UC BERKELEY
PRINCIPAL LEADERSHIP INSTITUTE

PLI

IMPACT REPORT
Experiments with Student Outcomes

W. NORTON GRUBB

PATRICK LIAO

REBECCA CHEUNG[1]

EXECUTIVE SUMMARY

In the last two years, UC Berkeley's Principal Leadership Institute (PLI) has published three impact reports: one focused on the breadth of regional impact and strength of program model, another focused on alumni and district partner feedback, and the third on collective leadership of all PLI alumni teams. This final report focuses on our experiments with evaluating impact using student outcomes. In its fifteenth year, the PLI now has 491 graduates, 88% of whom work in the Bay Area; 50% work in PLI's four partner districts. In response to growing calls from districts, government agencies, and foundations to increase accountability for preparation programs to "prove" their impact, PLI conducted an experimental evaluation related to school level outcomes focused on alumni who served as principals for a minimum of three years. We learned that our alumni as a group are more diverse than California's school leadership. PLI alumni principals were also more likely to work in schools that serve the most disadvantaged student populations. We were unable to draw conclusions about our program related to standardized test scores and suspension rates. However, we did find that PLI-led schools had stronger school cultures, especially related to developing positive relationships with students as measured by the California Healthy Kids Survey. Finally, we were able to identify major shortcomings and barriers related to rigorous analysis related to both insufficient data and methodological challenges.

[1]The authors are respectively the Faculty Coordinator; Research Associate; and the Program Coordinator of the Principal Leadership Institute, the University of California, Berkeley.

University of California, Berkeley
Principal Leadership Institute
GRADUATE SCHOOL OF EDUCATION

INTRODUCTION

In this age of accountability and evaluation, teachers have been subject to competency exams in the 1980s and now value-added measures. States imposed accountability systems on schools during the 1990s, and the federal government created accountability for districts in No Child Left Behind. Now accountability and more formal evaluation have spread to leaders, because they may be second in importance only to teachers.

Closer to home, foundations have asked our program—the Principal Leadership Institute (PLI) at the University of California, Berkeley—how we "prove" our impact, including effects on test scores; many foundations do not fund programs without outcome evidence. Federal grants ask grantees to quantify their effects, including those on student outcomes. In California an ongoing discussion concerns the need to "improve" and "monitor" leadership programs, though measures of quality have not emerged. The evaluation of principals and leadership programs is firmly on the agenda of districts, governments, and foundations.

This fourth Impact Report chronicles our efforts to measure the effectiveness of the PLI via school level indicators. Evaluating the impact of leadership programs based on student outcomes proved extremely complex. Therefore we consider our efforts to be *experimental* because of insufficient data and methodological challenges. Rarely were our results comprehensive enough to make conclusions about effectiveness. They were useful for uncovering barriers to rigorous analysis. Therefore, we believe that our process is instructive and points to important challenges and issues for both preparation programs as well as policy makers and/or foundations who promote this type of accountability.

EXISTING EVALUATIONS OF THE PLI

Since the PLI opened in 1999, it has collected data systematically on its graduates such as work placement history, demographic information, and the like. However, such information is purely descriptive and does not fully measure impact. With our program serving the 15th cohort, we decided to supplement the demographic data with three other impact reports in the last two years (all reports are available at **principals.berkeley.edu**):

- A 2012 report, "Breadth of Regional Impact and Strength of Model," analyzed data on the program's graduates. Based on a questionnaire about what positions they occupied, 98% worked in education, 66% in administrative positions and 24% as teacher-leaders. The report also described the PLI model.

- Cheung's "Alumni and District Partner Feedback" (2013), an example of "stakeholder evaluation," described alumni surveys and feedback from four partner districts: San Francisco, Oakland, Berkeley, and West Contra Costa. Of all respondents, 94% agreed that PLI provided strong preparation for their positions, and 100% of district partners provided positive feedback over 10 years.

- Grubb and Cheung's "Collective and Team Leadership: Preparation for Urban Schools" (2014) was based on interviews with graduates working in schools with two or more PLI alumni. It revealed many benefits of having multiple PLI graduates at one school, a practice we labeled "collective leadership," and it confirmed that many dimensions of PLI are valuable on the job.

