Teacher Evaluation, Development, and Dismissal in California

Part 1 – The Measurement Gap: Investigating the Accuracy of Teacher Performance Evaluations

By Michael Kent, Dennis Li, and Chris Frank

Students in California's primary and secondary public schools are underachieving compared with students in other states. In the 2012 annual report published by *Education Week,* California public schools received a "D" for K-12 student achievement and ranked 35 out of the 50 states.[1]

Some analysts attribute low student achievement to the quality of California's public school teachers as they are the single most important predictor of student outcomes.[2] Numerous factors affect teacher quality in an education system; professional development and dismissal policies are two widely discussed factors that have recently come under heightened scrutiny.

Professional development allows teachers to acquire new teaching techniques to improve their instruction. The Peer Assistance and Review (PAR) programs across California, for example, provide inexperienced or underperforming teachers with guidance from more seasoned colleagues in their districts. However, due to budget constraints, California has slashed more than 90 percent of the funding for teacher professional development in the past decade.[3]

Dismissal policies are intended to facilitate the removal of ineffective teachers when they are unable to improve with professional development. Yet some observers have suggested that provisions in the current teacher dismissal policy, such as the four opportunities for unsatisfactory teachers to appeal a dismissal, have led to a prevalence of underperforming teachers in the California public school system.[4] In extreme cases, such as one that occurred at Miramonte Elementary in Los

---

1   State Report Cards. *Education Week* – Quality Counts.

2   Public Policy Institute of California. "Public Schools Serving Poor Students Are Being Shortchanged In California, Study Finds."

3   In 2002, California allocated $250 million for PAR programs; by 2012, that amount dropped to $24 million.

4   Dawson, T. C., & Billingsley, K. L. 2000. "Unsatisfactory Performance: How California's K-12 Education System Protects Mediocrity and How Teacher Quality Can Be Improved." Pacific Research Institute.

Angeles[5], critics claim current dismissal policies allow even teachers who sexually abuse students to remain in the classroom, posing great risks to students at the expense of taxpayers.

Our research attempted to better understand how current teacher development and dismissal policies affect the quality of teaching in the California public school system. Because formal evaluations generally precede the decision to further develop or dismiss teachers, we began by investigating the teacher evaluation system.

We interviewed more than two dozen school and district administrators, educators, and California Teacher Association members (representing California's largest teachers' union). We also designed and disseminated a survey to more than 10,000 school administrators throughout the state, of whom approximately 11 percent responded. The analysis presented in these policy briefs is largely based on these interactions with California

school administrators. This brief (part 1 of 3) focuses on the teacher evaluation system and the phenomenon we call "The Measurement Gap."

The Measurement Gap

During our interviews with administrators, we collected information on the teacher evaluation process. Though the frequency and methods of evaluation varied by district, the majority of interviewees indicated a binary rating system is used, i.e. teachers receive either a satisfactory or unsatisfactory rating at the end of the school year.

Throughout the course of our interviews, a large majority of administrators expressed having a higher number of underperforming teachers than the number of assigned unsatisfactory ratings. In other words, administrators claimed to provide unsatisfactory ratings only to a subset of the population of teachers they believed to be ineffective. This administrative practice allows underperforming teachers to remain in the classroom without identifying them as in need for improvement or removal—we refer to this phenomenon as the measurement gap.

The existence of such a measurement gap was corroborated by the results of

our administrator survey. In aggregate, principals in our survey reported to have 2,131 underperforming teachers out of the 16,386 total teachers represented in the sample. Thus, according to principals, an average 13.0 percent of teachers were "not effectively teaching students up to expectations" during the 2011-12 academic school year. Superintendents reported having a similar percentage of underperforming teachers in their districts, responding with an aggregate average of 12.7 percent.

Despite indicating 13 percent of their staff as underperforming, principals assigned unsatisfactory ratings to just 2.9 percent of teachers (496 out of 16,386). Figure 1 shows the measurement gap, the aggregate breakdown of underperformers versus unsatisfactory ratings.

One possible explanation for the measurement gap is that principals believe some of their underperforming teachers can be developed into quality instructors. Since assigning unsatisfactory ratings is generally perceived as the first step to dismissing an underperforming teacher, principals could choose to assign satisfactory ratings to underperforming teachers they believe can be developed.

