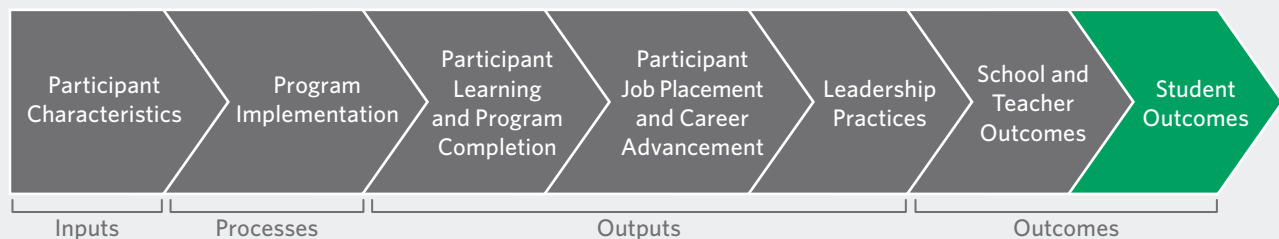


## 6. MEASURING STUDENT OUTCOMES

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 final step in the pathway, student outcomes.

Using student outcomes (achievement test results, graduation, attendance, etc.) for evaluation of principal preparation program evaluation is a critical component of examining program outcomes. However, the use of student outcomes can be somewhat challenging because the data used for these analyses can be flawed and results can be difficult to interpret. In fact, student outcome analyses may not be feasible or appropriate for some programs due to lack of data availability or small sample sizes.

The primary challenge with student outcome results is that they can be an imperfect measure of a program's impact. Despite limitations, New Leaders has found that when results are interpreted with caution and they are used as part of a more comprehensive program evaluation, student outcomes can be an important and useful component of its program evaluation.

### VALUE OF USING STUDENT OUTCOMES

As described in the [first paper in this series](#), high-quality program evaluation includes examination of final outcomes. In the case of principal preparation programs, the final outcomes are student outcomes. New Leaders has experienced three major benefits from examining student outcomes:

- 1. Internal accountability.** Focusing on student outcomes has helped New Leaders maintain a commitment to rigorous program design and execution. For example, like many programs, New Leaders experiences many different pressures from funders and partner district to endorse as many program participants as possible. Our commitment to assessing and reporting the student outcomes of schools led by endorsed New Leaders principals drives us to only endorse program completers who (according to their assessment scores) are likely to be successful in improving student achievement—even when withholding endorsement disappoints funders (who want to maximize the yield for their investments), district partners (who want as many qualified candidates for open principal positions as possible), and the program participants (who invested significant time and energy into the program).
- 2. Program improvement.** While New Leaders has demonstrated statistically significant positive impact on student outcomes for its Aspiring Principals Program, our outcome analyses show substantial variation underneath that overall positive effect. These results became important starting points for further investigation into root causes for the variation in results. Those inquiry processes have generated important insights and learning lessons that have been the basis for major program redesign, such as a completely new approach to program admissions.
- 3. Communicating the value of the program.** When student outcomes are positive, they can provide powerful messages to candidates and partners that the program is achieving its goals. Documenting positive student outcomes has helped New Leaders build support to sustain and expand programs.

## RESPONSIBLE USE OF STUDENT OUTCOMES DATA

Users of student achievement analysis results should take care to understand the strengths and limitations of the data and always triangulate the results with additional data sources—especially prior to making high stakes decisions. Responsible use includes:

- 1. Understanding the limitations of the data and analyses.** Great care is needed because the accuracy and validity of the data (as an indicator for program impact) can be uncertain. Some state tests are widely criticized regarding the extent to which they are a valid measure of whether students have learned the skills they need to be successful in college, careers, and citizenship. Even if tests are valid measures of what students know, school-level results may not be a valid measure of whether the school has improved student learning. For example, in one instance New Leaders' school-level analyses showed that a school had double digit losses in student achievement, but student-level results showed that the school made statistically significant gains. Further investigation revealed that the school-level losses were largely attributable to major shifts in student population caused by boundary line changes. Just like medical tests that have high false-positive or false-negative results can be quite useful when triangulated with other data, imperfect student outcome data can be quite useful to preparation programs when triangulated with other data, appropriately interpreted, and used in the manner described below.<sup>1</sup>

---

i. Note: this paper is intended for preparation program providers, not states or accreditation organizations. We recommend a much higher bar for using student outcomes data for high-stakes accountability, as described in our report [Change Agents: How States Can Develop Effective Leaders](#)