## EXPERIMENTAL EVALUATIONS OF PLI USING STUDENT OUTCOMES

Since none of the prior impact reports focused on school level indicators, we decided to carry out an outcome evaluation despite the fact that we anticipated (and subsequently faced) many challenges. For example, data are largely limited to those available in state and district databases, especially math and English test scores, rather than the many other consequences important for students. We also collected data on suspension and truancy rates, and student perceptions measured by the California Healthy Kids Survey. Because factors other than leadership affect outcomes, our analysis required variables reflecting such factors, particularly the percentage of low-income students (measured crudely by the proportion eligible for free and reduced-price meals); teachers with full and emergency credentials; racial minority students in eight different categories; students in the Gifted and Talented Education program; migrant students; English learners, and those redesignated as English proficient; and students in special education. The data available to us, and to virtually all preparation programs, are quite limited and include little information about other dimensions of schools such as family and community partnerships, student expectations, and the like.

In considering the effects of principals, it is also crucial to confine the analysis to a specific district. Districts vary enormously in their hiring practices, as well as expectations and support for leaders. In addition, it takes time for principals to improve schools. Therefore, initially, we defined the study population to include graduates of the PLI in the Oakland Unified School District (OUSD) with at least 3 years of experience at the same school. Unfortunately, while PLI has about 500 graduates, this restricted sample included only 24 alumni from 2001 to 2013, too few for any statistical analyses. Therefore, in later parts of the analysis we used samples unrestricted by experience.

Experienced Alumni Principal Profile in OUSD

The average age of the study population was 44 years old. 60% were female alumni and 54% were people of color (46% white), with five ethnicities represented. Just under 71% of graduates were principals at elementary schools, 12.5% were in middle schools, and 26.7% were in high schools (see **Figure 1**). **PLI alumni working as administrators in OUSD are substantially more diverse that the average California administrator population (74% white) as well as the average administrator population in Alameda County (63% white)[2].**

Figure 1
RACE/ETHNICITY OF PLI ALUMNI PRINCIPALS IN OUSD

- AFRICAN AMERICAN
- LATINO
- ASIAN
- MIXED RACE
- WHITE

---
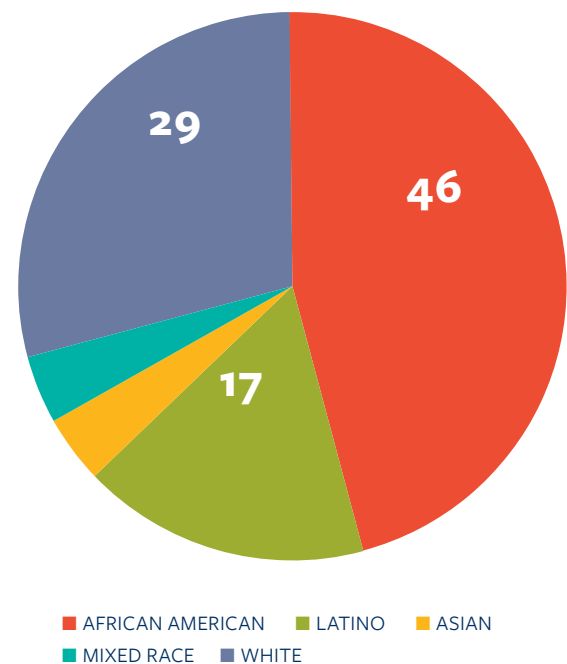
[2] Data retrieved from California Department of Education DataQuest.

We also used the Academic Performance Index (API) Similar Schools Ranking system, a system used by the California Department of Education to compare schools that serve students with similar demographics, as an additional way to describe the schools served by PLI alumni principals. Similar Schools are ranked from 1 to 10 with one representing the lowest performing. In analyzing Similar Schools ranking information, we discovered that in Oakland Unified, PLI alumni are more likely to be working in lower-ranked schools. While 47% of the schools fall into the lowest third of rankings (SIM rank 1-3), **over 68% of PLI alumni lead OUSD schools with the most vulnerable student populations (SIM rank 1-3): 42% at SIM rank 1, 18% at SIM rank 2, and another 9% at SIM rank 3.** Because of the more extreme challenges of improving lower ranked schools, we added an additional layer to our analysis. **Figure 2** illustrates the disproportionate placement of alumni in the lowest ranked schools.

