to understand the factors that contribute to program participation and quality.
What methods do programs use to measure progress?

- Programs reported using a variety of traditional methods to measure outcomes, including quantitative methods—like administering surveys and questionnaires, conducting formal assessments, and tracking attendance rates—and qualitative methods such as conducting interviews and focus groups with young people. Their use of quantitative methods was well-documented, but their use of qualitative methods less so.

- Programs also used nontraditional methods, such as regular check-ins with participants and reviewing participant journals and portfolios, along with creative, informal ways to track SEL outcomes, including games and award systems. As with their use of qualitative measures, use of these nontraditional methods was not well-documented.

What measurement challenges do programs face?

- The programs we interviewed did not consistently identify specific outcomes of interest that they were unable to measure, although some said they lacked the tools they needed to track equity- or SEL-related outcomes or program quality.

- Programs pointed to broader types of outcomes that were challenging to measure, including longer-term outcomes like college matriculation, career attainment, and participation in civic life; behavior change (e.g., whether participants in a civics program exercise their right to vote); school outcomes like test scores and grades that require a data-sharing agreement with the district; and the relationship between improvements in staff members skills and knowledge and youth outcomes.

- Programs reported that the process of collecting data could be burdensome for both participants, who have their fill of testing at school, and staff, who in many cases do not have the training or time to do the work. Some pointed specifically to the burden of fulfilling the reporting requirements of various funders, which often involved recording duplicative information using multiple tools and databases.

Recommendations for OST programs

Select a small number of high-priority outcomes to measure.

Programs often cast a wide net for data and end up not using the data they collect, which can be frustrating for staff and participants. To avoid this, programs should focus on data they will be able to use and share with participants, families, and the broader community.

Ensure that the selected outcomes fit into the program’s logic model or theory of change.

Many programs reported they did not use their logic model or theory of change to decide what outcomes to measure or guide their data collection efforts. Programs should review the outcomes they target for consistency with their logic model or theory of change and, if necessary, make changes to one or the other.

Document the type, source, and purpose of each piece of data collected.

Programs were more likely to keep a record of the data they collected using traditional, quantitative methods like surveys or formal assessments than data collected using qualitative or nontraditional methods like check-ins with participants or portfolios of their work. Programs should thoroughly document all the different ways they collect data to ensure that all data get used and to better understand what each method contributes to their work.
Make room for youth perspectives in decision making about what outcomes to measure and how.

It is important that programs find out what young people want to get out of their participation (e.g., to learn a particular skill or to get a job in a particular field) and their progress toward those goals. Programs can solicit input from participants not only in deciding what outcomes to measure but also in assessing how well the data collection process is working and in determining the meaning of data collected.

Incorporate measures of diversity, equity, and inclusion into data collection efforts.

Programs interested in tracking their progress on issues of equity should consider disaggregating data by race or other demographic variables to identify disparities in participant outcomes; incorporating outcomes specifically related to equity (e.g., the development of civic knowledge and social justice skills) into their logic models or theories of change; revising their curricula to promote those outcomes; and selecting qualitative and, if possible and appropriate, quantitative methods of measuring those outcomes.

Assess the capacity and effort required to collect data and take steps to minimize burnout.

Programs should expect finding the right approach to data collection to be an ongoing process of trial and error. One way to reduce the burden on staff is to schedule data collection for a few specific points in the calendar when data will be most relevant rather than trying to collect it continuously.

Recommendations for funders

Connect programs to additional training in key aspects of data collection and use.

Many staff members who oversee their programs’ data efforts are trained in youth development and are learning to work with data on the job. Interviewees spoke of the need for more training in assessing which measurement tools are valid and reliable; developing new databases or navigating the systems they have; and using data to support program improvement.

Create networks of programs to provide informal technical assistance.

One program secured a grant to create a network of OST programs that could support and learn from one another. Through this network, the program learned how other programs have solved common problems and worked with those programs’ technical assistance specialists to troubleshoot its own particular challenges. This program noted it would also benefit from participation in networks focused on developing specific skills (e.g., collecting longitudinal data, identifying outcomes of interest and ways to measure them).

Provide funding for additional staffing.

Tight budgets make it difficult for programs to hire dedicated staff for data collection and analysis, and the existing staff are focused on providing the services youth need and don’t have sufficient time to support data-related tasks. Additional resources would allow programs either to hire staff whose primary responsibility is data or reduce staff-to-participant ratios so all relevant staff members have more time for data activities.

Partnerships between practitioners and researchers can be mutually beneficial. Researchers can provide programs with technical assistance in identifying what outcomes to measure, selecting valid and reliable measurement tools, and analyzing the data they collect, while programs can provide researchers with real-world data to analyze and an opportunity to field-test the tools they develop.
Make reporting requirements as flexible as possible.

Almost every program in our study mentioned that meeting the requirements of their various funders was a time- and resource-intensive process that involved recording information that was duplicative or not meaningful to them, using multiple tools and databases. Funders should be sensitive to the strain that data collection can place on programs and be as flexible as possible in terms of how and in what format grantees demonstrate progress toward outcomes of interest.

Support digital data collection.

As many programs moved online during the COVID-19 pandemic so did their data collection efforts. Even as they return to serving young people in person, programs could benefit from additional funding for the development of mobile data collection processes and training to do this work.

Recommendations for further research

Expand the existing study to include more programs and voices.

The current study could be expanded by including a) more programs in each of the categories we prioritized; b) programs in additional categories (e.g., STEM, physical fitness, identity development) to increase the generalizability of the findings; and c) perspectives from different types of people involved with OST programs, particularly youth participants.

Further explore how programs make decisions about their data efforts.

Data collection and use require program leaders to make difficult tradeoffs in terms of money and staff time. To help programs navigate these tricky decisions, future research could focus on the following:

- How do programs identify outcomes meaningful to them and prioritize those outcomes?
- How do they choose the most appropriate measurement tools for their needs?
- How do they balance spending on improved data collection with spending on programming?
- What types of support for data collection and use (e.g., training, financial, staffing) do they find most helpful?

Highlight innovative solutions to common problems.

Future research could identify how exemplary programs tackle common data-related problems. Topics could include forming successful data-sharing partnerships, minimizing the burden of data collection on staff, building staff capacity to collect and analyze data, and securing funding for data efforts.

Investigate strategies for collecting longitudinal data.

Programs expressed interest in tracking long-term participant outcomes but had difficulty doing so. Many programs specifically mentioned the challenge of collecting data from alumni who may no longer be connected to the program. Future research could identify strategies for effectively measuring long-term outcomes by conducting a formal review of best practices in the field and gathering qualitative evidence from programs that have done it successfully.

Catalog existing measurement tools and develop and test new tools as needed.

Through our literature scan, we identified a robust set of tools for measuring program quality and SEL. Yet some programs were unaware of them. Researchers could help address this disconnect by cataloging the
tools already in use in the field and assessing their validity and reliability. The literature scan also revealed a lack of tools relevant to content-specific, equity-related, and other types of outcomes. Researchers could develop new tools to fill this gap. Funders could support this work by putting out a call for research in field publications; organizing panels or conference sessions focused on this topic; or forming a network of practitioners, funders, researchers, and content experts to identify and share promising approaches. Funders could also spearhead efforts to get new and existing tools into the hands of programs.

Introduction

“What gets measured gets done,” the saying goes. The idea that collecting and using data can play a powerful role in shaping an organization’s work certainly applies to out-of-school-time (OST) programs. OST programs and their funders rely on sound information to make decisions about everything from professional development and student recruitment to selecting activities to offer students.

The OST field, which includes afterschool, summer, weekend, and before school programs, is broad and diverse. Programs operate in a variety of locations (e.g., schools, community-based organizations, city parks and recreation centers), are run by a variety of entities (e.g., government agencies, private community organizations), and receive funding from a variety of sources (e.g., government, philanthropy), each of which may be interested in a different set of outcomes and come with its own reporting requirements. For example, the 21st Century Community Learning Centers program focuses primarily on academic outcomes, while Child Care and Development Block Grant funders may be more interested in understanding the extent to which programs help meet the needs of working parents. The breadth and diversity of OST programs, funders’ requirements, and characteristics result in significant variation in the types of data programs collect.

Historically, there has been a mismatch between the type of benefits OST programs typically confer and the type of data they collect. The field has long emphasized the social and emotional development of young people, and yet programs primarily tracked outcomes related to academic achievement. Recognizing this mismatch, The Wallace Foundation, a long-standing funder of OST, along with other funders, has invested in developing, testing, and cataloging measures of social and emotional outcomes for OST programs.

In 2019, Wallace commissioned Child Trends to study the kinds of youth outcomes OST programs are interested in measuring, the tools they use to measure those outcomes of interest, and the challenges they experience in doing so. The study included a literature scan and interviews and surveys of OST program leaders and staff members. It expands on past research by a) focusing on programs that work in specific content areas (e.g., the arts, civic engagement and social justice, career and workforce development) and b) covering both quantitative approaches (i.e., tracking numerical data) and qualitative approaches (gathering descriptive information through surveys, interviews, etc.) to data collection. It is common for OST programs to use qualitative methods but less common for them to use such methods in formal program evaluations.

While its findings apply to OST programs in general, the study focused on particular types of programs (i.e., afterschool, summer, online) and particular content areas, as well as programs that serve school-age children and adolescents from marginalized communities, those that support students’ social and emotional learning, those that serve systems-involved youth, and those that focus on promoting equity, for example by training staff to recognize and overcome personal bias or by recruiting and retaining leaders and staff who reflect the diversity of the participants served.

This report is organized into the following chapters:

- Chapter 1 provides an overview of the methods used to collect information for this study.
- Chapter 2 focuses on findings related to outcomes valued and measured by the OST programs interviewed for the study and the measurement techniques used to gather information on outcomes.
Chapter 3 highlights factors programs considered when measuring outcomes.

Chapter 4 focuses on challenges experienced by programs as they collected or tried to collect data on outcomes of interest.

Chapter 5 includes key findings from the study and recommendations for the field.