However, our analysis does not support this theory.

---

5 In 2011, third grade teacher Mark Berndt was arrested on 23 charges of lewd conduct. Berndt objected to the district's dismissal and requested a formal hearing. Despite the legal evidence against him, the Los Angeles Unified School District decided to avoid the formal dismissal process and reached a settlement agreement with Berndt. After receiving $40,000 in back pay and legal fees, and retaining his full pension and retirement health benefits, Berndt resigned from his position.

Figure 1:
The Measurement Gap – Aggregate Underperformers and Unsatisfactory Ratings

Legend:
- Satisfactory Teachers
- Underperforming Teachers
- Underperforming Teachers Receiving Satisfactory Rating
- Underperforming Teachers Receiving Unsatisfactory Rating
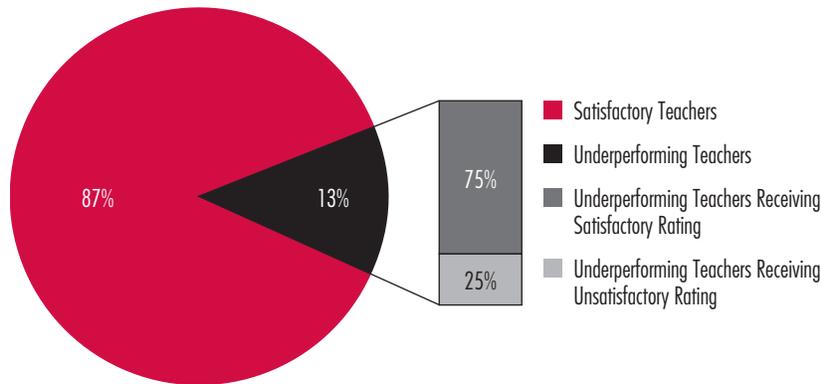
Source: Practicum Administrator Survey (2013)

Table 1:
Average Numbers of Underperforming Teachers, Teachers Principals Wish to Remove, and Unsatisfactory Ratings per School

	Underperforming	Wish to Remove	Unsatisfactory Ratings
Mean	3.51	2.98	.91

Number of Observations: 653
Source: Practicum Administrator Survey (2013)

Principals reported having an average of 3.5 underperforming teachers on staff. Of these underperformers, principals indicated that three should be removed from the classroom, suggesting that only a small subset of underperformers could benefit from professional development. Finally, when asked how many unsatisfactory ratings were given in the past year, the average response was fewer than one per principal.

Table 1 summarizes these findings. In short, principals responded that they wish to remove the majority of their underperforming teachers, but the evaluation process results in assigning unsatisfactory ratings to only about one in four teachers they deem to be underperforming.

## The Relationship Between Teacher Underperformance and Student Achievement

Our research indicates that principals do not assign unsatisfactory ratings to all the teachers they deem underperforming. To understand whether this behavior has any measurable impact on the quality of instruction, we built a model that used teacher underperformance, as indicated by principals on our survey, to predict student test scores. We then compared the accuracy of this model with another that used unsatisfactory ratings on formal evaluations to predict test scores. *Our analysis found teacher underperformance, as measured by our survey, was a better predictor of student test scores in both language arts and mathematics.* Hence, assessment of underperforming teachers by principals provides a better estimate of student achievement than the number of assigned unsatisfactory ratings they issue.

