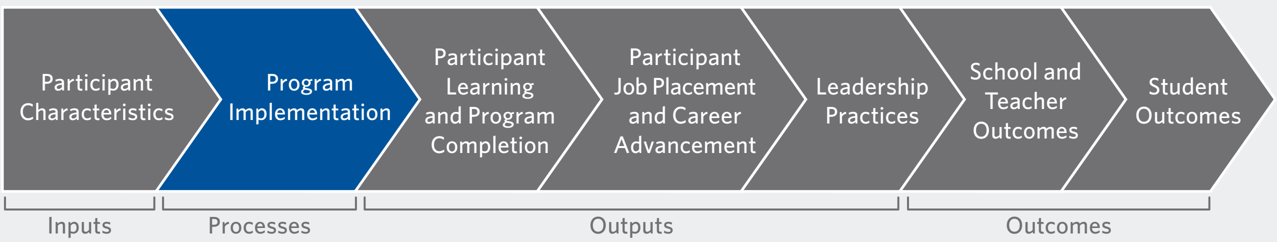
3. MEASURING PROGRAM IMPLEMENTATION

High quality evaluation of principal preparation programs examines inputs, processes, outputs, and outcomes. These components are depicted in the evaluation pathway framework below.

Evaluation Pathway for Leadership Preparation Programs (adapted from [UCEA, 2013](#)):



This paper is part of a series documenting lessons learned by New Leaders in utilizing this framework to guide internal evaluation of its Aspiring Principals Program (APP). The paper focuses on the second step in the pathway, program implementation.

Program implementation studies examine whether the program was implemented as intended, including whether participants received the full dosage of the program "treatment" (i.e., program experiences and support) and with what level of quality. Measuring program implementation can: 1) clarify assumptions embedded in program design; 2) guide program improvements; and 3) help explain and contextualize the impact results (Resnick, et. al, 2005; O'Donnell, 2008; Century, et. al, 2010).

Program implementation evaluation should track components of the program that designers expect will have the most substantial effect on program goals.

Examples of typical principal preparation program activities include:

- Curriculum and coursework content (quality / alignment with standards)
- Content delivery (dosage and quality)
- Assignments (completion and quality)
- Coaching and feedback (dosage and quality)
- Internship/Residency (length and/or quality)
- Instructional strategies and practice (including dosage)
- Knowledge of faculty/instructors, including coaching (quality)
- Technical assistance, such as placement assistance (dosage and quality)

Evaluators often need to engage in conversations with program designers to clarify the program activities, including their intended dosage and characteristics of quality. This clarification process can help to make design assumptions explicit. For example, New Leaders coaches are expected to provide high quality feedback on authentic work products submitted (e.g., video of leader practice or work artifacts) without visiting schools. Although it can be time consuming to surface the theory of action and design assumptions, the process can be helpful to program designers. When program designers make their implicit theories explicit, they are often better able to clearly communicate the critical features to program implementers, which in turn can improve the quality of implementation. For example, New Leaders program design is grounded in socio-cultural learning theory, which suggests that one of the most important drivers of participant learning is high quality feedback on authentic opportunities for practice. In order to track whether this aspect of the program (high quality feedback) was being implemented, the program designers had to define high quality feedback. This work not only informed the program evaluation, it also informed training for the program instructors whose role it was to provide the feedback.

The evaluation should also pay careful attention to program features that are likely to vary in implementation—even if the provider does not fully control that feature—in order to examine the relationship between implementation and results. Implementation might vary due to different levels of staff capacity or due to different contexts. For example, New Leaders historically did not track the quality of the participant's residency site, even though it is a critical program feature, because it had limited control choosing the sites. However, the organization decided to start measuring and tracking resident site quality so that it could study whether and how that aspect of the program was related to program outputs, such as participant learning and program completion. It is important to monitor activities which may have variability so as to understand the program as precisely as possible and document possible patterns across sites, participants, or timeframes.

Program implementation often involves measuring both fidelity (e.g., was the program component implemented?) and quality (e.g., was the program component implemented well?). Measuring fidelity often involves creating dosage tracking systems, such as attendance at courses, assignment completion, and coaching logs. Quality, on the other hand, can be measured via participant perceptions (often captured via survey) and/or program quality review (which often entails rating a component against a rubric). New Leaders uses a set of surveys created by the University Council of Educational Administration (UCEA), called the INSPIRE Leadership Suite of program evaluation surveys, that programs can use to assess common program elements. These annual measures of program implementation can be particularly useful for examining the relationship between implementation and program outputs. The INSPIRE surveys provide the added benefit of being able to benchmark results to national norms. New Leaders also creates and administers its own interim surveys five to ten times throughout the year to monitor quality of implementation.

