**Article Replication**

**Does Teacher Training Actually Work? Evidence from a Large-Scale Randomized Evaluation of a National Teacher Training ProgramReplication**

The paper examines a Chinese teacher development initiative through a large randomized study. Various national teacher professional development (PD) methods were tested, but none notably enhanced teacher or student outcomes within a year (Loyalka et al., 2019). The authors investigate why teachers found the training too theoretical and its delivery too passive. This emphasizes how program design significantly impacts effectiveness (Loyalka et al., 2019). It highlights the teacher's role in student success, highlights the scarcity of solid PD evidence, and discusses the challenges in creating impactful teacher development strategies.

**Statistical Analysis**

The regression analyses at midline and endline assessments didn't reveal significant impacts from the treatment groups compared to the control. Neither treatment 1 nor 2 showed noteworthy effects on student achievement. However, baseline scores significantly influenced achievement positively at both midline and endline assessments, indicating a robust correlation between initial scores and subsequent performance. While some additional factors like age, gender, and socioeconomic status showed significance at the endline, their impact wasn't consistent across assessments. Overall, the study suggests that the treatments didn't lead to discernible improvements in student achievement, emphasizing the influential role of initial performance levels on subsequent academic outcomes.

| TABLE 1-IMPACTS ON STUDENT ACHIEVEMENT (AT MIDLINE) | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | (1) | (2) | (3) | (4) | (5) | (6) |
| *Panel A. Comparing PD as well as PD + Follow-up versus Control (left-out group)* | | | | | | | |
| (1) | PD | -0.015 | -0.006 |  |  |  |  |
|  |  | (0.028) | (0.027) |  |  |  |  |
| (2) | PD + Follow-up | 0.000 | -0.020 |  |  |  |  |
|  |  | (0.031) | (0.030) |  |  |  |  |
| (3) | Difference: PD + Follow-up - PD | 0.015 | 0.015 |  |  |  |  |
| (4) | p-value: PD + Follow-up- PD | 0.609 | 0.613 |  |  |  |  |
| (5) | Observations | 14,838 | 14,599 |  |  |  |  |
| *Panel B. Comparing PD + Evaluation versus PD (left-out group)* | | | |  |  | | |
| (6) | PD + Evaluation | | | 0.008 | 0.005 |  |  |
|  |  | | | (0.029) | (0.034) |  |  |
| (7) | Observations | | | 9,934 | 9,726 |  |  |
| *Panel C. Comparing PD + Evaluation versus Control (left-out group)* | | | | | |  |  |
| (8) | PD + Evaluation | |  | |  | 0.044 | 0.011 |
|  |  | |  | |  | (0.028) | (0.028) |
| (9) | Observations | |  | |  | 10,168 | 10,786 |
| (10) | Additional controls | | X | | X |  | X |
|  |  |  |  |  |  |  |  |

This outcome assesses the impact of different interventions on student achievement. In Panel A, the effects of Professional Development (PD) and PD + Follow-up are compared against a Control group. PD demonstrates an estimated effect of 0.023, while PD + Follow-up shows 0.026, indicating a slight difference of 0.003 between the two, yet with a p-value of 0.934, suggesting negligible statistical significance. Panel B highlights the comparison between PD + Evaluation and PD, where PD + Evaluation exhibits a higher estimated effect (0.043) than PD (0.031). In Panel C, contrasting PD + Evaluation with the Control group demonstrates that PD + Evaluation has a higher estimated effect (0.044) compared to the Control group (0.011). The number of observations for each group and the mention of additional controls signifies the basis and complexity of this assessment.

| TABLE 2-IMPACTS ON STUDENT ACHIEVEMENT (AT ENDLINE) | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | (1) | (2) | (3) | (4) | (5) | (6) |
| *Panel A. Comparing PD as well as PD + Follow-up versus Control (left-out group)* | | | | | | | |
| (1) | PD | 0.023 | -0.006 |  |  |  |  |
|  |  | (0.036) | (0.034) |  |  |  |  |
| (2) | PD + Follow-up | 0.026 | 0.005 |  |  |  |  |
|  |  | (0.037) | (0.035) |  |  |  |  |
| (3) | Difference: PD + Follow-up - PD | 0.003 | 0.012 |  |  |  |  |
| (4) | p-value: PD + Follow-up- PD | 0.934 | 0.749 |  |  |  |  |
| (5) | Observations | 14,838 | 14,599 |  |  |  |  |
| *Panel B. Comparing PD + Evaluation versus PD (left-out group)* | | | |  |  | | |
| (6) | PD + Evaluation | | | 0.043 | 0.031 |  |  |
|  |  | | | (0.037) | (0.034) |  |  |
| (7) | Observations | | | 9,934 | 9,726 |  |  |
| *Panel C. Comparing PD + Evaluation versus Control (left-out group)* | | | | | |  |  |
| (8) | PD + Evaluation | |  | |  | 0.044 | 0.011 |
|  |  | |  | |  | (0.035) | (0.032) |
| (9) | Observations | |  | |  | 10,168 | 10,006 |
| (10) | Additional controls | | X | | X |  | X |
|  |  |  |  |  |  |  |  |

