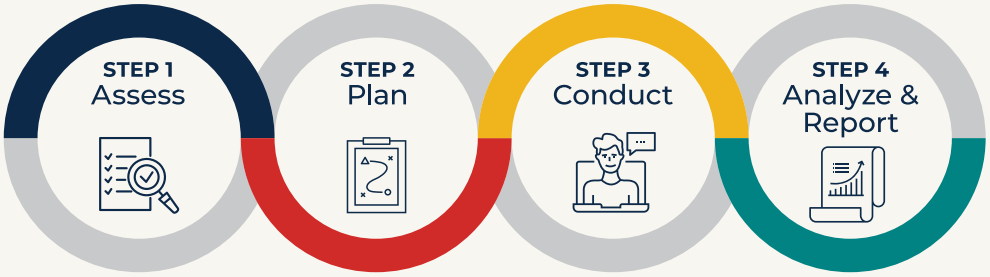
Sample Size Guide

This tool is part of Mathematica's suite of measurement and evaluation (M&E) tools, which provides a road map for generating timely and actionable evidence about what works for whom, and in what context. The tools were designed to promote rapid innovation and scaling of promising solutions (such as programs, practices, or products). The Sample Size Guide is used in Step 2 of the M&E process.



Learn more about the M&E process and other tools here:
https://www.mathematica.org/features/advancing-educational-equity

Who should use the Sample Size Guide?

This guide is designed for funders and organizations implementing solutions with support from a research partner. A research partner could be staff within the organization who have expertise in evaluation, or an external organization, technical assistance provider, or consultant.

What is the Sample Size Guide?

This guide provides an overview of the sample size considerations for each of the evidence-building phases: Design the Solution (Phase 1), Refine the Solution (Phase 2), Assess for Early Evidence of Success (Phase 3), and Validate Effectiveness (Phase 4). Organizations can use this document to inform their study's sample sizes as they begin recruitment and planning for measurement and evaluation activities. In Phase 3 (Assess for Early Evidence of Success) and Phase 4 (Validate Effectiveness), organizations can use this guidance as a starting point but should work with a research partner to determine the specific sample size needed for their study.

Acknowledgments

Mikia Manley, Virginia Knechtel, and Gregory Chojnacki developed this guide. John Deke provided guidance and conducted the power analyses that informed the sample size recommendations. Alex Resch and Kate Place reviewed the content and provided feedback. Grantee partners and Mathematica technical assistance liaisons also provided feedback on early versions of this guide. Sheena Flowers provided design and production support, and Jennifer Brown provided editorial support. This publication was prepared for the Bill & Melinda Gates Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation.

Sample Size Guide

Overview

This guide provides an overview of the sample size considerations across all evidence-building phases, as well as guidance for each phase. Sample size is one important consideration in planning for a strong evaluation, but other factors, including evaluation design, measures, and quality of implementation, play equally important roles. The Measurement and Evaluation Checklist included in this toolkit provides more guidance on the type of research designs recommended or required in each phase.

Recommendations across all phases

- **Establish the purpose of the evaluation, along with the research design and methods.** Different evaluation goals, such as testing a solution's usability versus measuring the effect of a program on student outcomes, require very different sample sizes. The Measurement and Evaluation Checklist provides more guidance on the type of research design recommended or required in each phase, and in Table 1, we list recommended sample sizes for different evaluation goals.
  - Focus groups, interviews, and observations: We recommend including at least 8 participants (for example, students, teachers, school leaders) when seeking qualitative feedback during the study. Examples of qualitative methods include focus groups, interviews, and observations (for example, observing classroom instruction or teacher professional learning).
  - Usability testing: We recommend including at least 5 students or other users for each round of usability testing.
  - Quantitative research methods: We recommend including at least 50 users to reasonably identify trends or patterns if analyzing administrative, program, or survey data. If study participants will be assigned to treatment and comparison groups, we recommend including at least 50 in each group. When conducting analyses within subgroups, we recommend using a larger sample size. See details below on the importance of planning for subgroup analysis.
- **Account for participant consent and sample loss when determining sample sizes.** Some organizations may need to recruit additional schools, teachers, or students because not all those who are recruited may consent to participate in your study or remain in your study through data collection.
  - The approach to obtaining consent (either passive or active consent) is one factor to consider as you determine how many more participants to recruit. Passive consent requires potential participants (or their parents/guardians) to opt out if they do not want to participate, while active consent requires potential participants to opt in if they do want to participate; passive consent rates are nearly always higher than active consent rates. Consult local, state, and federal agency guidelines, as well as Institutional Review Board guidance, to identify applicable consent requirements. Organizations should recruit at least 5 to 25 percent more individuals than needed. For example, if you are aiming to have a sample size of 100 students, recruit at least 105 students if you will seek passive consent or at least 125 students if you will seek active consent.

  - It is also common for some students or teachers to leave the study. They may drop out of the program or leave the participating schools or districts. The longer the study, the more loss you are likely to experience. Organizations should work with their research partner to determine how to plan for this loss.