Our regression analysis shows that of two schools (with all else being equal, i.e., student demographics, local population density, percentage of teachers formally rated as unsatisfactory), the school with more underperforming teachers (as perceived by principals)
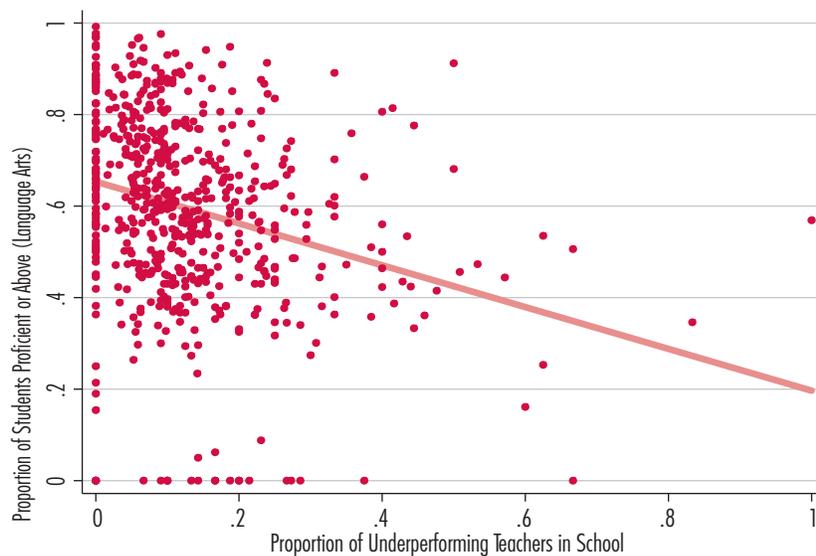
Figure 2:
Proportion of Students Proficient in Math by Proportion Underperforming Teachers in School, with Fitted Regression Line
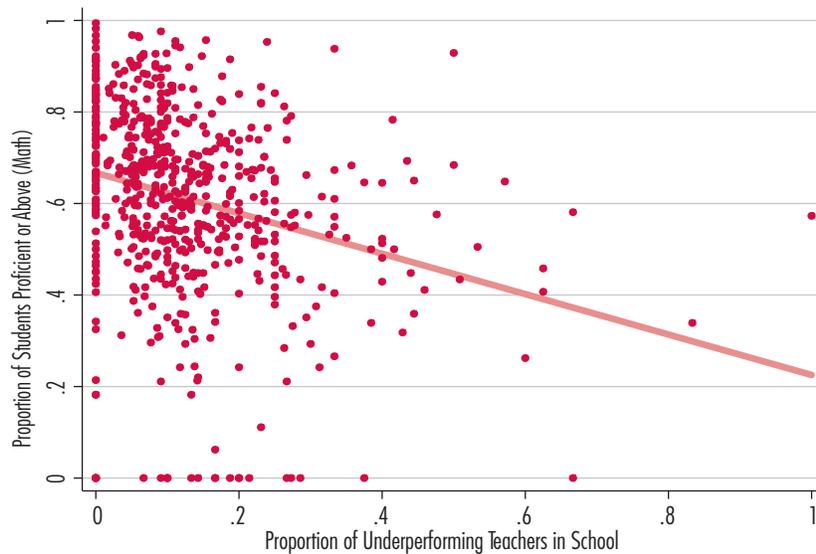
Note: Fitted regression line is based on Ordinary Least Squares regression, regressing proportion of underperforming teachers in a school on proportion of students proficient and above in Math in that school
Source: Practicum Administrator Survey (2013), Educational Results Partnership

Figure 3:
Proportion of Students Proficient in Language Arts by Proportion Underperforming Teachers in School, with Fitted Regression Line

Note: Fitted regression line is based on Ordinary Least Squares regression, regressing proportion of underperforming teachers in a school on proportion of students proficient and above in Language Arts in that school
Source: Practicum Administrator Survey (2013), Educational Results Partnership

will have lower student test scores. For example, take two elementary schools, each with 20 teachers. One school has two underperforming teachers, while the other has four. Using our models, we expect that the school with four underperforming teachers to have an approximately 1.3 percentage point decline in the proportion of all students proficient in mathematics in the whole school. Because we used school-level data in our analysis, we hypothesize that classroom-level data would yield even stronger effects of teacher underperformance. Figure 2 shows the relationship between underperforming teachers at a school with student proficiency in mathematics, while Figure 3 shows the relationship between underperforming teachers and student proficiency in language arts.

Where Do the Underperformers Teach?

Although more than 85 percent of principals in our sample reported having at least one underperforming teacher in their school, our data indicate that schools with increased percentages of socioeconomically disadvantaged students are likely to have higher proportions of underperforming

teachers. That is, there is a positive trend between the number of underperforming teachers administrators reported and the percentage of students who are socioeconomically disadvantaged (see Figure 4). This trend suggests that as the percentage of socioeconomically disadvantaged students in a school increases, so does the number of underperforming teachers.