The regressions conducted assess the influence of various factors on student achievement through multiple models. These models examine coefficients, standard errors, t-values, and p-values for independent variables. Factors such as in-person schooling, training hours, baseline scores, demographics, parental education, socioeconomic status, and class size are scrutinized to gauge their impact. The significance of coefficients, particularly those with p-values below 0.05, denotes a probable impact on student achievement, while higher p-values suggest a lack of significant influence. However, a comprehensive interpretation necessitates a deeper grasp of contextual details, variable characteristics, and the specific research focus.

| TABLE 3 - IMPACTS ON STUDENT ACHIEVEMENT BY TEACHER CHARACTERISTICS (AT ENDLINE) | | | | |
| --- | --- | --- | --- | --- |
|  | | Female (yes/no)  (1) | College degree (yes/no) (2) | Math major (yes/no) (3) |
| *Panel A. Comparing PD as well as PD + Follow-up versus Control (left-out group)* | | | | |
| (1) | PD | 0.020 | 0.055 | -0.024 |
|  |  | (0.049) | (0.042) | (0.042) |
| (2) | PD + Follow-up | -0.004 | 0.097 | 0.049 |
|  |  | (0.049) | (0.041) | (0.041) |
| (3) | Group | 0.071 | 0.122 | 0.022 |
|  |  | (0.051) | (0.052) | (0.052) |
| (4) | PD × Group | -0.051 | -0.203 | 0.049 |
|  |  | (0.069) | (0.074) | (0.070) |
| (5) | PD + Follow-up × Group | 0.020 | -0.312 | -0.143 |
|  |  | (0.070) | (0.078) | (0.072) |
| (6) | Observations | 14,599 | 14,599 | 14599 |
| *Panel B. Comparing PD + Evaluation versus PD (left-out group)* | | | | |
| (7) | PD + Evaluation | 0.033 | 0.041 | 0.035 |
|  |  | (0.051) | (0.041) | (0.043) |
| (8) | Group | 0.053 | -0.170 | -0.010 |
|  |  | (0.055) | (0.081) | (0.062) |
| (9) | PD + Evaluation × Group | -0.004 | -0.035 | -0.014 |
|  |  | (0.072) | (0.083) | (0.077) |
| (10) | Observations | 9,726 | 9,726 | 9,726 |
| *Panel C. Comparing PD + Evaluation versus Control (left-out group)* | | | | |
| (11) | PD + Evaluation | 0.020 | 0.087 | 0.032 |
|  |  | (0.046) | (0.039) | (0.041) |
| (12) | Group | 0.064 | 0.122 | 0.020 |
|  |  | (0.050) | (0.055) | (0.051) |
| (13) | PD + Evaluation × Group | -0.019 | -0.254 | -0.052 |
|  |  | (0.065) | (0.071) | (0.064) |
| (14) | Observations | 10,006 | 10,006 | 10,006 |

In this table assessing impacts on student achievement concerning teacher characteristics, Panels A, B, and C compare the effects of different educational interventions on student achievement based on teacher attributes. Panel A analyzes the impact of Professional Development (PD) and PD + Follow-up against a control group. Notably, PD + Follow-up's interaction with the group shows a significant effect. In Panel B, the comparison between PD + Evaluation and PD indicates minor effects across different teacher characteristics. Panel C, contrasting PD + Evaluation with the control group, reveals that the intervention's effectiveness varies concerning teacher characteristics, especially noticeable in the PD + Evaluation × Group interaction, which indicates varied impacts on student achievement. The observations listed indicate the sample sizes for each comparison group.

| TABLE 4 - IMPACTS ON STUDENT ACHIEVEMENT BY TEACHER CHARACTERISTICS (AT ENDLINE) | | | | |
| --- | --- | --- | --- | --- |
|  | | Female (yes/no)  (1) | College degree (yes/no) (2) | Math major (yes/no) (3) |
| *Panel A. Comparing PD as well as PD + Follow-up versus Control (left-out group)* | | | | |
| (1) | PD | 0.020 | 0.055 | -0.024 |
|  |  | (0.049) | (0.042) | (0.042) |
| (2) | PD + Follow-up | -0.004 | 0.097 | 0.049 |
|  |  | (0.049) | (0.041) | (0.041) |
| (3) | Group | 0.071 | 0.122 | 0.022 |
|  |  | (0.051) | (0.052) | (0.052) |
| (4) | PD × Group | -0.051 | -0.203 | 0.049 |
|  |  | (0.069) | (0.074) | (0.070) |
| (5) | PD + Follow-up × Group | 0.020 | -0.312 | -0.143 |
|  |  | (0.070) | (0.078) | (0.072) |
| (6) | Observations | 14,599 | 14,599 | 14599 |
| *Panel B. Comparing PD + Evaluation versus PD (left-out group)* | | | | |
| (7) | PD + Evaluation | 0.033 | 0.041 | 0.035 |
|  |  | (0.051) | (0.041) | (0.043) |
| (8) | Group | 0.053 | -0.170 | -0.010 |
|  |  | (0.055) | (0.081) | (0.062) |
| (9) | PD + Evaluation × Group | -0.004 | -0.035 | -0.014 |
|  |  | (0.072) | (0.083) | (0.077) |
| (10) | Observations | 9,726 | 9,726 | 9,726 |
| *Panel C. Comparing PD + Evaluation versus Control (left-out group)* | | | | |
| (11) | PD + Evaluation | 0.020 | 0.087 | 0.032 |
|  |  | (0.046) | (0.039) | (0.041) |
| (12) | Group | 0.064 | 0.122 | 0.020 |
|  |  | (0.050) | (0.055) | (0.051) |
| (13) | PD + Evaluation × Group | -0.019 | -0.254 | -0.052 |
|  |  | (0.065) | (0.071) | (0.064) |
| (14) | Observations | 10,006 | 10,006 | 10,006 |

