

## WWC Review of the Report “An Evaluation of the Chicago Teacher Advancement Program (Chicago TAP) After Four Years”<sup>1,2</sup>

The findings from this review do not reflect the full body of research evidence on Chicago TAP.

### What is this study about?

The study examined whether the Chicago Public Schools’ Teacher Advancement Program (Chicago TAP), which provides mentoring, leadership opportunities, and financial incentives to teachers, improved student academic achievement and teacher retention.

The study used two designs to answer distinct research questions. Under the first design, a randomized controlled trial, the authors examined the academic achievement of more than 7,600 students in grades 4–8 from 34 public schools in Chicago. In the spring of 2007 and again in the spring of 2009, groups of schools were randomly assigned either to participate in Chicago TAP during the coming school year or to serve as a comparison group for a year and participate in Chicago TAP during the following school year.

The effect of Chicago TAP on academic achievement after one year of implementation was estimated by comparing the spring math, reading, and science achievement of students in Chicago TAP schools to the achievement of students in schools that had not yet implemented the program.

Using the second design, a quasi-experiment, the study examined teachers’ retention rates, defined as remaining in the same school from year to year. The effect of Chicago TAP on teacher retention was assessed by comparing the retention of teachers in Chicago TAP schools with the retention of a matched sample of teachers in non-TAP Chicago public schools (sample sizes varied across years).

### Features of the Chicago Public Schools’ Teacher Advancement Program (Chicago TAP)

Chicago TAP is a local adaptation of the Teacher Advancement Program (TAP), a schoolwide reform that has been implemented in more than 200 schools nationwide. TAP provides annual performance bonuses to teachers based on a combination of their value added to student achievement and observations of their classroom teaching. High-performing teachers can earn additional bonuses by serving in mentor or master teacher positions, which include salary increases of \$7,000 and \$15,000, respectively.

The Chicago TAP model includes weekly meetings of teachers and mentors. It also includes regular observations of teachers’ classrooms and instructional delivery by a school leadership team. Unlike the national model, teacher value added is not measured for individual teachers, but for teachers in the same school and grade. Also unlike the national model, Chicago TAP offers performance bonuses to principals and other staff.

The first three cohorts of Chicago TAP teachers received an average bonus of \$1,100 in the first year the school implemented the program. The fourth cohort received an average first-year bonus of \$1,400. Across all cohorts, average bonuses increased to approximately \$2,500 in the second and third years of Chicago TAP implementation, and were \$1,900 in the fourth year of implementation.

### WWC Rating of the Analysis of Student Academic Achievement

#### ***The analysis of student academic achievement meets WWC evidence standards with reservations***

**Strengths:** The analysis was based on a randomized controlled trial.

**Cautions:** Because the authors were not able to identify the students who were enrolled in study schools at the time of random assignment, the study cannot receive the highest WWC rating of meets standards without reservations. However, the authors were able to demonstrate the equivalence of the analytic samples, which is sufficient to meet standards with reservations.

In addition, the analysis includes students who enrolled in Chicago TAP and comparison schools after random assignment had been conducted. Therefore, the estimated effects on student achievement could reflect both the effect of the intervention on students who were exposed to it and changes in the composition of the student body.

### **What did the study find about student achievement?**

After one year of implementation, students attending Chicago TAP schools did not score significantly differently in math, reading, or science achievement, as measured by the Illinois Standards Achievement Test (ISAT), than students attending comparison schools.

### WWC Rating of the Analysis of Teacher Retention

#### ***The analysis of teacher retention meets WWC evidence standards with reservations***

**Strengths:** Schools that participated in Chicago TAP were well-matched with comparison schools on a number of demographic and academic characteristics.

**Cautions:** Although the study matched Chicago TAP schools to comparison schools in the district based on several observable characteristics, it is possible that there were other differences between the two groups that were not accounted for in the analysis; these differences could have influenced teacher retention rates.

### **What did the study find about teacher retention?**

Sixty-seven percent of teachers who were employed in schools that first implemented Chicago TAP in the fall of 2007 were still teaching in the same school in the fall of 2010. In contrast, 56% of teachers employed in non-TAP public schools were retained during the same period. This 12 percentage point difference in three-year teacher retention rates between the original cohort of Chicago TAP and non-TAP schools was statistically significant.