Race characteristics were also found to be associated with the percentage of underperforming teachers in a school. Figure 5 shows that as the percentage of white students in a school increases, the average number of underperforming teachers decreases. This suggests that as the proportion of minority students in a school increases, so does the number of underperforming teachers.

Furthermore, regressions indicate the percentage of African-American students in a school is a significant predictor of teacher underperformance. That is, if the percentage of African-American students in a school increases by 10 percentage points (say 5 to 15 percent) then the proportion of underperforming teachers in that school will increase by approximately 12 percentage points *when all other*
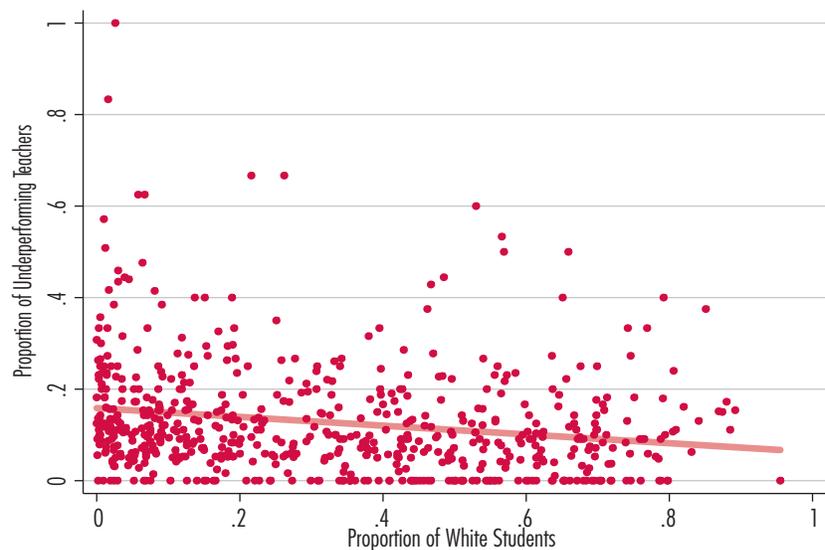
Figure 4:

Proportion Underperforming Teachers in School by Proportion Socioeconomic Disadvantaged in School, with Fitted Regression Line
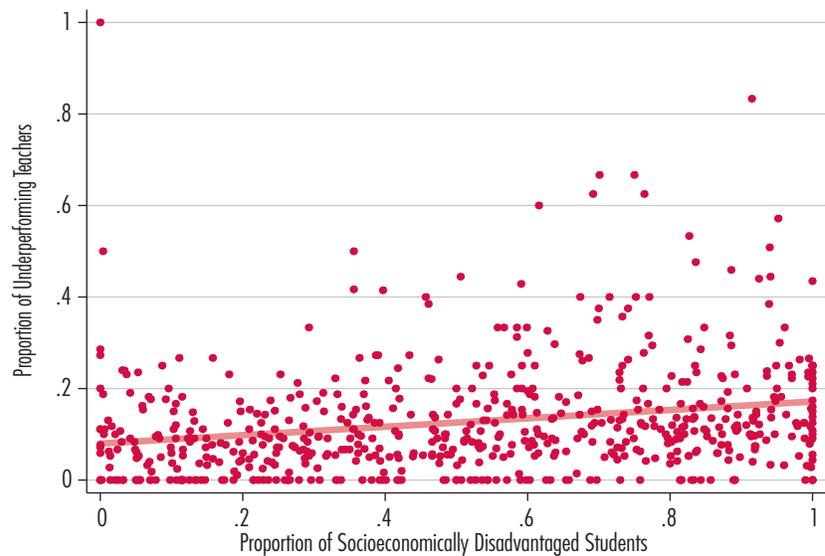
Note: Fitted regression line is based on Ordinary Least Squares regression, regressing proportion of socioeconomic disadvantaged on proportion of underperforming teachers
Source: Practicum Administrator Survey (2013), Educational Results Partnership

Figure 5:

Proportion Underperforming Teachers in School by Proportion of White Students in School, with Fitted Regression Line

Note: Fitted regression line is based on Ordinary Least Squares regression, regressing proportion of white students on proportion of underperforming teachers
Source: Practicum Administrator Survey (2013), Educational Results Partnership

demographic factors about the school are held constant. In short, our models indicate that larger proportions of African-American students while holding other variables constant will predict statistically significant increases in the proportion of underperforming teachers in schools.