## 2. Maximizing the validity of results—to the extent possible—by:

- **Examining changes in student outcomes.** Analyses that do not examine change (e.g., gains in test scores, improvements in graduation rates) are likely to be more descriptive of the where the program graduate was placed than their impact.
- **Examining results after 3 years.** Research conducted by our external partners (the RAND Corporation and Mathematica) and in the broader field suggests that the indirect nature of principal effects mean that they often are not measurable until the principal has three years of tenure in the same school.
- **Setting minimum thresholds for sample size.** Programs may not be able to make valid interpretations of student outcomes analyses until programs reach a sample size of at least five to ten participants who have at least three years of usable data from the same school.
- **Using student-level data when possible** because it allows for the most rigorous analyses. Student-level data provides an individual record for each student which includes: actual test scores; personal characteristics, such as free and reduced price lunch (FRPL) eligibility, English language learner (ELL) status, special education services, and ethnicity; it usually can be linked across years to create a longitudinal data set for individual students; and it can be “rolled up” to classroom- or school-level. Student-level data, however, are not publicly available and programs need to arrange for permission to access the data. Access might be costly if the district or state charges a fee for the data or has data safeguarding requirements that are costly to implement.

3. **Using multiple indicators and metrics, including those that are aligned with organizational goals.** While there are some basic measures that might be appropriate for all programs to use, programs may want to select additional measures based on their unique programmatic framework, mission, and goals. Programs should identify research questions and metrics that are aligned with their theory of how they expect the program to impact student achievement. For example, New Leaders mission and goals are very focused on closing the achievement gap. Program design is focused on that goal and therefore it is important for us to identify metrics that help us measure impact on achievement gap.

4. **Triangulating results with other data and information.** In addition to examining multiple indicators of student outcomes, users should also triangulate student outcomes with other data and information, such as principal supervisor perceptions of principal effectiveness and other data collected through the program evaluation regarding program inputs, implementation, and outputs.

These recommendations should improve the validity of student outcome results but the results are still likely to be less than perfect. Yet, they can still be valuable. Nearly all types of data, including survey data and assessment data, have limitations. Student outcome data is no different. The field of medicine routinely uses less than perfect data. For example, during annual exams, doctors monitor indicators of general health such as weight, blood pressure, and visual scans of eyes, ears, and mouth, etc. They seek additional information when an indicator is abnormal or changes substantially. For example, if a patient has high blood pressure, responsible doctors do not immediately assume that the patient has a blocked artery and schedule bypass surgery. Instead, before beginning any kind of medical intervention, doctors gather more information by asking questions about symptoms, checking the patient’s medical history, confirming family history, and possibly running more tests.

Likewise, imperfect student achievement results can be useful to programs. Individual results can be a flag for further investigation. Patterns over large samples and/or over time can be combined with other information to decide when program design changes are warranted. For example, when New Leaders gathered more information about program graduates with poor student outcomes results, we found a theme of weak skills related to adult leadership. This insight led us to design a new teacher leader program focused on adult leadership skills that participants complete prior to the Aspiring Principals APP, which gives them an additional year to develop and practice those skills.

**SUMMARY OF NEW LEADERS APPROACH TO STUDENT OUTCOME EVALUATION**

The New Leaders Aspiring Principals APP is designed to prepare principals to close the student achievement gap and improve overall student outcomes. The program is also designed to outperform other preparation programs by providing a high-quality preparation experience with a robust residency component.

New Leaders contracted with the RAND Corporation in 2006 to conduct a comprehensive evaluation of the implementation of the New Leaders program and its effects on student achievement and other outcomes, such as graduation rates and attendance rates. The RAND evaluation tracked program effects on student achievement outcomes through a value-add analysis of individual student achievement data provided directly to RAND by school districts.

Since RAND’s analyses of student-level data take time and resources, New Leaders also has a robust internal program evaluation to provide quick-turnaround results that can inform program improvements and demonstrate effect. The New Leaders internal program evaluation tracks student achievement outcomes with school-level student achievement data that is publically available on state websites. On an annual basis, the internal research and evaluation department downloads the publicly available student achievement data for each district or state in which there are at last five principals who have been endorsed by New Leaders and who are currently serving as principals.

New Leaders uses these data to build a longitudinal database of student achievement and student learning gains data, including both New Leaders-led schools and other schools in our partner district and charter schools. This database enables New Leaders to map student achievement results with other programmatic data, including placement information, recruitment and selection data, and participant assessment data.

The table below provides examples of indicators and metrics that New Leaders uses to measure progress toward student outcome goals.