Figure 2
PLI ALUMNI PLACEMENT IN OUSD BY SIMILAR SCHOOLS RANK (BY PERCENT)

- SIM RANK 1
- SIM RANK 2
- SIM RANK 3
- SIM RANK 4-10
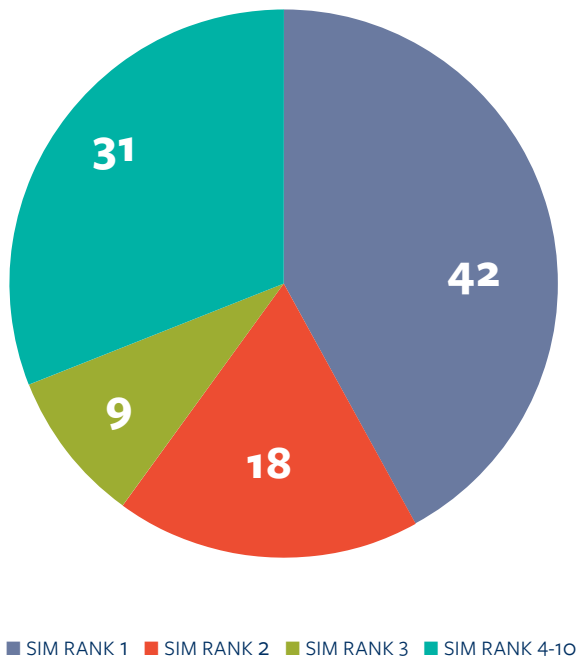
Academic Performance Index

The most widely-used test score measure in California is the Academic Performance Index (API), a re-scaling of California State Tests to a range of 200 – 1,000.

PLI-led schools had slightly higher API scores than other schools. However, higher scores may reflect students' family backgrounds, their preparation levels, or school characteristics only partially influenced by principals. To adjust for these influences, we used a statistical technique—regression analysis—to determine which other variables might influence API scores.

When we allowed for effects on test score *growth* rather than *levels*—a value-added approach—the coefficient associated with a PLI leader was positive but small. In addition, the difference was statistically insignificant, suggesting that PLI graduates may not have greater effects than graduates of other programs. Instead, scores were higher in schools with higher-income students, a higher proportion in gifted programs, lower proportions of African-Americans, Latinos, Pacific Islanders, English Learners, special education students, and students redesignated as English proficient. Finally, scores were higher in schools with more fully and emergency credentialed teachers, compared to those lacking credentials. These results are unsurprising, since they are patterns that commonly appear in most analyses of student test scores.

California Standards Tests

We analyzed the California Standards Tests (CST) by subject as well as Similar Schools Rankings. First, we collapsed the five achievement bands to three: At or above proficient, Basic, and Below Basic. Then, we analyzed the distribution of student scores. This analysis demonstrated that students in PLI-led schools had better test score distributions than those in schools with leaders from other programs. On average over 10 years*, schools with PLI alumni principals had 34% of students scoring at or above proficient while schools led by non-alumni averaged only 26% at or above proficient (+8%). Schools with PLI alumni principals also had 11% fewer students scoring below basic than their counterparts (10 year average = 37% versus 46%). **Table 1** displays the comparison of PLI alumni-led and non-alumni-led SIM Rank 1 schools.

Suspension and Truancy

Suspension and truancy rates are important because they are indicators of lost instructional time which, generally speaking, negatively impacts students' academic progress. When we analyzed these using the same value-added formulation as for test scores, roughly the same results emerged. There were no statistically significant differences between PLI-led schools and those with principals from other programs. Instead, suspension rates were higher in schools with more low-income students and more African-American students, reflecting the common finding that African-American boys are more likely to be disciplined for offenses for which other students are not. Truancy rates were higher in schools with more African-Americans, more low-income students, and more students re-designated fluent English Proficient. Unfortunately, we did not have access to other student discipline data that might have informed our analysis such as office referrals, in school detentions, and/or other alternatives to suspension.

Table 1

CST SCORES FOR SIM RANK 1 SCHOOLS WITH AND WITHOUT PLI ALUMNI PRINCIPALS IN OUSD

	PERCENT PROFICIENT OR ADVANCED	PERCENT BELOW OR FAR BELOW BASIC
PLI ALUMNI PRINCIPAL	34	37
NON PLI PRINCIPAL	26	46
DIFFERENCE	+8%	+11%

*10 year average 2002-2012 for ELA and Math Combined

California Healthy Kids Survey

The California Healthy Kids Survey (CHKS) collects information about students' perceptions of belonging at school, leadership, and home lives. Schools with negative responses are assumed to be poor places to learn. To analyze the CHKS, we focused on key topics and questions that are significantly influenced by school leadership. **Student responses to the indicators "I plan to graduate from high school" and "I know where to go for help" were much more positive for PLI-led schools in OUSD. Students in PLI-led schools also had significantly more positive responses to questions about their sense of belonging and support.**

Another revealing way to examine the CHKS data is to focus on SIM Rank 1 schools, those with the most challenging demographic compositions and students who are conventionally thought to require more support from teachers and leaders. In **Table 2** we see PLI-led schools faring better across every question, as shown by the higher proportion of students responding "very much true." The differences are particularly striking for the responses about "teachers and others who really care about students," and "teachers who listen when students have something to say."