Another statistically significant finding is that schools in suburban and town areas are predicted to have smaller proportions of underperforming teachers than schools in rural areas.[6] Moving from a rural school to a school in a town is expected to decrease the proportion of underperforming teachers by 24 percentage points, while moving to a suburban school decreases the proportion by 25 percentage points, holding other variables constant.

Larger schools were also predicted to have significantly

—————————
6   p < .01 and p < .05, respectively.

smaller proportions of underperforming teachers. For every 100-student increase, a school's proportion of underperforming teachers is expected to drop by 2.8 percentage points. Our results do not support the idea that socioeconomic status is the driving force behind greater proportions of underperforming teachers in schools. Instead, factors such as percent African American, population density, and school enrollment are the greatest predictors of the proportion of underperforming teachers.

Conclusion

Through qualitative interviews and quantitative survey data, our research found a disparity between the numbers of underperforming teachers and assigned unsatisfactory ratings. We refer to this as the measurement gap. This gap has implications with respect to student achievement,

socioeconomic status, and race characteristics. Our results suggest that the number of underperforming teachers, as reported by principals, can better predict student proficiency scores in language arts and mathematics when compared with formal unsatisfactory ratings. Furthermore, we find that underperforming teachers are more prevalent in lower socioeconomic status areas and correlated with particular race characteristics, such as in schools with larger proportions of African-American students.

This brief aimed to examine the extent of the measurement gap and its impact on student achievement. Our next brief will focus on the implications of accurate evaluation systems on teacher development and dismissal practices.

About The Authors

Chris Frank is a joint PhD in Sociology and Master of Public Policy student at Stanford. Before coming to Stanford, Chris worked as a market research analyst in New York after graduating with a BA in Sociology from Cornell University. Chris is the co-founder of TopCorner.org, an online platform to promote and facilitate civic engagement and community change. Upon graduation, Chris plans to begin work as a quantitative data analyst at Google.

Michael Kent is a Master of Public Policy student concentrating in Government Innovation and Technology Policy. Michael received his BA in Public Policy from Stanford University. He has had a variety of research and professional experience including at the Stanford Graduate School of Business, the Stanford Institute for Economic Policy Research (SIEPR), Booz Allen Hamilton, the Omidyar Network, and Dodge & Cox Mutual Fund. After graduation, Michael plans to work as an iShares Analyst at BlackRock in San Francisco. He hopes to volunteer for a local official and eventually work as a policy advisor and strategist.

Dennis Li is a Stanford graduate student pursuing a joint MS/MPP degree in Management Science & Engineering and Public Policy. Before coming to Stanford, Dennis worked as a Marketing Analyst for Vistaprint in Lexington, Mass., specializing in web analytics. He has completed internships in the financial services, consulting, and technology industries. Dennis received his BS in Computer Science and Operations Research from Cornell University. Upon graduation, he will work as a Graduate Fellow at Education Pioneers.

This policy brief is the first of a three-part series analyzing teacher evaluation, development, and dismissal policies in California. All three briefs analyze results of a survey disseminated to California administrators across the state. This first brief analyzes and discusses administrators' responses about the evaluation process.

SIEPR *policy brief*

SIEPR

About SIEPR

The Stanford Institute for Economic Policy Research (SIEPR) conducts research on important economic policy issues facing the United States and other countries. SIEPR's goal is to inform policymakers and to influence their decisions with long-term policy solutions.

Policy Briefs

SIEPR policy briefs are meant to inform and summarize important research by SIEPR faculty. Selecting a different economic topic each month, SIEPR will bring you up-to-date information and analysis on the issues involved.

SIEPR policy briefs reflect the views of the author. SIEPR is a non-partisan institute and does not take a stand on any issue.

For Additional Copies

Please see SIEPR website at *http://SIEPR.stanford.edu*.