Goal	Indicator	Metric
Close Achievement Gap	Percent of students scoring proficient or greater in Math and Reading	Percent of K-8 schools on track to statewide averages in math, reading, and in both subjects
Close Achievement Gap	High school graduation rate	Percent of schools on track to statewide average graduation rates
Improve Overall Achievement	Change in percent of students scoring proficient or greater in Math and Reading	Percent of K-8 schools making math, reading, and combined proficiency gains

Goal	Indicator	Metric
Outperform	Performance on Local Accountability Metrics	The percent of New Leaders schools outperforming partner district performance on local accountability metrics
Outperform	Change in percent of students scoring proficient or greater in Math and Reading	<p>The average change in percent of students scoring proficient or greater in K-8 New Leader schools compared to the district and a matched sample of schools.</p> <p>The percent of K-8 New Leader schools making proficiency gains greater than the average proficiency gain of the partner district and a matched sample of schools.</p>
Outperform	Value Added Measures	Statistically significant effect sizes calculated by external RAND evaluation

New Leaders tracks graduate and alumni career trajectory information in Salesforce. This electronic database allows the organization to build a comprehensive record of program graduate placements that is used to conduct analyses for various stakeholders. The database also allows New Leaders to merge placement data with other outcomes databases (like the longitudinal student achievement database).

**RECOMMENDATIONS BASED ON LESSONS LEARNED**

Many programs have funders that require them to examine student outcomes. Government agencies and accreditation processes might also require reporting of student outcome results. Even if this expectation is not the case, we recommend that principal preparation programs use student outcomes for:

- Internal accountability
- Program improvement
- Communicating the value of the program

Programs should use student outcomes data responsibly by:

- Understanding its limitations
- Maximizing validity of results (to the extent possible within resource, time, and data availability constraints)
- Using multiple indicators and metrics, including those that are aligned with organizational goals
- Triangulating results with other data and information

We also recommend the following strategies for addressing challenges we have encountered in collecting and reporting student outcome data:

- **Principal tenure data.** Having updated information on where program graduates are placed is critical because it defines which schools to include in student achievement analyses. Unfortunately these data are often imperfect or absent because principal tenure data is usually not publicly available, district data files are often outdated, and programs often lose touch with program graduates. When tenure data is missing or its accuracy is unknown, it can severely hinder the usefulness of examining student achievement results.

- Invest in meticulous tracking of program participant placement and tenure, including maintaining relationships with alumni. See the paper in this series on tracking graduate job placement for more information.
- **Data accessibility.** Often, it is difficult to either obtain or use the desired data. Preparation programs may not have necessary permissions to handle individual student-level data and processes for obtaining the data from districts may be slow, cumbersome, and sometimes costly. Publicly accessible data may be in formats that are difficult to use, have missing elements, or lack longitudinal consistency.
  - Build relationships with districts to aid in accessing student-level data (preferably with memorandums of understanding (MOUs)). New Leaders builds into all of its agreements with districts provisions for gathering student-level data, usually at minimal or no cost. New Leaders, whenever possible, builds personal relationships with the head of the district research and evaluation offices as well.
- **Cost to collect and analyze data.** Collecting and analyzing student outcomes can be time consuming and complicated given the limitations above, and organizations have limited human and financial resources to devote to the process.
  - Seek funding for rigorous, individual-level analyses.
  - Embed other program evaluation costs into the program cost model.
- **Time lapse between program participation, placement, and results availability.** The time between when our program participants complete our program and when their student outcomes results are available is typically two to four 2-4 years. While some program completers enter the principalship immediately, others might take a couple of years.
  - Try to set accountability metrics for three or more years in the principalship, while tracking interim data.
- **Stakeholder timelines.** In spite of logistical challenges, various stakeholders want results as soon as possible, leaving little time to address the lack of availability, the complexity of the data, or the time lapses in results. For example, many funders require updated results on an annual basis.
  - Make the investment in student-level data when possible, but use publicly available data as well to meet timeline requirements. New Leaders draws on both kinds of data to report on multiple metrics.
- **Comparability across states.** Since states typically have different state assessments, many metrics and indicators cannot be appropriately aggregated across states.
  - Develop metrics that do not refer to specific tests or performance indices (e.g., New Leaders uses percent of schools on track to meeting statewide proficiency averages).
- **Test changes over time.** This issue is particularly challenging when states and districts change tests, which makes gains or losses attributable in part to changes in the test rather than to the school or students.
  - Examine the school achievement relative to district averages from one year to the next.

## REFERENCES

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