Similarly, in **Table 3** reporting on student feelings in SIM Rank 1 schools, the responses are again more positive for PLI-led schools, with especially large differences for being happy at school, for feeling part of the school, and for feeling safe. **For example, 66% of students at PLI-led schools reported being happy at their school compared with 57% of students at non-PLI-led schools. 57% of students at PLI-led schools reported feeling safe compared to 47% of students at non-PLI-led schools.** Unlike other parts of our analysis, the differences in student responses to the CHKS in both **Table 1** and **2** are statistically significant according to Chi-square tests.

PLI emphasizes preparing graduates for improving low-performing schools that often have the common conditions of inexperienced teachers, harsh discipline policies, as well as class and racial discrimination. We like to think, therefore, that the differences in student responses are due to the preparation principals receive in PLI, though more detailed analysis would be necessary to demonstrate this.

Table 2
STUDENT PERCEPTIONS OF BELONGING AND SUPPORT, FOR SIM RANK 1 SCHOOLS

Question	PLI			
	Not at all true	A little true	Pretty much true	Very much true
There is a teacher or some other adult who really cares about me.	8.85%	27.84%	28.77%	34.53%
There is a teacher or some other adult who tells me when I do a good job.	22.37%	36.68%	5.87%	35.08%
There is a teacher or some other adult who notices when I am not there.	12.81%	28.91%	32.06%	26.22%
There is a teacher or some other adult who always wants me to do my best.	5.60%	16.28%	33.46%	44.66%
There is a teacher or some other adult who listens to me when I have something to say.	8.70%	25.36%	33.47%	32.47%
There is a teacher or some other adult who believes that I will be a success.	9.56%	20.79%	31.90%	37.75%
At school I do things that make a difference.	29.26%	34.52%	21.49%	14.73%
I plan to graduate from high school.	3.01%	4.91%	8.40%	83.68%
I know where to go for help with a problem.	5.63%	11.82%	23.07%	59.49%
	Non-PLI			
	Not at all true	A little true	Pretty much true	Very much true
There is a teacher or some other adult who really cares about me.	9.45%	27.45%	31.27%	31.83%
There is a teacher or some other adult who tells me when I do a good job.	7.03%	21.09%	35.74%	36.13%
There is a teacher or some other adult who notices when I am not there.	12.07%	25.10%	32.78%	30.05%
There is a teacher or some other adult who always wants me to do my best.	6.76%	15.98%	31.21%	46.04%
There is a teacher or some other adult who listens to me when I have something to say.	9.23%	23.86%	34.05%	32.86%
There is a teacher or some other adult who believes that I will be a success.	9.18%	19.28%	31.99%	39.56%
At school I do things that make a difference.	27.80%	33.89%	23.26%	15.05%
I plan to graduate from high school.	5.23%	5.58%	11.58%	77.61%
I know where to go for help with a problem.	7.65%	16.03%	25.14%	51.17%

Table 3
STUDENT FEELINGS ABOUT THEIR SCHOOLS, FOR SIM RANK 1 SCHOOLS

SIM Rank 1	PLI		
Question	Agree or strongly agree	Neither agree nor disagree	Strongly disagree or disagree
I feel close to people at this school.	58.78%	28.65%	12.57%
I am happy to be at this school.	66.06%	23.33%	10.60%
I feel like I am a part of this school.	63.98%	21.72%	14.30%
The teachers at this school treat students fairly.	58.94%	26.23%	14.83%
I feel safe in my school.	57.12%	25.20%	17.68%
	Non-PLI		
	Agree or strongly agree	Neither agree nor disagree	Strongly disagree or disagree
I feel close to people at this school.	53.96%	29.45%	16.60%
I am happy to be at this school.	56.88%	27.13%	15.99%
I feel like I am a part of this school.	54.79%	26.61%	18.60%
The teachers at this school treat students fairly.	53.78%	27.03%	19.19%
I feel safe in my school.	46.68%	31.34%	21.98%

CONCLUSIONS

Overall, our experiments with school level indicators found only a little evidence that PLI does a better job of preparing school leaders than other school leadership programs serving the Oakland Unified School District. There were no statistically significant differences in test scores, suspension, or truancy rates. Students in PLI-led schools had more positive conceptions of their schools as safe and supportive places (statistically significant).