Appendices are included at the end of the report to provide additional contextual information on our approach and findings.
Chapter 1: Methods

Research questions

The study was designed to answer the following three research questions:

1. What is the range of intended outcomes for children and youth attending OST programs, including, but not limited to, programs that focus on the arts, civic engagement and social justice, career and workforce development, and general or other services?

2. What are the different measurement approaches and instruments used to document OST program outcomes, including formative measures (i.e., assessment that occurs during the program) and summative measures (i.e., assessment that takes place at the end of the program)?

3. What are the gaps in and barriers to documentation and measurement where intended outcomes are either not measured or are not articulated because of a lack of a measurement strategy, skills, resources, or instruments?

The first step in answering these questions was to conduct a scan of the literature for studies of OST programs focused on measuring outcomes in the areas listed in Table 1 below. Studies were considered “relevant” if they focused on measuring outcomes. Studies were categorized as “not relevant” if they were about program quality or implementation or were descriptive in nature.

Table 1. Studies of OST Programs Focused on Measuring Outcomes, by Outcome Type

<table>
<thead>
<tr>
<th>Outcome Type</th>
<th># of Relevant Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social and Emotional Learning</td>
<td>93</td>
</tr>
<tr>
<td>Academics</td>
<td>74</td>
</tr>
<tr>
<td>STEM</td>
<td>60</td>
</tr>
<tr>
<td>Arts</td>
<td>23</td>
</tr>
<tr>
<td>Identity and equity</td>
<td>22</td>
</tr>
<tr>
<td>Civic engagement</td>
<td>11</td>
</tr>
<tr>
<td>Unclear from abstract</td>
<td>15</td>
</tr>
</tbody>
</table>

Nearly a decade has passed since the last comprehensive reviews focused on measurement tools and strategies and data use in OST programs were published. (Wilson-Ahlstrom et al., 2014; Hair et al., 2001; Bronte-Tinkew, Moore, and Shwalb, 2006). More recently, there have been several reviews of findings from OST programs (McCombs, Whitaker, & Yoo, 2017; Sloan McCombs et al., 2019; McCombs et al., 2017; Lantos et al., 2021). This study did not focus on the types of outcomes OST programs are likely to produce but rather on how programs identify and track those outcomes.

Sampling approach

Following the literature review, we set out to identify a sample of OST programs with the intention of conducting surveys and in-depth interviews with staff members to understand the outcomes they prioritize, how they measure those outcomes, the barriers to successful measurement, and the complex decisions they
must make regarding the selection and measurement of outcomes. The purpose of the brief, initial survey was to help the interviewers gather background information (including the program's theory of change or logic model and data collection tools), so they could probe with deeper, more complex questions during the interview.

We decided to target OST programs identified as strong by experts in the field. We hypothesized that strong programs would be more ready for evaluation and have greater capacity for data collection than the average program. Narrowing our focus to a subset of all OST programs would also help us minimize irrelevant variables and reach "saturation," the point at which we had conducted enough interviews to confidently identify common themes.

In order to explore how programs measure equity and any barriers to doing so, we used a purposive sample (i.e., a sample in which participants are selected based on desired characteristics), focused on programs serving diverse populations, communities of color, and justice-involved youth.

Child Trends asked 35 experts in the OST field—including leaders from national policy and advocacy organizations, national youth-serving organizations, researchers, and funders—to recommend two to five strong, high-quality programs to be part of this sample. (See Appendix III. for the list of the experts who provided program recommendations). Experts possessed broad knowledge of the OST field or specific knowledge about a segment of the field, such as summer programs, arts programs, or programs that serve youth engaged in the juvenile justice system. Outreach to the experts consisted of an email that briefly described the study, our purpose for reaching out to them, how they were referred to us, and the types of programs in which we were interested. We offered to have a videoconference, if necessary, to gather recommendations, but most experts provided recommendations by email. We felt it was important to allow each expert to define program quality independently. Thus, we did not define high quality for them in our outreach email and instead asked for “recommendations of strong programs to interview.”

Programs in the study

The 20 experts who responded to our request recommended 105 programs that they considered strong. We selected 45 programs out of the 105 recommended to contact for interviews using videoconferencing software. To select this first set of programs, we prioritized four criteria: 1) programs recommended by multiple people, 2) program types that Wallace identified as priorities (afterschool, summer, online, programs that serve diverse populations or focus on equity), 3) program types for which we had fewer recommendations from our experts overall (e.g., online), and 4) programs that did not have recent, rigorous evaluations. If two programs shared these criteria, we prioritized the ones that would add geographic variation to the sample. When responses from this first set of programs slowed, we reached out to 13 programs that were either online programs or career readiness programs, as these were the two program types for which responses were still low (only one or two responses each). In the end, 28 programs participated in interviews, and an additional 10 provided information by survey only. We also received fully or mostly completed surveys from 12 of the 28 interviewed programs. Our final distribution of interviewed programs can be found in Table 2.
### Table 2. Number of Programs Interviewed by Category

<table>
<thead>
<tr>
<th>Focus Area</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arts</td>
<td>7</td>
</tr>
<tr>
<td>Career &amp; workforce</td>
<td>5</td>
</tr>
<tr>
<td>Civic engagement &amp; social justice</td>
<td>7</td>
</tr>
<tr>
<td>Social and emotional learning</td>
<td>5</td>
</tr>
<tr>
<td><strong>Programs with specific timing, location, or target population</strong></td>
<td></td>
</tr>
<tr>
<td>Online</td>
<td>3</td>
</tr>
<tr>
<td>Summer camps</td>
<td>3</td>
</tr>
<tr>
<td>Serves youth engaged in the juvenile justice system</td>
<td>5</td>
</tr>
<tr>
<td>Serves young school-aged children</td>
<td>11</td>
</tr>
</tbody>
</table>

**Note:** This table includes only the programs we interviewed, not the programs that completed surveys but were not interviewed. Also, the total here adds up to more than 28 because programs sometimes fell into multiple categories (e.g., social justice and arts). Some of the programs listed in the bottom half of the table (those with specific timing, location) overlap with the focus areas in the top half (e.g., online and social and emotional learning or juvenile justice and art).
Chapter 2: Outcome Measurements

Valued and measured outcomes

In this section, we discuss the outcomes on which programs focused. Specifically, by outcomes we mean the knowledge, attitudes, behaviors, or skills that programs want the young people they serve to develop. Many programs think about short-term outcomes (e.g., knowledge gained) and long-term outcomes (e.g., behavioral change) but measure only short-term outcomes.

Before summarizing the outcomes that programs reported measuring, we present definitions of key measurement-related terms used throughout the rest of this report, using “well-being” as an example of an intended outcomes:

Table 3. Definitions of Key Terms (Using “well-being” as an example)

<table>
<thead>
<tr>
<th>Domain</th>
<th>The broadest conceptual category of well-being. For youth-serving programs, domains may include educational well-being, social and emotional well-being, economic well-being, and physical health, among others. Within each domain, there may be several constructs (defined below).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construct</td>
<td>A conceptual category within the specific domain of well-being. For instance, a program may be interested in assessing educational constructs of academic performance, school engagement, or educational attainment or social constructs such as sense of belonging or collaboration.</td>
</tr>
<tr>
<td>Measure</td>
<td>The actual metric used to assess participant well-being. These may involve specific tools and methods for data collection and reporting, such as surveys, scales, observational codes, administrative records, etc. There may be several ways to measure a construct of interest, such as academic performance or sense of belonging, depending on the age group of the students in the program, among other factors. For younger children, a measure of belonging might entail a parent or teacher assessment. For older children, programs might use a self-report survey.</td>
</tr>
<tr>
<td>Output</td>
<td>Measure of program participation and services offered, such as the number of students served, the amount and type of sessions or supports provided, and the number and type of projects participants complete.</td>
</tr>
<tr>
<td>Outcome</td>
<td>Measure of the changes experienced by the target population as a result of participating in a program. Outcomes include changes in knowledge, attitudes, behaviors, and skills and may occur over a short or a long period.</td>
</tr>
<tr>
<td>Mediator</td>
<td>An intermediate output or outcome that influences whether a program achieves a targeted outcome. For example, frequency of participation in a program, duration of instruction, or program quality may serve as mediators for whether a program achieves a targeted outcome like improved social and emotional skills.</td>
</tr>
<tr>
<td>Moderator</td>
<td>A specific condition or characteristic of the target population that leads to variation in outcomes. For example, boys participating in a program may benefit more or less, or differently, than girls.</td>
</tr>
</tbody>
</table>
The term "measures" may refer to outputs, outcomes, mediators, and moderators. Each of these is important in informing programs’ efforts to reach their long-term goals and improve program quality.

Content-specific measures

The most easily identifiable measures were specific to the content of the program (e.g., arts, career and workforce readiness, civic engagement and social justice). See Table 4 for examples. Table 4 also includes commonly reported measures, such as social and emotional learning (SEL) and academics, that were not specific to a particular content area. Some programs identified other measures, such as physical activity, as a focus. Most programs spoke of these measures as either short- or long-term outcomes, though some could be thought of as mediators (e.g., improved math grades or test scores being a step toward graduation).