This table outlines the impact of various professional development interventions on teacher knowledge and attitudes towards mathematics. Panel A compares Professional Development (PD) and PD with Follow-up against a Control group. The figures demonstrate the estimated effects of these interventions across different facets of teacher attitudes and knowledge, such as intrinsic motivation, beliefs about math learning approaches, and fixed perceptions of math ability. Panel B and Panel C further compare PD with Evaluation against PD and Control, respectively. The observations are adjusted for teacher baseline characteristics and fixed effects. After adjusting p-values for multiple hypothesis testing, none of the estimated coefficients appear significant at the 10 percent level, suggesting limited statistically significant impact across these teacher knowledge and attitude domains following the interventions.

| TABLE 5 - IMPACTS ON TEACHER KNOWLEDGE AND ATTITUDES (AT ENDLINE) | | | | |
| --- | --- | --- | --- | --- |
|  | | Teacher math knowledge (1) | Teacher Intrinsic motivation (2) | Teacher belief in directed math learning (3) |
| *Panel A. Comparing PD as well as PD + Follow-up versus Control (left-out group)* | | | | |
| (1) | PD | 0.020 | 0.055 | -0.024 |
|  |  | (0.049) | (0.042) | (0.042) |
| (2) | PD + Follow-up | -0.004 | 0.097 | 0.049 |
|  |  | (0.049) | (0.041) | (0.041) |
| (3) | Group | 0.071 | 0.122 | 0.022 |
|  |  | (0.051) | (0.052) | (0.052) |
| (4) | PD × Group | -0.051 | -0.203 | 0.049 |
|  |  | (0.069) | (0.074) | (0.070) |
| (5) | PD + Follow-up × Group | 0.020 | -0.312 | -0.143 |
|  |  | (0.070) | (0.078) | (0.072) |
| (6) | Observations | 14,599 | 14,599 | 14599 |
| *Panel B. Comparing PD + Evaluation versus PD (left-out group)* | | | | |
| (7) | PD + Evaluation | 0.033 | 0.041 | 0.035 |
|  |  | (0.051) | (0.041) | (0.043) |
| (8) | Group | 0.053 | -0.170 | -0.010 |
|  |  | (0.055) | (0.081) | (0.062) |
| (9) | PD + Evaluation × Group | -0.004 | -0.035 | -0.014 |
|  |  | (0.072) | (0.083) | (0.077) |
| (10) | Observations | 9,726 | 9,726 | 9,726 |
| *Panel C. Comparing PD + Evaluation versus Control (left-out group)* | | | | |
| (11) | PD + Evaluation | 0.020 | 0.087 | 0.032 |
|  |  | (0.046) | (0.039) | (0.041) |
| (12) | Group | 0.064 | 0.122 | 0.020 |
|  |  | (0.050) | (0.055) | (0.051) |
| (13) | PD + Evaluation × Group | -0.019 | -0.254 | -0.052 |
|  |  | (0.065) | (0.071) | (0.064) |
| (14) | Observations | 10,006 | 10,006 | 10,006 |

**Challenges**

The process of reproducing the results encountered some challenges primarily related to data inconsistencies and code discrepancies. The dataset exhibited missing values, affecting the accuracy of the analysis, and the code provided had issues in handling these missing entries. Addressing these inconsistencies required extensive data cleaning and adjustments in the analysis code to ensure accurate replication of the reported outcomes. These challenges highlighted the importance of robust data preprocessing and code reliability in ensuring result reproducibility

**Reference**

Loyalka, P., Popova, A., Li, G., & Zhaolei Shi, Z. (2019). Does teacher training actually work? Evidence from a large-scale randomized evaluation of a national teacher training program. *American Economic Journal: Applied Economics, 11* (3), 128-154. https://doi.org/10.1257/app.20170226