The rigorous evaluation of principals and leadership preparation programs in relation to school level outcomes is highly problematic. Problems that we identified that cannot be readily resolved with existing data and methods include:

- The need to confine analysis to one district and to principals with several years of experience, reduces the sample size dramatically.

- Outcome measures are limited, usually to test scores. In most states, test score measures are complex and opaque. They are rarely vertically equated so that a point in one year means the same as a point in another year; strictly speaking they cannot be used to measure value-added. The lack of variables measuring progress through schooling is particularly serious because test scores and progress respond to different school resources.[3]

- Just as value-added measures for teachers vary from year to year[4], principal effects on outcomes may vary as students come and go, district and state policies shift, funding increases or declines, and different statistical models and test data are used. With unstable measures, judging the effectiveness of principals and leadership preparation becomes unreliable.

- Evaluations use the wrong model of leadership. Principals don't affect students directly; teachers do that through their instruction. Instead, leaders affect students indirectly, through effects on teachers, climate, the curriculum and reforms adopted (though these are restricted in many districts); through linkages to parents, external organizations, and community services; and through their success in capturing additional resources. These resources in turn affect student outcomes. The right way to describe the effects of principals is therefore to estimate two kinds of equations: one describing the influences of principals on school processes and resources, and the second analyzing the effects of these resources plus student characteristics on outcomes. But if the two equations are combined (as in all existing analyses), then all coefficients including those describing the effects of principals are certain

[3] Grubb, W. N. (2009). *The Money Myth: School Resources, Outcome, and Equity* (New York: Russell Sage Foundation).
[4] Darling-Hammond, L. et al. (2012). Evaluating Teacher Evaluation. *Phi Delta Kappan,* February 29.

to be biased. (For this demonstration, see the longer paper on which this report is based, on the **PLI website**.)

- In some schools, efforts to measure the influence of principals mistakes the nature of leadership. The conventional approach assumes that leadership is embodied in the principal only, with a top-down hierarchy. But many schools have different practices: high schools with leadership teams; those adopting distributed leadership, with responsibilities shared among principals, assistant principals, teacher leaders, and coaches; [5] or schools with collective leadership based on the similar backgrounds and values of several leaders. Evaluation must then focus on a school's leadership team or its leadership environment, as the CALL system does.[6] But existing evaluations do not reflect examples of distributed or collective leadership.

- The analysis of student outcomes can tell only whether leaders, or principal preparation programs, affect outcomes, not how they do so. Such evaluations are useful only for threatening principals with dismissal or leadership programs with closure, not for improving them. An emphasis on punishment rather than improvement cannot make schools better places for students.

Where does this leave us? Unfortunately, **we conclude that it is impossible at this stage to use student outcomes to evaluate principals and leadership programs; educators ought to move away from the narrative that leadership and preparation can be judged by test scores.** Instead we must fall back on more conventional but more informative kinds of evaluation, collecting information about practices that presumptively influence outcomes. These might include, for example, content evaluations examining a program's

courses, syllabi, and practica; satisfaction surveys asking graduates how well they have been prepared; stakeholder surveys asking teachers, district officials, and perhaps parents and students about the effectiveness of leaders. Some of these even have distinct advantages over outcome evaluations: They can more readily identify what dimensions of leadership need to be improved and whether disagreements exist among stakeholders. Their results are more comprehensible and transparent than complex statistical analyses. In the absence of much better data and methods, these traditional forms of evaluation will have to do, despite the pressures of this age of accountability.

CREDITS

For further information about how you can become engaged in the work of the Principal Leadership Institute, contact us at:

University of California, Berkeley
Principal Leadership Institute
Graduate School of Education
3653 Tolman Hall #1670
Berkeley, CA 94720-1670
(510) 643-7458
http://principals.berkeley.edu

[5] Spillane, J. (2006). *Distributed Leadership*. San Francisco: Jossey-Bass.

[6] Halvorsen, R., Kelly, C., & Shaw, J. (2014). A CALL for Improved School Leadership. *Phi Delta Kappan*, March.