Table 4. Examples of Program-Specific Measures

<table>
<thead>
<tr>
<th>Focus Area</th>
<th>Examples of Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arts Programs</td>
<td>Completion of a performance; creative thinking skills; artistic ability (dance, music, mural painting, etc.); public speaking skills; self-expression; community engagement; knowledge of how to use the arts to engage others</td>
</tr>
<tr>
<td>Career &amp; Workforce Programs</td>
<td>Employment; earnings; knowledge of career options; completion of sample resumes or cover letters; positive reviews from supervisors; completion of an internship; credentials earned; job readiness skills (professional communication, timeliness, job-specific technical skills, etc.); development of a professional network</td>
</tr>
<tr>
<td>Civic Engagement &amp; Social Justice Programs</td>
<td>Knowledge about government systems and processes; positive attitude toward civic engagement; intention to participate in civic life (e.g., attend a city council meeting or email an elected representative); ability to perform root cause analyses (defined below); knowledge of political and social systems; ability to listen to other points of view</td>
</tr>
<tr>
<td>Online Programs (Note: Online programs in our sample were mostly STEM-focused.)</td>
<td>Technical skill mastery (robotics, computer science, gaming, e-sports, etc.); teamwork; age-appropriate math skills; digital citizenship (defined below)</td>
</tr>
<tr>
<td>SEL Outcomes</td>
<td>5Cs +1 (competence, confidence, connection, character, caring, and contribution); self-esteem; self-efficacy; cooperation; teamwork; empathy; problem solving; decision-making skills; resilience</td>
</tr>
<tr>
<td>Academic Outcomes</td>
<td>Graduation; on-time advancement to next grade; college enrollment; passion for learning; school engagement; ability to identify universal themes in literature and theater; English and math skills</td>
</tr>
<tr>
<td>Other Outcomes</td>
<td>Life skills, physical activity, reduced recidivism, equity-focused outcomes (see more below)</td>
</tr>
</tbody>
</table>

Below we describe two categories of measures—SEL and equity—in more detail.
Measurement of social and emotional learning (SEL)

Our interviews with OST programs clearly supported that SEL has been a growing priority in the OST field over the past decade. SEL outcomes (e.g., relationship-building skills, emotion management, communication skills, empathy) were a point of emphasis for the programs in our sample, regardless of program content. Arts, career readiness, and even online programs focused on video gaming incorporated SEL. For example, a theater program might stress the importance of teamwork, while a civics program might make a point of building students’ empathy. One interviewee noted that, while civics and SEL are closely linked, it is easier to measure—and get support for measuring—SEL:

“[SEL] is naturally embedded in what […] civic development goals are anyway, like developing critical thinking… From a measurement standpoint, the field of SEL is much more developed and has a lot more standardized [measures] we can pull from and also has much more interest and investment from funders.”

Some programs explicitly measured SEL outcomes by weaving specific SEL measures (either validated measures from other surveys or questions they developed on their own) into the other surveys or assessments they already conducted.

Several programs that focus on younger children did identify SEL as their primary content area. These programs often had curricula geared toward developing these skills, and measures were explicitly linked to those curricula. For example, Wings for Kids, an SEL program that works with elementary school-aged children in three southern states, focuses on and measures five core SEL competencies: self-awareness, self-management, responsible decision making, social awareness, and relationship skills. The program uses the mini-Devereux Student Strengths Assessment (DESSA mini) twice a year to measure these five competencies. (For a short time, they also collected weekly assessments of students; however, this was burdensome for staff and ultimately resulted in poor quality data.) The DESSA is a 72-question screen that takes 5-10 minutes to complete and measures eight SEL competencies, while the DESSA mini consists of eight questions and takes one minute to complete (see here for more information about both instruments). Wings for Kids also collected some qualitative data, primarily through interviews and focus groups, to better understand how program quality and implementation are related to outcomes. Multiple programs said they use a different measurement tool, Hello Insight, to collect data on SEL.

Measurement of diversity, equity, and inclusion

It was clear from our interviews that the rise of the racial justice movement in 2020 prompted some programs to think critically about their curricula and the outcomes they do or do not measure. Whether this reflection led to changes in measurement strategies varied from program to program. Many interviewees spoke of the need for programmatic changes related to diversity, equity, and inclusion (DEI) and said that those changes would help them determine what to measure. Following are three other themes about how programs measure DEI that arose from our interviews:

Disaggregating data by race/ethnicity

One way programs assess their progress on issues of equity is by determining whether participants from different backgrounds experience equitable outcomes. Specifically, by disaggregating data by race and ethnicity or other demographic variables (age, gender, neighborhood, socioeconomic status, etc.), programs can begin to understand whether their program is helping to reduce disparities, having no impact, or even exacerbating them. Disaggregating data is an analytic strategy, not a measurement strategy; however, it is related to measurement because it necessitates collecting data on race and ethnicity or other demographic variables of interest. Some programs did this well, while other programs noted that most participants were youth of color and thus disaggregating by race was not feasible. These programs sometimes identified other factors that they could disaggregate such as income, neighborhood, and length of time in the program. These
Factors could serve as proxies for risk or vulnerability in the same way that other programs sometimes used race. Many programs did not yet have a clear process for responding to lingering or worsening disparities they identified.

**Tracking development of equity-related knowledge and skills**

Some programs, particularly arts, civic engagement, juvenile justice, and camp or recreation programs, have the development of civic knowledge or social justice skills among their intended school outcomes. These skills involve learning how to discuss racism, power, inequity, and their personal experiences with these complex social dynamics and, in some cases, engaging with members of the community to address the effects of racism. The programs sometimes assess these outcomes using traditional methods such as surveys but more often using nontraditional methods like discussions, the art students produce, or the local action campaigns they carry out. In addition, equity is often embedded in the content of programs that do not make it an explicit focus of their work, for example by choosing equity-related topics for a theater production, facilitated group sessions, or local campaign work. Just as a program that does not explicitly focus on SEL might track SEL outcomes informally, programs that do not explicitly focus on equity might ask staff members if they feel there are any equity issues to discuss or whether disparities appear to be arising.

**Revising curricula and/or outcomes measures based on equity reviews or perspectives**

In redesigning their curricula in 2020 and 2021 to better address racial equity issues, some programs more clearly described how they would track students’ knowledge of these issues. For example, a civics education program spoke about how an exercise called root cause analysis, which encourages students to think backwards from an issue to its underlying causes, had always been a key part of its curriculum but would now explicitly include discussion of inequities in power and resources. For example, students might identify differences in high school graduation rates for different groups of people and then work out an explanation for this outcome. Teachers would guide those conversations so that power, inequity, resource allocation, and representation were explicitly included in their analyses. An interviewee described the new approach:

“We’ve been working with some DEI consultants, and we have some great people on staff also who are fully overhauling the curriculum… [so] a lot of the civics knowledge aspects and the civics skills aspects are more clearly grounded in students’ learning around what systems of oppression are, how they function, how their identity and intersectional identity shape their experience in classes and communities. And then, hopefully through that, supporting a much more grounded root cause analysis and policy-action approach for our youth.”

For this program, increased knowledge of how power and resources are used, and subsequent behavior change are long-term outcomes; whether discussion of these issues takes place is a mediator. In other words, the program’s assumption is that students’ knowledge about equity-related issues will increase if discussion of those issues is part of their root cause analyses.

**Measurement tools and techniques**

Programs reported that they use a variety of traditional and nontraditional methods to measure targeted program outcomes with varying degrees of success. Traditional methods of data collection include observations, surveys and questionnaires, interviews, focus groups, formal assessments, observation of participant performances, self-assessments, tracking attendance rates, and collecting school data. Nontraditional methods, including check-ins and reviewing participant journals and portfolios, often blend more seamlessly into programming than traditional methods because they are generally designed to be instructional tools first, evaluation tools second. In fact, while many programs have used check-ins, journaling, and portfolios for years, in some cases for quality improvement purposes, some still do not recognize that these methods can be used to measure student outcomes. Many stated that they think outcomes measurement needs to be more formal or quantitative.
Traditional methods

Surveys
Almost all programs reported that they use surveys to collect participant data. Programs used a variety of surveys, including validated tools (i.e., tools that have been developed by experts and shown to be effective in the field), and more informal surveys that they developed themselves. (Programs that reported developing their own survey explained that they did so because other surveys did not ask the questions they needed.) While most programs administered surveys at the end of the program or sporadically during the program, many administered a pre- and a post-survey at the beginning and end of the program.

For online programs, in which participants never meet at the same time, it may not be practical to administer a survey at a specific point in the program to the entire population. One online program used quick randomized pop-up surveys that participants, and often staff, completed during their online sessions. A few programs also administered a survey to parents.

Programs also used tools validated by universities and other organizations for participants to conduct self-assessments. These tools varied by program and, in some cases, changed from year to year.

Formal assessments of knowledge and skills
Some programs, including programs focused on STEM, SEL, and juvenile justice, used formal assessments to measure not only what and how well the participants learned but also their proficiency in applying what they learned. These formal assessments varied from program to program and included short quizzes, skills tests, and questionnaires.

Fine arts programs that focused on poetry, music, literature, and dance typically used performances at the end of the program—such as recitals or choreographed dances—to measure whether and how well participants learned relevant skills. In one theater program in Chicago, participants worked together to write their own play drawing on stories from their peers, communities, and their own lives. The final performance demonstrated the extent to which the participants had developed storytelling and other skills.

Observations
Just fewer than half of the programs in our study reported using observations by staff members to measure participant outcomes. In many cases, this method of measuring outcomes entailed observing participants as they worked on projects individually or in teams to gauge their progress toward a specific goal (e.g., developing interview skills) and including this information in reports throughout the program or at the end of the program. As noted above, arts programs in our study commonly observed performances to measure participants’ progress in developing particular skills.

Interviews and focus groups
Several programs conducted interviews and focus groups to collect information from participants, staff, and parents that they then used for quality improvement purposes and for measuring youth outcomes. Some programs used interviews and focus groups to learn why youth might be struggling or to better understand how to support them. Additionally, a few programs, mostly arts programs, used interviews and focus groups to gather participants’ own perceptions of what they had learned.

School data
Many OST programs focus on academic skills. Therefore, almost all the programs we spoke to wanted access to school data—administrative records, student grades, or teacher reports—that they could use for quality improvement and evaluation purposes. Very few, however, actually had access to this data from the schools or school district because it tends to be carefully protected and rarely shared. Programs that did have data use agreements with schools or school districts were able to use that data to track participants’ progress in
school over the course of the program. A few programs that did not have such agreements asked students or parents to self-report grades or test scores.

**Information programs tracked beyond outcomes**

Many programs noted that they collected and measured information other than participant outcomes. For example, many programs reported tracking attendance, an output that is usually considered a mediator of behavioral outcomes like skill development, knowledge, or behavior change. As with other program outputs, attendance is easier to track than outcomes like increases in knowledge or changes in behavior. Some programs used attendance as a proxy for engagement (i.e., the degree of young people’s participation in the program) and as part of their reporting to funders.

Many programs also reported measuring program quality using assessments such as the Weikart Youth Program Assessment (YPQA) or the After-School Quality (ASQ) Observation Tool.