However, there were no statistically significant differences in teacher retention rates between Chicago TAP schools and comparison schools after one year (among three cohorts of schools, fall 2009–fall 2010) or two years of implementation (among two cohorts of schools, fall 2008–fall 2010).

### Appendix A: Study details

Glazerman, S., & Seifullah, A. (2012). *An evaluation of the Chicago Teacher Advancement Program (Chicago TAP) after four years. Report prepared for The Joyce Foundation. Washington, DC: Mathematica Policy Research.*

- Setting** The study was conducted in the Chicago Public Schools starting in the 2007–08 school year and continued through the 2010–11 school year.
- Study sample** A total of 34 public elementary schools in Chicago participated in the randomized controlled trial part of the study.<sup>3</sup> More than 90% of the students in these schools were African American, and more than 95% were eligible for free or reduced-price lunch.
- In the spring of 2007, 16 elementary schools were randomly assigned to begin Chicago TAP in fall 2007 (eight schools in the Chicago TAP group [Cohort 1]) or in fall 2008 (eight schools in the comparison group [Cohort 2]). In spring 2009, 18 additional elementary schools were randomly assigned to begin Chicago TAP in fall 2009 (nine schools in the Chicago TAP group [Cohort 3]) or in fall 2010 (nine schools in the comparison group [Cohort 4]).
- Students in grades 4–8 were included in the analysis of student achievement: 7,661 students in the reading analysis sample (3,717 Chicago TAP students and 3,944 comparison students); 7,656 students in the math analysis sample (3,714 Chicago TAP students and 3,942 comparison students); and 1,717 students in the science analysis sample (808 Chicago TAP students and 909 comparison students), which is smaller than the others because standardized test data in science were only collected for students in grades 4 and 7 in two cohorts.
- For the analysis of teacher retention, the 34 Chicago TAP schools were matched to other schools in the district that did not participate in Chicago TAP during the study period on measures such as school size, teacher retention, student race/ethnicity, student achievement, student poverty, student special education status, student language proficiency, and charter school status. The authors used a propensity score matching procedure where TAP schools were matched to their nearest five neighbors, with replacement. Altogether, the teacher retention sample as of fall 2010 included 612 Chicago TAP teachers in 21 schools and 2,082 comparison teachers in 77 schools after one year; 370 Chicago TAP teachers in 12 schools and 1,509 comparison teachers in 51 schools after two years; and 166 Chicago TAP teachers in five schools and 615 comparison teachers in 20 schools after three years.
- Intervention group** Under TAP, teachers can earn extra pay and responsibilities through promotion to mentor or master teacher and can earn annual performance bonuses based on a combination of their value added to student achievement and observations of their classroom teaching. The Chicago TAP model includes weekly meetings of teachers and mentors, regular classroom observations by a school leadership team, and pay for principals who meet implementation benchmarks. In the first year of implementation, teachers in Cohorts 1, 2, and 3 (i.e., those implementing Chicago TAP in 2007–08 through 2009–10) received an average bonus of \$1,100; teachers in Cohort 4 received an average bonus of \$1,400 in 2010–11. Average bonuses increased to approximately \$2,500 in the second and third years of implementation, and were \$1,900 in the fourth year of implementation. Teachers and mentors met weekly, and mentors received an additional \$7,000 per year. Master teachers received \$15,000.

### **Comparison group**

For the randomized controlled trial portion of the study, comparison schools were in a business-as-usual condition for a year and subsequently participated in Chicago TAP. For the quasi-experimental portion of the study, comparison schools were in a business-as-usual condition and did not receive Chicago TAP at any point during the study period.

### **Outcomes and measurement**

Standardized test data on student achievement were obtained from the Chicago Public Schools, including scores on three parts of the Illinois Standards Achievement Test: Reading (grades 4–8), Math (grades 4–8), and Science (grades 4 and 7). Teacher retention was defined as remaining in the same school from year to year, and was measured at one, two, and three years after Chicago TAP implementation. For a more detailed description of these outcome measures, see Appendix B.