SUMMARY OF NEW LEADERS APPROACH TO MEASURING PROGRAM IMPLEMENTATION

The following chart summarizes the indicators and data sources that New Leaders uses to track implementation of its key program components.

Program Component	Indicator	Data Sources
Formal Learning Sessions	Session content quality	Participant session surveys; Participant comprehensive surveys (mid-year and end-of-year); Session evaluation by national staff
	Session facilitation quality	Participant session surveys; Participant comprehensive surveys (mid-year and end-of-year); Session evaluation by national staff
	Number of sessions delivered	Program calendar (national and site-specific); Participant participation tracker
	Participant attendance	Participant participation tracker
Residency	Quality of experience	Residency site assessment; Participant comprehensive surveys (mid-year and end-of-year)
Coaching	Quality of coaching	Participant comprehensive surveys (mid-year and end-of-year); observation ratings by managers
	Quantity of coaching	Coaching log (duration & type of support)

New Leaders conducts descriptive quantitative and qualitative analyses immediately upon availability of each new data source and makes results available to relevant program and city staff, often through facilitated data conversations that allow performance feedback to be turned into useful information for adjusting tactics to meet program goals. For example, a facilitated data conversation highlighted evidence that monthly assessments were consuming considerably more time than expected for participants to complete and for program staff to consistently score and provide quality, timely feedback. New Leaders therefore decided to reduce the number of assessments, which in turn yielded more thorough completion from participants and more thoughtful, timely feedback from program staff.

The process of using data for performance feedback—i.e., rich cycles of data collection, analysis, and action—is deeply embedded in the work of all New Leaders staff. For example, local city teams meet at least once per month to assess APP participant and school progress, and to strategize for immediate improvement. Executive Directors from each city meet quarterly with national staff to analyze targeted data streams and to share learning lessons from across program sites, making adjustments for the current year's participants and for future trainees. Finally, each programmatic area holds at least two step backs each year to review data, identify lessons learned, and identify implications for planning for the next year.

RECOMMENDATIONS BASED ON LESSONS LEARNED

Recommendations based on New Leaders' experiences measuring program implementation are:

- Ensure that program implementation evaluation is a **collaborative process** by encouraging program staff to provide input into evaluation tools. For example, New Leaders staff responsible for delivering the Aspiring Principals Program help determine the content of New Leaders-developed surveys. This ensures that participants are responding to items that clearly connect to the program initiatives and hypotheses. Collaboration not only increases the accuracy of the data tools, but also allows program staff to see the value of investing time and energy in collecting the data— -- which in turn increases the likelihood that the data collected may yield information that will inform effective programmatic changes.

- Improve **survey response rates** by administering surveys (of all kinds) during training. We recommend using an online survey platform, but beware that this approach necessitates participants in having internet devices and access during training. It is also a good idea to have back-up paper survey versions available for improved response rates.

- When using **survey administration software**, ensure that the program or service offers suitable functionality to the tasks at hand. For example, New Leaders uses Qualtrics to administer surveys, as it allows for confidential survey responses, as well as the linking of responses to key respondent data (such as demographic data). It also offers a larger variety of survey question types (such as matrix questions and sliding scale questions), and allows for skip or display logic to be used based on previous responses or respondent demographic data. The reporting function also offers several options for filtering and sharing results. The software is also able to translate the data into our core database (Salesforce) for storage and analysis.

- **Examine results regularly** (at least annually and more often if possible). Use results to consider whether and how fidelity of implementation can improve (single loop learning) AND to consider whether and how to adjust the theory of action (double loop learning).

REFERENCES

Blase, K. and Fixsen, D. (2013). Core intervention components: *Identifying and operationalizing what makes programs work, ASPE Research Brief*. Office of Human Services Policy, US Department of Health and Human Services: Washington, DC

Century, J., Rudnick, M., and Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*. 31(2): 199-218.

O'Donnell, C. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in k –12 curriculum intervention research. *Review of Educational Research*. 78(1): 33-84.

Resnick, B., Bellg, A. J., Borrelli, B., DeFrancesco, C., Breger, R., Hecht, J., et al. (2005). Examples of implementation and evaluation of treatment fidelity in the BCC studies: Where we are and where we need to go. *Annals of Behavioral Medicine*, 29, 46–54.

University Council for Educational Administration. (2013). *Developing evaluation evidence: A formative and summative evaluation planner for educational leadership preparation programs*. Charlottesville, VA: Author.