**Nontraditional methods**

**Informal measurement of SEL outcomes**

Programs, particularly summer camps, found creative ways to track SEL outcomes. One was to use games or an award system. For example, **Fiver**, a summer camp and school-year program, developed a system of “dog tags” that youth get for exemplifying specific competencies such as trustworthiness. The youth wear the dog tags proudly, even when they return home from camp or go to college. Giving awards for key program outcomes made it possible for Fiver to track participants’ progress over the course of the summer without additional assessments or surveys and incentivized youth to make further gains. The program could also track the acquisition of dog tags on aggregate year over year, but it was unclear whether it did so.

Online programs also informally tracked SEL outcomes that were important to them. Staff and coaches at **Connected Camps**, an online program that partners with schools to support students’ participation in an e-sports league, met regularly to discuss students’ progress toward SEL outcomes like teamwork, sportsmanship, commitment to racial and gender equity, sense of community, and digital citizenship (i.e., treating people respectfully online) and to identify ways to help students struggling in these areas. One staff member commented: “We have a big emphasis on healthy gaming, reducing toxicity...conflict solutions, collaboration...a lot of those ‘soft skills’ within the digital citizenship and SEL categories.” Connected Camps staff added that these skills are essential to STEM in particular and career readiness in general. They explained that many of the young people they serve were unable to represent their schools on a sports team, so e-sports offered a vital opportunity for building SEL skills.

**Regular check-ins**

Many programs reported using one-on-one or group check-ins with youth throughout the program to ask them how they were doing. Some programs scheduled check-ins, while others engaged with their participants informally. Many programs also used staff meetings or professional development sessions as informal opportunities for staff to assess progress toward SEL- and equity-related goals.

**Journaling**

The two programs that reported using journaling as a method of collecting data said participants were instructed to respond to prompts given by staff or to free-write. Staff members reviewed the journals throughout the program to gauge participants’ learning.

**Portfolios**

Several programs, including arts, career and workforce readiness, and juvenile justice programs, used portfolios collecting participants’ artwork, job records, or other relevant materials to measure outcomes.
Chapter 3: Factors Programs Consider When Deciding What Outcomes to Measure

This chapter focuses on how programs decide what outcomes they want to measure. These decisions are driven not only by the programs' theories of change (i.e., their understanding of how their work influences participants' knowledge acquisition, skill development, or behavior change) but also practical considerations like the cost of data collection. Program staff reported several essential criteria for deciding what outcomes to measure:

- Does the outcome fit into the program's logic model or theory of change?
- Is the data usable?
- Is it shareable?
- Does it incorporate youth voice?
- How much effort and capacity are required to collect the data?
- Are available measures valid and reliable?
- Does measuring the outcome promote equity?

Logic models or theories of change

Many programs reported having logic models or theories of change that influence what outcomes they are interested in measuring. Many logic models and theories of change also include intermediate outcomes (mediators) or factors that might explain variations in outcomes for different groups of participants (moderators). These mediators or moderators can often help programs understand the context for the longer-term outcomes for which they want to be held accountable. For example, the degree of participants' attendance in and engagement with a program might play a role in whether they develop the skills that are the program's focus. In a workforce program, mediators like certification or skill development might lead to a job.

Many programs that had a logic model or theory of change noted that the model or theory was developed early in the history of the program or at some other key strategic juncture. Staff may have used it initially to determine what outcomes to measure, but it no longer served that purpose. Nor did staff appear to use it to guide program implementation or data collection for ongoing quality improvement. They were more likely to use it in grant applications or to explain the program to funders.

Usability

Program leaders want to prioritize data that will inform how the program is run, how it is performing, which youth are benefitting and which are not, and what might make it better. One program leader put it this way:

“Before you invest in the outcome, know what you’re measuring. [...] If data comes back and says, ‘no improvements in caring,’ what are [program staff] going to do about that?...How are you truly using data for improvement? [...] If you’re not using it to showcase impact and you’re not using it for program improvement and it’s just sitting in the corner of your office, it’s not that helpful.”

For some programs, determining what data is usable is a process of trial and error that involves collecting one set of data, discovering that it raises new questions, and then adjusting accordingly. One program described the process as a “journey” in which they have developed a better understanding of how much
money and time it takes to collect particular pieces of data, how to explain the importance of data to staff, and how to tweak what they are measuring so that staff are able to act on the information.

Another organization spoke of the importance of collecting data that is valuable not only to program staff but also participants and avoiding data collection that comes off as critical of youth:

“Evaluation is also instructional, so we try to use student-friendly language so that students who do take the time to digest [the question/s] get some guidance on things that they could be working on. [...] Those are framed developmentally, so that students don’t feel we are judging them. [...] We try to focus, too, on behaviors and attitudes that are malleable, because we want students to feel like they are growing.”

For many programs, data collection can be a source of frustration for staff—either because the data they collect isn’t usable or the logistics of collecting it are burdensome. Staff with Our Piece of the Pie, a multi-service program that provides career readiness services, noted that every contract they manage comes with its own reporting requirements. On their wishlist was a system that would allow them to enter data using their phones while they are already working with youth so they wouldn’t have to schedule a separate time for data collection and entry.

Shareability with participants and community

Many program staff also spoke of the importance of sharing data with youth, parents, schools, and funders. Data sharing—with attention to privacy requirements—can serve a number of purposes, including engaging families and the community and helping stakeholders and program participants track their own progress. The New York City Administration for Children’s Services (ACS NYC), an agency that serves youth involved in the city’s juvenile justice system collects portfolios of students’ work while they are in detention. It shares these portfolios with the students’ home school and the transition specialist. This builds trust between ACS NYC, the transition specialists, and the school system and helps the students establish themselves at their home schools.

Another primary purpose of data sharing is to communicate program impact and effectiveness, which can help with fundraising. Staff at Mikva Challenge spoke about wanting to tell the stories of their youth participants:

“We really want to show that young people are being transformed by our programming. That they are becoming more civically engaged. Not only when they are part of our program but carrying that on into adulthood. [...] We want to be storytellers because our stories are amazing.”

Program staff felt that stories of this kind were more compelling (and better illustrated young people’s agency in achieving desired outcomes) than the time- and resource-intensive quantitative reports that some funders require.

Youth voice

Incorporating youth voice into selecting outcomes and data collection methods was a priority for many programs in our study. In general, arts and juvenile justice-focused programs did a good job collecting youth feedback by conducting regular check-ins to 1) make sure participants were getting what they needed from the program; 2) build participants’ communication and other SEL skills; 3) regularly gauge participants’ progress; and 4) engage participants in making meaning of data or assessing data collection strategies.

Programs mentioned that soliciting youth voice also means learning about participants’ personal goals and development. For example, a young person in a career readiness program may be interested in a job in a
specific field. Staff in arts, juvenile justice, and workforce programs emphasized that working with youth to identify and take steps toward a goal that is meaningful to them is, in many cases, a more important outcome than the goal itself. In this sense, youth voice is not simply a process to get to an outcome but an outcome itself. Data collection in this instance is about recording participants’ goals and progress toward those goals, typically using a database or customer relationship management system. Staff at the Our Piece of the Pie career readiness program explained how the process works:

“[Participants] can have a short-term goal—[which] they have to complete every thirty days—an intermediate goal—[which] might be three to six months—and then a yearly goal. So, we are basically tracking what they are saying they want to achieve and how they are progressing towards that...[W]hen you go and set that goal, you can put action steps within that and set dates you hope they hit, and you can check them off as you go.”

**Capacity and effort required**

Most organizations do not have staff dedicated to data collection and analysis. This responsibility is usually one of many and consequently can end up as a lower priority. One organization mentioned this directly: “I wrote yesterday to the Board about our need to hire a fulltime Director of Impact... There’s so many different things...that we wish we were measuring, and we just don’t have the capacity for it.”

In some cases, lack of capacity can force programs to make hard choices about what to measure. As one program put it:

“Our problem is we measure a lot—a lot a lot—and we need to focus a bit more...We’re trying to narrow down the five core outcomes we want to see...[We need] staffing, staffing to do the research, staffing to figure it out, staffing to implement it, staffing to analyze it. You know, we sometimes have to say, we don’t want to do this survey because we’re not going to be able to analyze it for another year, and that’s not helpful.”

In the absence of dedicated personnel and resources, data collection can be a significant strain on staff, participants, other stakeholders from whom the program collects data, and program budgets. One program called this “survey fatigue,” but the issue is much broader than that label suggests. Some programs reported that staff will stop collecting data or even quit their jobs if the process becomes too burdensome or alienates youth. Most staff members who work directly with youth got into the field because they care about them deeply. Being forced to choose between spending their time engaging with young people in fun, productive ways or collecting data (particularly data they do not see being used) can lead to dissatisfaction and burnout. One career readiness program told us:

“I’ve had staff leave because they just got frustrated with the data, the stuff that was forced on them. And I’ve had staff say, ‘I feel rude when I’m talking to youth and I say, ‘I’m going to type this up as we’re talking,’ because they feel like if they don’t get it into the system [they’ll] forget to go back or won’t get accurate [data]. And they don’t like working like that because these people generally are people-persons who want to work with individuals, and they want to give them their attention.”

Program participants can also experience burnout because of excessive data collection. This burnout can compromise the quality of the data if participants are not willing to cooperate in the process. An even greater concern is that young people, particularly those who have been marginalized at school or part of the juvenile justice system, may be re-traumatized by intrusive data collection. One program leader said, “It can be a lot. We’ve been working with the program staff to balance what is needed versus what is not needed, just so we don’t get [youth] burnout. We’ve really worked to, [in] the last couple of years, scale back where we can or consolidate where we can.”
Program leaders must try to strike the right balance between protecting valuable staff time and program dollars on one hand and ensuring that the program has usable, high-quality data to work with on the other. As with identifying usable data, this balancing act can be a continuous process with programs trying one approach and then adapting it to work better for everyone involved. A staff member at Wings for Kids spoke of the adjustments the program had made in its approach to data collection:

“We tried to do it much more regularly, and it just is overwhelming. You know, we have staff trying to do a Devereux Student Strengths Assessment (DESSA) on 12 kids, and we found that when they were doing it more often, they were just kind of answering and not really putting real thought into the actual assessment of the kids.”