### **Support for implementation**

The Chicago TAP model provides for observations of teachers by the principal, lead teachers, and mentor teachers, all of whom undergo training and certification in using the skills, knowledge, and responsibilities (SKR) rubric. SKR scores are based on observed classroom performance in four domains: designing and planning instruction, learning environment, instruction, and responsibilities.

### **Reason for review**

This study was identified for review by the WWC by receiving significant media attention.

### Appendix B: Outcome measures for each domain

Math achievement	
<i>Illinois Standards Achievement Test (ISAT): Mathematics Assessment</i>	The ISAT Mathematics Assessment is a standardized statewide test administered to students in grades 3–8. Assessment scores were obtained from the Chicago Public Schools (CPS).
Reading achievement	
<i>ISAT: Reading Assessment</i>	The ISAT Reading Assessment is a standardized statewide test administered to students in grades 3–8. Assessment scores were obtained from the CPS.
Science achievement	
<i>ISAT: Science Assessment</i>	The ISAT Science Assessment is a standardized statewide test administered to students in grades 4 and 7. Assessment scores were obtained from the CPS.
Teacher retention	
<i>Teacher retention</i>	Retention was defined as remaining in the same school between the fall of the baseline year and the fall of follow-up years (one, two, and three years post-implementation). The one-year retention rate included Cohorts 1–3 (i.e., schools that began Chicago TAP implementation in the fall of 2007, 2008, or 2009 and their comparison schools), and measures whether teachers remained in the same school after one academic year. The two-year retention rate included Cohorts 1 and 2 (i.e., schools that began Chicago TAP implementation in the fall of 2007 or 2008 and their comparison schools), and measures whether teachers remained in the same school after two years. The three-year retention rate was calculated only for Cohort 1 (i.e., schools that began Chicago TAP implementation in the fall of 2007 and their comparison schools), and measures whether teachers remained in the same school after three years, from fall 2007 to fall 2010. The authors calculated teacher retention rates using CPS administrative data on the employment status of teachers.

Appendix C: Study findings for each domain

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
<b>Math achievement</b>								
<i>ISAT: Mathematics Assessment</i>	Grade 4–8 students after one year of Chicago TAP	34 schools/ 7,656 students	233.4 (25.1)	234.3 (28.8)	–0.9	–0.03	–1	> 0.10
<b>Reading achievement</b>								
<i>ISAT: Reading Assessment</i>	Grade 4–8 students after one year of Chicago TAP	34 schools/ 7,661 students	221.3 (26.5)	221.0 (27.0)	0.3	0.01	0	> 0.10
<b>Science achievement</b>								
<i>ISAT: Science Assessment</i>	Grade 4 and 7 students after one year of Chicago TAP	34 schools/ 1,717 students	204.3 (31.0)	200.6 (31.0)	3.7	0.12	+5	> 0.10
<b>Teacher retention</b>								
<i>Teacher retention rate</i>	one year (fall 2009–fall 2010)	98 schools/ 2,694 teachers	0.81	0.81	0.00	0.01	0	> 0.10
<i>Teacher retention rate</i>	two years (fall 2008–fall 2010)	63 schools/ 1,879 teachers	0.71	0.68	0.03	0.09	+4	> 0.10
<i>Teacher retention rate</i>	three years (fall 2007–fall 2010)	25 schools/ 781 teachers	0.67	0.56	0.12	0.30	+12	< 0.01

**Table Notes:** For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. Means and standard deviations were provided by the author after an inquiry by the WWC; the mean for science achievement presented here is a correction of an error in Table IV.1 of the report. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the change (measured in standard deviations) in an average student’s outcome that can be expected if the student is given the intervention. The improvement index is an alternate presentation of the effect size, reflecting the change in an average student’s percentile rank that can be expected if the student is given the intervention. The WWC did not compute average effect sizes for the three retention outcomes because they were measured with different samples at different time periods and are therefore considered to be in separate domains. ISAT = Illinois Standards Achievement Test.

The study is characterized as having indeterminate effects on mathematics, reading, and science achievement, since none of the effects in these domains were statistically significant or substantively important. The study is characterized as having a statistically significant positive effect for teacher retention because univariate statistical tests are reported for each outcome measure, the effect for at least one measure within the domain is positive and statistically significant, and no effects are negative and statistically significant.