- **Plan for subgroup analyses before recruitment begins.** A central principle in the Equitable Evaluation Initiative's guiding framework is that research should aim to understand connections between systemic drivers of inequity and the program or practice being studied (Equitable Evaluation Initiative 2021). As part of this effort, and consistent with the Measurement and Evaluation Checklist guidance, study teams should plan for subgroup analyses whenever possible—that is, when the sample size is large enough and the sample is comprised of diverse identities. Subgroups may include students or teachers, across types of context, or by user characteristics (race/ethnicity, English learner status, gender, socioeconomic status, grade in school). Decisions about whether to conduct subgroup analyses and for what purpose will influence the sample size needed for the study.
  - If you will use subgroup analyses to test a central hypothesis of the study (rather than simply to provide additional context or suggest avenues for future investigation), organizations should work with their research partner to account for this when deciding how large the sample should be. An example of this would be if an organization wants to implement your program among a racially diverse population, but the study's key hypothesis is that the program benefits the subgroup of participating students who are Black and Latino.
  - A larger sample size is required to conduct subgroup analyses.
- **Include the right people in your evaluation.** When designing equity-focused evaluations, *who* is in the sample matters as much as *how large* the sample is. We recommend reflecting on the following questions when planning for sample recruitment: Does your sample include a diverse group of participants from the community in focus? Are the intended users of the solution (including the proposed implementers of the solution) included in the sample? Who is not included in your sample? What efforts are made to include perspectives from various collaborators or understand differences in outcomes for members of the community?

Table 1. Sample size requirements and research designs and goals, by phase of evidence building

Phase	Sample size requirement	Research design and goal
Phase 1 Design the Solution	**No sample size requirement**. Please review the recommendations across all phases, including the reflection questions listed above.	• No research design (solution is being designed in this phase) • Organizations work to unpack the problem and identify potential solutions • **Goal:** Develop a solution based on a well-defined theory of change for how that solution is expected to lead to improved outcomes for students in a community in focus
Phase 2 Refine the Solution	Organizations should include **at least one school and engage at least 5 students (or other single analysis units)** in their study to understand implementation (Nielsen 2000). The final sample size will depend on how organizations plan to assess implementation.	• No comparison group required • Low burden research methods are used to (1) understand the implementation context; (2) learn if the solution is usable, utilized, and useful; and (3) learn if the solution may lead to improved outcomes • **Goal:** Refine solution and develop descriptive evidence that the solution was successfully implemented in the community in focus
Phase 3 Assess for Early Evidence of Success	Organizations should plan to include 100 to 200 students, if they are assigning individual students to treatment and comparison groups or at least 50 students if they are using a correlational analysis. Alternatively, they should plan to include about 6 schools if they are assigning schools to treatment and comparison groups or 10 to 20 classrooms if they are assigning classrooms. These estimated sample sizes are for planning purposes, and organizations should determine the final sample size in collaboration with their research partner. Sample size will be influenced by which analytic approach is used, whether the organization plans to analyze subgroup differences, and how small of an effect the organization wants to be able to detect, among other factors. (Smaller effects can only be detected in larger samples.) The size of the effect you want to be able to detect should take into account the cost and intensity of the solution.	• Small pilot • Randomized controlled trial (RCT) or quasi-experiment design (QED) preferred; correlational analysis (pre/post) with statistical controls also acceptable • **Goal:** Demonstrate that the solution is associated with improved outcomes for students in the community in focus

Phase	Sample size requirement	Research design and goal
Phase 4 Validate Effectiveness	**Required:** Use formal power calculations to determine the minimum sample size needed **to detect statistically significant differences on outcomes.** Organizations should plan to include at least 350 to 500 students[1] if they are assigning individual students to treatment and comparison groups. Alternatively, they should plan to include at least 20 to 50 schools in total if assigning schools to treatment and comparison groups, or 50 to 100 classrooms if assigning teachers. Organizations should determine which range to use based on whether the solution is offered at the school, classroom, or student level. These ranges are for planning purposes only, and organizations should determine the final sample size in collaboration with their research partner. Sample size will be influenced by the level of randomization or matching, which analytic approach is used, whether the organization plans to analyze subgroup differences, and how small of an effect the organization wants to be able to detect, among other factors. The size of the effect you want to be able to detect should consider the cost and intensity of the solution.	• Large pilot with a comparison group • RCT or QED required • **Goal:** Generate causal evidence that the solution leads to improved outcomes for students in the community in focus

[1] A study must have a minimum sample of 350 participants to be eligible to meet the ESSA Tier 1 Strong Evidence level or the ESSA Tier 2 Moderate Evidence level (Regional Educational Laboratory Midwest 2019).

References

Nielsen, J. (2000). *Why you only need to test with 5 users*. Nielsen Norman Group. **https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/**

Regional Educational Laboratory Midwest. (2019). *ESSA tiers of evidence: What you need to know*. Institute of Education Sciences, U.S. Department of Education. **https://ies.ed.gov/ncee/edlabs/regions/midwest/pdf/blogs/RELMW-ESSA-Tiers-Video-Handout-508.pdf**

Equitable Evaluation Initiative. (2021). *Shifting the evaluation paradigm: The Equitable Evaluation Framework.* **https://www.equitableeval.org/blog-main/shifting-the-evaluation-paradigm-the-equitable-evaluation-framework-eef**