Validity and reliability of measures

In order to effectively collect and use data, programs need access to measures that are valid (i.e., accurate) and reliable (i.e., yield consistent results). This can be challenging for two reasons. First, proprietary measures developed by researchers can be prohibitively expensive for many programs. Second, researchers don’t always report on the validity or reliability of the measures they develop, and many staff members don’t have the training to assess the quality of a given measure on their own.

Mikva Challenge, a civic engagement program in Chicago and Washington D.C., uses measures developed by the Center for Information and Research on Civic Learning and Engagement (CIRCLE) at Tufts University because they have undergone statistical analysis for validity. These measures are not publicly available, but Mikva Challenge has been able to use them because CIRCLE is interested in helping civic engagement organizations determine what to measure and how to measure it and in testing those measures. Organizations that are not well-connected enough to form a partnership of this kind may struggle to find high-quality measures.

Equity

As programs have become more interested in understanding whether they have disparate effects on different groups of participants, they have put more thought into how to identify these disparities and their causes. Many of the programs in the study did this by breaking down the data they collected by race and other demographics. Some civic engagement and arts programs developed outcomes specifically related to equity, such as increased knowledge of how power and inequity operate in society.

Because equity continues to be a topic of rising importance, programs’ thinking about it is changing from year to year. It will be important to revisit the question of how OST programs measure equity in 3-4 years as they begin to settle on and codify their approaches.
Chapter 4: Measurement Challenges

What measurement challenges did OST programs experience?

The programs we spoke with noted several measurement challenges. Some were specific to the organization, while others were driven by external factors, such as COVID-19. The measurement challenges programs reported fell into three categories: 1) lack of available or tested measurement tools; 2) difficulty measuring specific types of outcomes; and 3) the burden associated with data collection.

Lack of available or tested measurement tools

Few OST programs reported that they lacked the tools they needed to measure the outcomes of interest to them. The ones that did report a lack of tools identified cultural competence, appreciation for diversity, and changes in anti-racist practices as outcomes they were unable to measure with the tools available. An arts program said, "[we would love to measure our equity mindset, [and] how are we developing it." Another program said it would "love to measure [the] impact of celebrating and appreciating youth’s culture in programs.”

Some programs were unaware of research-backed surveys and assessments designed to track the development of SEL skills and related outcomes, such as creativity, identity formation, sense of belonging, self-concept, and confidence. One program said it was interested in "soft skills that can’t be measured in a meaningful way," while another said of this same category of skills, "not sure where [the measurements] are, [but] we are sure they are out there." It was not within the scope of our current study to review the tools available for measuring particular outcomes, the tools used by programs, or the quality of those tools. We did learn that many programs used untested tools developed by their own staff.

Specific outcomes that are difficult to measure

Long-term outcomes

Many programs expressed interest in tracking outcomes one or more years after participants left the program. For instance, career and workforce development programs were ultimately interested in understanding if the training, job readiness skills, and work experience participants gained through the program led to better outcomes in terms of future employment, earnings, and benefits. STEM programs mentioned wanting to know if program participation resulted in higher rates of high school graduation and college matriculation, influenced college major and/or career choices, and built long-term trust in scientific principles. For civics programs, long-term outcomes of interest included voting or involvement in a political or advocacy campaign. SEL programs had questions about how developing social and emotional skills affects academic achievement; career development; or relationships with friends, families, or partners. Arts programs were interested in whether their participants were more likely to find lasting happiness, achieve their goals, and participate in civic life.

Programs looking to collect this kind of longitudinal data had trouble keeping track of participants after they left the program, lack of access to administrative records, lack of consent to collect long-term data, and limited staff capacity to do the work. One program described the challenge this way:

“If we’re funding this program, what are the long-term [benefits]? Not just, ‘is this child gaining credits toward graduation this semester for youth in our afterschool program, […] but what does success look like three months, six months, nine months down the road?’ And we’re not doing that right now because we don’t have a system for getting that information from the school district or those families after the kid has graduated from our program. […] We need more capacity to maintain that contact longer-term and actually see real outcomes.”
Another said, “We use [the] National Student Clearinghouse to report on college enrollment [and] persistence.” This program was interested in knowing if its participants were employed and, if so, “where are they employed, how much are they making, what the job is. That would be the dream. [We have] explored different avenues such as Equifax work numbers...That is hard to get.”

Behavior versus knowledge

Many programs reported measuring participants’ knowledge attainment but were unable to determine whether and how knowledge affected behavior. For example, a civics program might track what participants learned about voting but not whether they actually voted. An SEL program might measure the development of participants’ communication skills but have no sense of whether they struggle to put those skills into practice in their daily lives.

School outcomes

While some programs that focused on academic achievement conceived of themselves as extracurricular and therefore were not interested in tracking participants’ progress in school, most wanted access to school data. Many programs did not have data-sharing agreements with school districts and other public agencies. As one put it, “[We] have a plan. [It’s] just sitting on someone’s desk waiting for signature.” In the absence of formal agreements, programs had to gather the information from the student themselves. Outcomes like promotion to the next grade and high school graduation were relatively easy to measure in this way because programs could ask all participants about them at the same time. Test scores, grades, and behavior were more of a challenge: Participants did not always want to share the information or did not share it in a timely way, making it difficult for programs to act on it effectively. For example, if a program knew that a student failed an algebra test as soon as it happened, it could provide more tutoring or guidance on study habits before that student failed the entire course. In many cases, however, programs missed that opportunity. Many programs said they needed help engaging and forming trusting partnerships with schools and districts, and developing formal data sharing-agreements to meet their measurement objectives.

Relationships between staff and youth

Many of the programs we spoke to wanted to understand whether and how improvements in a staff member’s skills or knowledge were associated with improved youth outcomes. For example, in one civics program, the focus on democratic classrooms—in which students guided discussions and chose topics to focus on—required understanding on the part of the teachers. Analyzing the relationship between staff and youth learning was a challenge because it required high-quality measurement of both and an understanding of statistical modeling.

Data collection challenges

As noted in Chapter 3, data collection can be a burden on staff, who in many cases are not trained to do the work, and program participants. Some programs in our study were reluctant to ask students to fill out more forms given the heavy testing culture they experience in school. They were also concerned that data collection activities took staff away from delivering program content.

Multiple interviewees noted that their program already collected more data than it could use, saying staff lacked the time for data entry and the expertise to manage the collection process, analyze the data, and create the kind of visual presentations that make it usable and shareable. Some pointed to the burden of fulfilling the reporting requirements of various funders, which often involved recording duplicative information using multiple tools and databases.

We spoke to programs in early 2021 when some were operating online due to COVID-19 and figuring out what data they could collect under the circumstances and how they could collect it. One noted an interest in
“capturing things being said on social media about youth’s experiences with program successes.” Another spoke of trying to “do discourse analysis of students’ writing and online chat artifacts.” Barriers to online data collection included participants’ lack of stable internet access and privacy protections and concerns related to tracking information online.

Other data collection challenges noted less frequently include the following:

- Issues of parental consent
- A need to incentivize data collection
- Concerns about social desirability bias (i.e., participants giving the answers they think staff want to hear)
- Lack of funding
- Difficulty gathering information from a representative sample of the participant population
- Poor response rates
- Lack of “kid-friendly” data collection tools
- The ceiling effect (i.e., most participants report positive outcomes at the start of the program making it difficult to measure progress)
- Difficulty tracking outcomes for participants who are in the program for a short period of time

Finally, some programs that do have funding to hire dedicated staff for data collection activities noted that turnover can be a challenge. Data teams can be as small as one person, meaning turnover can bring data collection to a grinding halt.

Measuring equity outcomes can be challenging.

Programs that measure equity outcomes (such as eliminating disparities between demographic groups or increasing participants’ knowledge about how power changes social outcomes) often do so in informal ways, such as staff meetings, staff check-ins, and discussions with participants. These informal methods can make it challenging to quantify progress. Some programs, such as Generation Citizen, are working on updating their logic models to more clearly explain how they will know whether they are achieving their equity goals, but for many this work is new and challenging.

Other programs expressed an interest in accessing data that would help them gauge the effectiveness of their approaches to equity. For instance, Sherwood Forest, a summer camp in which young people participate from 1st grade through high school, was interested in knowing if its outreach and retention efforts have yielded equitable results—“[Are the] graduating class demographics the same as when they entered? Did [we] lose kids who are worse off? Is the class whiter? This isn’t easy to track.”
Programs can benefit from tapping outside expertise in data collection and use.

For programs looking to, as one put it, "change the culture around data...to be less fearful of data and more data-driven," bringing in outside experts to help facilitate data collection efforts can make a big difference. One program, Project Morry, discussed being part of a fellowship with Youth INC., a New York City-based, nonprofit organization that helps programs plan for and carry out data collection and analysis. Youth INC. focuses specifically on SEL outcomes but also helps organizations think broadly about data. Project Morry staff told us:

"...it's a competitive program and an application process. And what they do is help nonprofits with creating systems for measurement and getting yourself set up to really do this on an on-going basis...So, we've been working with them for the last two years and through them we have been trying to really focus our organizational culture on measurement and program evaluation... [SEL outcomes were] something that was always intrinsic to what we were trying to accomplish. [Youth INC.] helped us formalize what we were always trying to do and gave us a way to really measure it. ... We're still on the journey in terms of...understanding what to do with the data once we get it and how to utilize it to actually make programmatic decisions. But we are moving in that direction. We are surveying our kids. We use Hello Insight, which is an online tool which has made it really easy to do surveys twice a year and get the data and the analytics."
Chapter 5: Key Findings and Recommendations

The programs in this study made it clear that they cared deeply about measuring participant outcomes and program quality and using what they learned to make improvements and achieve their missions. While many of the staff members we interviewed got into the OST field to work with young people, not data, their desire to help young people develop new skills and achieve success and well-being motivated them to learn a lot about how to collect and use data. This bodes well for the continued commitment of OST programs to measurement and evaluation.

At the same time, the burdens of, and obstacles to, data collection and use that they described were wide-ranging and daunting relative to the resources available to them. Programs need more time, financial support, training, and dedicated staff to ensure that the data they collect is high-quality and that they can act on it effectively.