**Study Notes:** No corrections for clustering or multiple comparisons were needed. The p-values reported here were reported in the original study

Appendix D: Supplemental findings by domain

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
<b>Math achievement: quasi-experimental design</b>								
<i>ISAT: Mathematics Assessment</i>	Grade 4–8 students after one year of Chicago TAP	41,580 students	235.9 (29.9)	235.5 (29.1)	0.4	0.01	+1	> 0.10
<b>Reading achievement: quasi-experimental design</b>								
<i>ISAT: Reading Assessment</i>	Grade 4–8 students after one year of Chicago TAP	41,580 students	222.0 (26.6)	222.4 (26.5)	–0.4	–0.02	–1	> 0.10

**Table Notes:** The results presented above are from a quasi-experimental analysis of academic achievement, which included 8,097 students in Chicago TAP schools and 33,483 students in comparison schools. The study authors did not specify the number of schools included in the analysis. Because this analysis was conducted to validate the results of the randomized controlled trial, results are presented as supplementary findings. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. Means and standard deviations were provided by the author after an inquiry by the WWC. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the change (measured in standard deviations) in an average student’s outcome that can be expected if the student is given the intervention. The improvement index is an alternate presentation of the effect size, reflecting the change in an average student’s percentile rank that can be expected if the student is given the intervention. ISAT = Illinois Standards Achievement Test.

**Study Notes:** No corrections for clustering or multiple comparisons were needed. The p-values presented here were reported in the original study. Results for the quasi-experimental study of science achievement did not meet WWC evidence standards because the analytic sample exhibited baseline differences in reading and mathematics achievement that required a statistical adjustment that the authors did not perform. Therefore, the results of this analysis are not reported.

### Endnotes

<sup>1</sup> Single study reviews examine evidence published in a study (supplemented, if necessary, by information obtained directly from the author[s]) to assess whether the study design meets WWC evidence standards. The review reports the WWC's assessment of whether the study meets WWC evidence standards and summarizes the study findings following WWC conventions for reporting evidence on effectiveness. This study was reviewed using the single study review protocol, version 2.0. A quick review of this study was released on April 9, 2012, and this report is the follow-up review that replaces that initial assessment.

<sup>2</sup> Absence of conflict of interest: This study was conducted by staff from Mathematica Policy Research. Because Mathematica operates the WWC, this study was reviewed by staff from subcontractor organizations.

<sup>3</sup> A separate analysis of academic achievement was conducted on the quasi-experimental sample, which included 8,097 students in Chicago TAP schools and 33,483 students in comparison schools. Because this analysis was conducted to validate the results of the randomized controlled trial study, results are presented as supplementary findings in Appendix D.

### Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, February). *WWC review of the report: An evaluation of the Chicago Teacher Advancement Program (Chicago TAP) after four years*. Retrieved from <http://whatworks.ed.gov>.



### Glossary of Terms

<b>Attrition</b>	Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study.
<b>Clustering adjustment</b>	If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary.
<b>Confounding factor</b>	A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.
<b>Design</b>	The design of a study is the method by which intervention and comparison groups were assigned.
<b>Domain</b>	A domain is a group of closely related outcomes.
<b>Effect size</b>	The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.
<b>Eligibility</b>	A study is eligible for review if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.
<b>Equivalence</b>	A demonstration that the analysis sample groups are similar on observed characteristics defined in the review area protocol.
<b>Improvement index</b>	Along a percentile distribution of students, the improvement index represents the gain or loss of the average student due to the intervention. As the average student starts at the 50th percentile, the measure ranges from -50 to +50.
<b>Multiple comparison adjustment</b>	When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.
<b>Quasi-experimental design (QED)</b>	A quasi-experimental design (QED) is a research design in which subjects are assigned to intervention and comparison groups through a process that is not random.
<b>Randomized controlled trial (RCT)</b>	A randomized controlled trial (RCT) is an experiment in which investigators randomly assign eligible participants into intervention and comparison groups.
<b>Single-case design (SCD)</b>	A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.
<b>Standard deviation</b>	The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample are spread out over a large range of values.
<b>Statistical significance</b>	Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ( $p < 0.05$ ).
<b>Substantively important</b>	A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

Please see the [WWC Procedures and Standards Handbook \(version 2.1\)](#) for additional details.