Summary of findings

The following key findings address the three research questions that were the starting point for the study:

Research question 1: What is the range of intended outcomes for children and youth attending OST programs, including, but not limited to, programs that focus on the arts, civic engagement and social justice, career and workforce development, and general or other services?

- Programs in the study measured outcomes that were closely related to the content they delivered (e.g., arts, career and workforce readiness, civic engagement and social justice). They also generally measured SEL outcomes and other outcomes required by funders.
- Programs thought critically about how to measure equity outcomes. Approaches to measuring equity included disaggregating data by race and other demographic variables and tracking the development of equity-related knowledge and skills.
- Programs used several criteria to decide what outcomes to track, including consistency with the program’s logic model or theory of change, usability and shareability of the data, the effort and capacity required to collect the data, availability of valid and reliable measures, youth interest, and whether measuring a given outcome would promote equity.
- In addition to outcomes, programs consistently measured outputs (i.e., steps that lead to desired outcomes), particularly participation and program quality. This is consistent with findings from our literature scan that public and private funders have invested heavily in quality assessment tools and participation tracking systems. Comparatively few programs used qualitative methods to understand the factors that contribute to program participation and quality.

Research question 2: What are the different measurement approaches and instruments used to document OST program outcomes, including formative and summative measures?

- Programs reported using a variety of traditional methods to measure outcomes, including quantitative methods—like administering surveys and questionnaires, conducting formal assessments, and tracking attendance rates—and qualitative methods like conducting interviews and focus groups with young people. Their use of quantitative methods was well-documented, use of qualitative methods less so.
- Programs also used nontraditional methods, such as regular check-ins with participants and reviewing participant journals and portfolios. As with their use of qualitative measures, use of these nontraditional methods was not well-documented.
- Programs found creative, informal ways of tracking SEL outcomes, including games and award systems.
Research question 3: What are the gaps in and barriers to documentation and measurement where intended outcomes are either not measured or are not articulated because of a lack of a measurement strategy, skills, resources, or instruments?

- The programs we interviewed did not consistently identify specific outcomes of interest that they were unable to measure, although some said they lacked the tools they needed to track equity- or SEL-related outcomes or program quality.
- Programs pointed to broader types of outcomes that were challenging to measure, including longer-term outcomes like college matriculation, career attainment, and participation in civic life; behavior change (e.g., do participants in a civics program exercise their right to vote); school outcomes like test scores and grades that require a data-sharing agreement with the district; and the relationship between improvements in staff members skills and knowledge and youth outcomes.
- Programs reported that the process of collecting data could be burdensome for participants, who have their fill of testing at school, and staff, who in many cases do not have the training or time to do the work. Some pointed specifically to the burden of fulfilling the reporting requirements of various funders, which often involved recording duplicative information using multiple tools and databases.

Recommendations

The following recommendations are divided into three main categories: recommendations for OST programs, recommendations for funders of OST programs, and recommendations for further research. All three sets of recommendations address the challenges experienced by the programs in the study and, in some cases, were suggested by the interviewees themselves.

Recommendations for OST programs

These recommendations for OST programs call for a shift in organizational priorities and approach as opposed to additional funding. In some cases, though, programs may benefit from technical assistance to help them successfully adopt these ideas.

Select a small number of high-priority outcomes to measure.

Programs often cast a wide net for data and end up not using what they collect, which can be frustrating for staff and participants. To avoid this, programs should focus on data they will be able to use and share with participants, families, and the broader community. They may also want to identify outcomes they consider “exploratory”—that is, outcomes they value but choose not to measure for practical or other reasons.

Ensure that the selected outcomes fit into the program’s logic model or theory of change.

Many programs reported that they did not use their logic model or theory of change to decide what outcomes to measure or guide their data collection efforts. Programs should review the outcomes they target for consistency with their logic model or theory of change and, if necessary, make changes to one or the other.

Document the type, source, and purpose of each piece of data collected.

The programs in our study were more likely to keep a record of the data they collected using traditional, quantitative methods like surveys or formal assessments than data collected using qualitative or nontraditional methods like check-ins with participants or portfolios of their work. Programs should thoroughly document all the different ways they collect data to ensure that all data gets used and to better understand what each method contributes to their work.
Make room for youth perspectives in deciding what outcomes to measure and how.

Ultimately, OST programs collect data to better serve young people. Therefore, it is important that programs find out what young people want to gain from their participation (e.g., to learn a particular skill or to get a job in a particular field) and their progress toward those goals. Programs can solicit input from participants not only in deciding what outcomes to measure but also in assessing how well the data collection process is working and in making meaning of what’s collected.

Incorporate measures of diversity, equity, and inclusion into data collection efforts.

Programs interested in tracking their progress on issues of equity should consider disaggregating data by race or other demographic variables to identify disparities in participant outcomes; incorporating outcomes specifically related to equity (e.g., the development of civic knowledge and social justice skills) into their logic models or theories of change; revising their curricula to promote those outcomes; and selecting qualitative and, if possible and appropriate, quantitative methods of measuring those outcomes.

Assess the capacity and effort required to collect data and take steps to minimize burnout.

Programs should expect finding the right approach to data collection for staff and participants to be an ongoing process of trial and error. In addition to selecting a small number of high-priority outcomes (as recommended above), programs can help reduce the burden on staff by scheduling data collection for a few specific points in the calendar when data will be most relevant rather than trying to collect it continuously over the course of the program.

Recommendations for funders

To truly get the most out of data collection efforts, OST programs need outside support in a number of forms: additional funding, training and technical assistance, access to high-quality tools, and opportunities to work with researchers and learn from peers. The following recommendations touch on the ways funders can partner with programs on their data journeys.

Connect programs to additional training in key aspects of data collection and use.

Many staff members who oversee their programs’ data efforts are trained in youth development and are learning to work with data on the job. Interviewees spoke of the need for more training in assessing which measurement tools are valid and reliable; in developing new databases to replace systems they describe as “clunky” and “cumbersome” or better understand how to navigate the systems they have; and in using data to support program improvement.

Create networks of programs to provide informal technical assistance.

One program in our study secured a grant to create a network of OST programs that could support and learn from one another. Through this network, the program has learned how various types of programs have solved common problems and has worked with other programs’ technical assistance specialists to troubleshoot its own particular challenges. This program noted that it would also benefit from participating in networks focused on developing specific skills (e.g., collecting longitudinal data, identifying outcomes of interest and ways to measure them). Conferences serve this function to some extent, but many programs cannot send multiple staff members to conferences due to travel costs or the time it would take away from their regular responsibilities. Virtual meetings or webinars that give participants the opportunity to engage with one another on specific topics of interest are one alternative to in-person conferences.
Provide funding for additional staffing.

Tight budgets mean that a) programs are often unable to hire dedicated staff for data collection and analysis and b) the staff they do have are too busy providing all the services youth in the program need to give data work the attention it demands. Additional resources would allow programs to hire staff whose primary responsibility is data or to reduce staff-to-participant ratios so all relevant staff members have more time for data activities.

Facilitate partnerships with researchers.

Partnerships between practitioners and researchers can be mutually beneficial. Practitioners need guidance on what to measure and how to measure it, while researchers need to field-test the measurement tools they develop or real-world data to analyze. Mivka’s Challenge for example has been able to use measurement tools developed by Tufts University, and Tufts researchers have been able to see how those tools work in practice. Research partnerships may also involve formal initiatives in which researchers provide program staff with technical assistance in identifying what outcomes to measure, selecting valid and reliable measurement tools, and analyzing the data they collect.

Make reporting requirements as flexible as possible.

Programs sometimes see their own goals for data collection as being at odds with the data reporting that funders require of them. Almost every program in our study mentioned at least once that meeting the requirements of their various funders was a time- and resource-intensive process that involved recording information that was duplicative or not meaningful to them using multiple tools and databases. Funders should be sensitive to the strain that data collection can place on programs and be as flexible as possible in terms of how and in what format grantees demonstrate progress toward outcomes of interest. As one interviewee put it, “If you hired someone to do a job at your house, to build a deck, and then you said, ‘Okay, I’ll hire you because I like you and the work you do, but I want you to use only my tools,’ it would be kind of ridiculous to go about it that way.”

Support digital data collection.

As many programs moved online during the COVID-19 pandemic so did their data collection efforts. Even as they return to serving young people in person, programs could benefit from additional funding for, and training in, developing mobile data collection processes (Cava Tadik et al., 2019). Staff at one program noted that the ability to enter data on their phones while they are already working with youth would eliminate the need to schedule a separate time for data collection and entry.

Recommendations for further research

As is often the case with research, this study raised additional questions the answers to which would help inform the field. These recommendations are aimed not only at researchers, but also at funders since the recommendations require resources to carry out.

Expand the existing study to include more programs and voices.

The current study could be expanded by including a) more programs in each of the categories we prioritized; b) programs in additional categories (e.g., STEM, physical fitness, identity development) to increase the generalizability of the findings; and c) perspectives from different types of people involved with OST programs, particularly youth participants.
Further explore how programs make decisions about their data efforts.

Data collection and use require program leaders to make difficult tradeoffs in terms of money and staff time. To help programs navigate these tricky decisions, future research could focus on the following questions:

- How do programs identify outcomes meaningful to them and prioritize those outcomes?
- How do they choose the most appropriate measurement tools for their needs?
- How do they balance spending on improved data collection with spending on programming?
- What types of support for data collection and use (e.g., training, financial, staffing) do they find most helpful?

Highlight innovative solutions to common problems.

This study focused in part on challenges OST programs face in collecting and using data. Future research could help programs address these challenges by identifying "bright spots"—innovative solutions to common problems from exemplary programs. Topics could include forming successful data-sharing partnerships, minimizing the burden of data collection on staff, building staff capacity to collect and analyze data, and securing funding for data efforts.

Investigate strategies for collecting longitudinal data.

Programs in our study expressed interest in tracking long-term participant outcomes but had difficulty doing so. Many programs specifically mentioned the challenge of collecting data from alumni who may no longer be connected to the program. Future research could identify strategies for effectively measuring long-term outcomes by conducting a formal review of best practices in the field and gathering qualitative evidence from programs that have done it successfully.

Catalog existing measurement tools and develop and test new tools as needed.

Through our literature scan, we identified a robust set of tools available for measuring program quality and social and emotional learning, yet some of the programs in our study were unaware of these tools. Researchers could help address this disconnect by cataloging the tools already in use in the field (those developed by researchers and other outside experts and those developed by programs themselves) and assessing their validity and reliability.

The literature scan also revealed a lack of tools relevant to other types of outcomes, including content-specific and equity-related outcomes. Researchers could develop new tools to fill this gap. Funders could support this work by putting out a call for research in field publications like *The Journal for Youth Development*; organizing panels or conference sessions focused on this topic; or forming a network of practitioners, funders, researchers, and content experts to identify and share promising approaches. Funders could also spearhead dissemination efforts to get new and existing tools into the hands of programs that stand to benefit from them. New and existing tools should be tested to ensure they are appropriate for the diverse populations OST programs serve.
Appendices

Appendix I. Additional References


Appendix II. Bibliography for OST Measurement Literature Scan

Bibliography for OST Measurement Literature Scan


Franklin, C. (2017). Effects of the Afterschool Program on Student Achievement of Students with Disabilities in a Rural Georgia Middle School. In *ProQuest LLC*.


Goff, J. E. (2016). Building Stronger Communities: The Reciprocity between University, Student, and Community through Service-Learning. In *ProQuest LLC*.


Johnson, J. A. (2019). The Effect of Online Cross-Age Peer Tutoring on Student Self-Efficacy in Middle School STEM. In *ProQuest LLC*.


Sucre, S. (2016). Teaching the Whole Child through Physical Education and Youth Development. *Strategies (Reston, Va.)*, 29(2), 42–.


## Appendix III. Expert Informants

List of experts that recommended programs to be included in interviews and survey

<table>
<thead>
<tr>
<th>Expert Name</th>
<th>Organizational Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gigi Antoni</td>
<td>The Wallace Foundation</td>
</tr>
<tr>
<td>Deana Around Him</td>
<td>Child Trends</td>
</tr>
<tr>
<td>Charles Barrios</td>
<td>Administration for Children's Services, NYC</td>
</tr>
<tr>
<td>Bronwyn Bevan</td>
<td>The Wallace Foundation</td>
</tr>
<tr>
<td>Deb Bialeschki</td>
<td>American Camp Association</td>
</tr>
<tr>
<td>Dale Blyth</td>
<td>University of Minnesota</td>
</tr>
<tr>
<td>Isaac Castillo</td>
<td>Youth Invest Partners (formerly Venture Philanthropy Partners)</td>
</tr>
<tr>
<td>Krista Collins</td>
<td>The David P. Weikart Center for Youth Program Quality, The Forum for Youth Investment</td>
</tr>
<tr>
<td>David Condliffe</td>
<td>Center for Community Alternatives</td>
</tr>
<tr>
<td>Elizabeth Devaney</td>
<td>The Children’s Institute</td>
</tr>
<tr>
<td>Sean Flanagan</td>
<td>America's Promise Alliance</td>
</tr>
<tr>
<td>Felipe Franco</td>
<td>The Annie E. Casey Foundation</td>
</tr>
<tr>
<td>Tracey Hartmann</td>
<td>Research for Action</td>
</tr>
<tr>
<td>Heather Ikemire</td>
<td>National Guild for Community Arts Education</td>
</tr>
<tr>
<td>Mimi Ito</td>
<td>University of California, Irvine</td>
</tr>
<tr>
<td>Ranita Jain</td>
<td>The Aspen Institute</td>
</tr>
<tr>
<td>Stephanie Jones</td>
<td>Harvard University</td>
</tr>
<tr>
<td>Suzanne Le Menestrel</td>
<td>National Academy of Sciences</td>
</tr>
<tr>
<td>Akiva Liberman</td>
<td>Child Trends</td>
</tr>
<tr>
<td>Ivan Lui</td>
<td>Brooklyn Bridge Alliance for Youth</td>
</tr>
<tr>
<td>Kelly Murphy</td>
<td>Child Trends</td>
</tr>
<tr>
<td>Rachel Oberg-Hauser</td>
<td>City of Minneapolis</td>
</tr>
<tr>
<td>Jen Rinehart</td>
<td>Afterschool Alliance</td>
</tr>
<tr>
<td>Peter Scales</td>
<td>Search Institute</td>
</tr>
<tr>
<td>Marc Schindler</td>
<td>Justice Policy Institute</td>
</tr>
<tr>
<td>Victor St. John</td>
<td>Child Trends</td>
</tr>
<tr>
<td>Noel Tieszen</td>
<td>Consultant in Public Policy and Nonprofit Management</td>
</tr>
<tr>
<td>Gina Warner</td>
<td>National AfterSchool Association</td>
</tr>
<tr>
<td>Roger Weissberg</td>
<td>University of Illinois Chicago</td>
</tr>
</tbody>
</table>
Appendix IV. Programs That Participated in the Study

Programs that participated in interviews and survey

- Administration for Children’s Services NYC
- Cheyenne River Youth Project
- Community Alternatives
- Connected Camps
- Creative Action
- Creative Solutions
- Department of Youth and Community Development
- Fiver
- Free Street Chicago Theater and Performance for Social Justice
- Generation Citizen
- Genesys Works
- Girls on the Run
- Hillside Work-Scholarship Connection
- Inside Out Literary Arts
- Kalamazoo Youth Development Network
- Mikva Challenge
- Mural Arts
- Northwestern Center for Development
- Our Piece of the Pie
- ourBRIDGE for Kids
- Playworks
- Project Morry
- Shakespeare Behind Bars
- Sherwood Forest
- Ucode
- Urban Alliance
- Wings for Kids
- Words Beat & Life
Programs that participated in survey only

- Boys and Girls Club of San Francisco
- Camp Fire USA
- Eastern Michigan University Bright Futures
- Ignite Afterschool
- Juma
- LA’s BEST
- Outward Bound USA
- RYSE Youth Center
- Say Yes Buffalo
- West End House Boys and Girls Club, Boston
Appendix V. Data Collection Protocols

Documenting Outcomes and Measures in OST: Expert Interview Protocol

Introduction

Hi [ ]. This is my colleague [Notetaker's name]. S/he’ll be on taking notes or may jump in with a question or two. Thanks for agreeing to talk with us today.

Let’s start with just a quick explanation of this project we’re doing: The Wallace Foundation has engaged Child Trends to conduct a study to help inform the development of the Foundation’s next initiatives related to out-of-school time (OST) programming for children and youth. More specifically, Wallace is interested in better understanding what outcomes programs measure in their OST programming, what tools or methods (qualitative or quantitative) they use to measure outcomes (if any), and what barriers they experience when trying to collect high-quality outcomes data in OST programs. We expect today’s discussion to take about 30 minutes. Do you have a hard stop at 1:30?

To gather information on intended outcomes, measurement tools, and gaps in measurement, Wallace recommended that we talk to strong programs to learn more about what they are doing. To help identify those programs, we developed a list of experts and are looking for recommendations you have of programs that would be good for us to interview.

These programs may or may not have been formally evaluated and they may or may not necessarily be well-known. It could be a large and well-known provider of services or a mid-sized program based in one locality.

Other things that might be useful for you to know is that Wallace is interested in programs that offer content in arts, civic engagement/social justice, career and workforce development, STEM, or general services, as well as programs focused on SEL and online programs.

In addition, they are interested in programs that serve school-age children and adolescents from marginalized communities and those that embed an equity lens in their work and are also hoping for us to talk to a subset of programs that serve systems-involved youth.

Can you please recommend 2-5 strong OST programs in the social and emotional learning sector for us to potentially interview?

We are also interested in gathering information from you about your thoughts about outcome measurement in OST.

1. Questions for Expert Informants
   - Would you say that there are any common outcomes that most or many programs in [area] aim to target?
   - What do you see is the main outcomes of focus for OST programs in [area]?
     - Probe on specific outcome areas
   - Have you noticed any changes in outcomes of interest to OST programs/camps over time? Any changes in what programs/camps are interested in measuring? If so, why?
     - Probe on specific outcome areas
     - Are there any resources on OST measurement in virtual settings?
• Are there recent compilations of OST measures and tools that you’re aware of? We are interested in global measures as well as those that might target a specific outcome or program type (STEM, SEL measures, arts, leadership, etc.)

• Do you find that most programs base outcomes measurements on a theory of change?
  o Probe
    ▪ How many programs have a theory of change in the first place?
    ▪ Go back to theory of change when analyzing outcomes?

• What would you say are the key gaps in measuring outcomes in OST?
  o Probe about the following:
    ▪ Lack of available measurement instruments for certain outcomes
    ▪ Weak measures
    ▪ Lack of testing across populations
    ▪ Ease of administration
    ▪ Time
    ▪ Financial resources
    ▪ Usefulness to practice

• Are there other people you think we should reach out to speak about OST outcomes and measurement documentation?

• Are there any organizations/agencies that you think are thinking about outcomes and measurement in creative or innovative or particularly strong ways?

• Are there any additional resources we should review on outcomes, measurement instruments, or reports that we should review for our research?

II. Conclusion

• Is there anything I didn’t ask about that you’d like to share about documenting and measuring outcomes in the OST field?

Thank you so much for taking the time to speak with us today, we really appreciate your insight.

Are there any tools for measuring outcomes that you or programs you’ve worked with have found useful? (Request a copy)
Documenting Outcomes and Measures in OST: Program Interview Protocol

Introduction

Hi, my name is [YOUR NAME] and this is [NOTETAKER’S NAME], we’re from Child Trends. I will be leading our discussion today and [MY COLLEAGUE, name] will be taking notes. Thank you so much for taking the time to talk with us today. Let’s start with just a quick explanation of this project we’re doing: The Wallace Foundation has engaged Child Trends to conduct a study to help inform the development of the Foundation’s next initiatives related to out-of-school time (OST) programming for children and youth. More specifically, Wallace is interested in better understanding what outcomes programs measure in their OST programming, what tools or methods (qualitative or quantitative) they use to measure outcomes (if any), and what barriers are in place to collecting high-quality outcomes data in OST programs. We expect today’s discussion to take 30-45 min.

Your participation in the interview is voluntary. We’re happy to skip any questions you prefer not to answer. What you say here will be kept private, which means that your name will not be shared or associated with what you tell us today without your permission. [For programs only: If it’s okay with you, your organization will be included in a list of programs from which we gathered information.]

Also, as a sign of our appreciation for the time you will take to share your insights and experiences, we will send a $50 gift card to you after the interview. This will be for a website called “Giftbit” and you’ll be able to use it for a few different vendors such as Amazon, Target, or Starbucks amongst others. Is the email address we used to schedule the interview an ok one to send the link for this giftcard?

We plan to audio record today’s conversation and may transcribe the recording to help us capture all the details from our discussion. The audio-recordings will be deleted once the study is complete. Is that ok with you, do we have your permission to record today’s discussion?

Interview Questions

I. Interviewee Role and Background

Note for interviewer: The main goal of this section is to build rapport and gather background information on the interviewee and their organization.

To begin, we would like to hear a little about your role at the organization.

1. Can you tell us your title and a bit about your role at the organization?
   a. [If unaddressed from the overarching question] How long have you worked at [name of organization]?

II. Background Information about the Program

2. Can you tell us a little about the OST programming offered at [name of program]?
   a. What types of programming are offered in your organization?
   b. What are the goals or focus of the programs? If there are too many programs to talk about, we can focus on 1-3?
   c. [IF SURVEY DOESN’T HAVE INFO] What are the program locations? Note: this is geographic but also TYPE of place (i.e., school, rec center, etc.)
3. What is the target population of your program?
   a. Can you tell me a bit more about the demographic backgrounds of the young people in your programs?

4. Do you consider your program or organization to have a strong focus on equity? If so, how does this focus show up?

III. Intended or Targeted Program Outcomes

5. What would you say are the intended or valued outcomes for participants in your program? (if they are having trouble because they have too many programs, you can ask for each program – if there are too many to talk about, work together to choose 1-3) → Probe: What does "success" look like for an individual in your program?
   a. In what ways do you hope participants will benefit (in terms of outcomes) from your program?
   b. How were these focal areas for outcomes identified?
   c. How have your intended or target outcomes for program participants changed over time if at all?

6. Do you use a theory of change or logic model to guide outcomes measurement?
   a. How do you use it?
   b. If yes, would it be possible to share your theory of change or logic model with us?

7. In addition to what you have shared so far, are there any additional or secondary outcomes of your OST program(s)? These might be ones that you HOPE to see change in, but they are less of an intentional focus of the program or are more secondary?

8. Have you established shorter term (or interim) and longer-term outcomes that you hope your program will effect?

9. Has your program ever asked young people who participate in the program how they hope to benefit from their participation in the program?
   a. Has youth provided input into the selection of program outcomes?

10. Do you track whether there are differences in outcomes across different groups or populations that you serve (race/ethnicity, age, gender, location, etc.)?

IV. Measurement of Outcomes

11. Have you gathered any evidence on what outcomes are resulting from your program? If yes, what have you gathered?

12. What type of data do you collect, if any, to measure whether participants are benefiting from the program?
   a. **Probes:** Do you use surveys, interviews, administrative records, focus groups, observational assessments, tests, or other data collection instruments to assess whether participant outcomes are improving as intended?
   b. Did you create the data collection tool(s) that you use to assess program outcomes? If not, was it an off-the-shelf tool or was it developed by an external evaluator?
13. Do you use any other methods to document whether or how participants are benefiting from your program? Please describe what you use?

14. How did you decide what measurement tool to use? What other tools did you consider?
   a. [Probe if they only mention quant tools] Do you only use quantitative measures (surveys, test scores) or do you also collect and look at qualitative measures? This could be focus groups or interviews or detailed observations? If yes, tell me more about that? Where? When? How?

15. [IF USING TOOLS] Would it be possible for us to have a copy of the tools you’re using to collect data on participant outcomes?

16. Have you undergone evaluation by an outside, third-party evaluator?
   a. If yes, what methodology was used to evaluate the program?
      i. Probes: outcomes, pre-post, comparison group, implementation study, etc.

17. [OKAY TO SKIP] Once you analyze data and have information about outcomes, in what formats and with whom is this information shared? Anything else? Anyone else?

18. [OKAY TO SKIP] Are youth ever involved in your data collection, planning for data collection, data analysis or in interpretation of data?
   a. If so, how?

19. [OKAY TO SKIP] Does equity show up at all in your measurement approaches/strategies? Do you disaggregate data by gender or by race/ethnicity? Or by other factors?

V. Gaps and Challenges in Measurement

Now we want to switch gears and ask for your perspective on measurement gaps, or challenges you have experienced as you collect data on participant outcomes.

20. Are there outcomes for program participants you want to measure that you haven’t been able to?
    a. PROBE: Why haven’t you been able to measure them? Have you not been able to find a good instrument to assess a specific outcome of interest? Or you don’t know what question to ask?

21. Are there any outcomes of interest that you’ve tried to measure but have struggled to do so in some way? In what ways have you struggled (e.g., people skip a specific question, poor measures, everyone gives the same response, etc.)?
    a. Is there anything you’ve tried to measure that didn’t work to measure for some reason? (e.g., lack of available measures, the best way to measure the concept is not clear, or the measures you’ve found don’t perfectly align with your program model)?

22. What are the limitations of the tools or strategies that you’ve used or tried to use?
    a. Probe: Do you feel like your ability to measure outcomes of interest has been hindered by lack of available tools, or by poorly designed tools?

23. Do you feel you have the staff capacity and/or other needed capacity to conduct the measurement you want to conduct?
    a. If not, which is lacking?
    b. Why (i.e., is it hard to hire? Hard to fund? Hard to make time for?)?
24. Have there been challenges or other issues with the measurement instruments or strategies you've used?

25. [OKAY TO SKIP] Are there any other tools for measuring outcomes that you or programs you've worked with have found useful? Please describe.

26. [OKAY TO SKIP] Aside from measuring participant outcomes, what other data does your program collect to assess its performance? Probes: implementation, quality, program fidelity, organizational capacity, staff satisfaction, parent/participant satisfaction, etc.

27. What kinds of support would improve the measurement of outcomes in your program?

VI. Conclusion

28. Is there anything I didn’t ask about that you’d like to share about documenting and measuring outcomes in your program? Anything I should have asked about?

As a reminder, we would love if you could send us the [logic models, tools, etc.] we discussed today. Thank you so much for taking the time to speak with us today, we really appreciate your insight.
Appendix VI. Survey

Program Outcomes Email Questionnaire

- **Overview**

  The Wallace Foundation has engaged Child Trends to help inform the development of initiatives in learning and enrichment, in youth arts, and in the OST field at large.

  Specifically, Wallace is interested in documenting and measuring outcomes in OST programs with a focus on arts, civic engagement & social justice, career & workforce development, and social and emotional learning. This includes online programs, programs that operate afterschool or during the summer, programs that serve systems-involved youth, and programs that embed an equity lens in their work. In addition, Wallace is interested in outcomes measurement in OST programs that promote positive identity formation, and programs that serve school-age children and adolescents from marginalized communities.

  To learn about intended outcomes, the breadth of measurement approaches and instruments, and measurement gaps and challenges, we are reaching out to programs that were recommended by OST field leaders. We would like to know whether and how your program measures outcomes for program participants. We want to learn about the measurement tools you’re using – or are interested in using – to assess participant outcomes, and any challenges you’re experiencing pertaining to outcome measurement.

  This information will inform work that Wallace does in the future to support OST programs across the country.

  Thank you!

- **Organization Name**

- **Name of program mentioned in the email, if different than organization name**

- **Thinking of the program mentioned in the email, how many sites does your program have?**
  - 1
  - 2
  - 3
  - 4
  - 5
  - More than 5

- **Who operates your program?**
  - Government entity
  - Private community organization
  - Other
• Program Location (Select all that apply)
  o Parks and Rec
  o Schools
  o Community-based organization
  o Other

Program 1 (pick your largest or most important program)

• Program Type (Select all that apply)
  o SEL
  o Arts
  o Civic engagement
  o Social justice
  o Leadership development
  o Identity-focused
  o Career/workforce
  o Sports/physical fitness/health
  o Stem
  o Other

• When is this program offered (select all that apply)?
  o Before school
  o After school
  o Summer
  o Other

• Is this an online program?
  o Yes, my program was designed to be an online program
  o Yes, we've adapted because of COVID-19
  o No
  o Not sure

• Age Groups Served (select all that apply)
  o 5-12
  o 13-17
  o 18-20
• Do you consider your program or organization to have a strong focus on equity?
  o Yes
  o No
  o Not Sure

• What are intended outcomes for your program?

• Have you undergone a formal evaluation?
  • Yes
    o Date of most recent evaluation
    o Who conducted the evaluation?
    o What were the results of the evaluation?
    o What methods were used to evaluate your program?
  • No

• Do you have a logic model?
  o Yes
  o No

• Do you have a theory of change?
  o Yes
  o No

• Do you have a performance management system (Efforts to Outcomes - ETO, Apricot, KidTrax, CitySpan, etc.)?
  o Yes
  o No

• Do you have a staff person who helps support data collection and/or analysis?
  o Yes
  o No

• Program Measurements
  • Do you use any tools to measure outcomes for your program’s participants? These could include, but are not limited to, quantitative tools such as survey instruments or assessment measures, as well as qualitative tools such as interview protocols or focus group guides, or other data collection tools that you’re using.
    o Yes
    o No
• If yes, what measurement tools does your program use (e.g., surveys, interviews, assessments, portfolio, performance)?

• What is the most challenging aspect of collecting data on youth outcomes for your program?
  o Are there any other youth outcomes you are unable to collect?

• What are the youth outcomes you’re unable to collect and why aren’t you able to collect them?

• What supports would be most useful to your program to improve your outcome measurement strategies?

If you have additional programs and would like to add background information for those programs, please select "submit" and then start over.

OR, press "finish" to submit your answers.
References


