NIET | NATIONAL INSTITUTE FOR EXCELLENCE IN TEACHING

# Comprehensive Educator Effectiveness Models That Work: Impact of the TAP System on Student Achievement in Louisiana

Joshua H. Barnett, Kellie C. Wills, & Peggy C. Kirby

November 2014

# Table of Contents



# Table of Tables



# Table of Figures

# Executive Summary

TAP™: The System for Teacher and Student Advancement was launched in 1999 as a comprehensive educator effectiveness model to offer career advancement and leadership opportunities for educators, a fair and transparent evaluation process linked to job-embedded professional development and performance-based compensation, which culminate in improved instructional practices and student achievement.

This study illustrates the impact of the TAP System in 66 Louisiana schools. These are primarily high-need schools, with average free/reduced price lunch eligibility of 86%, impacting more than 32,000 students each year. We report evidence of a positive effect of the TAP System on student achievement using two complementary analysis strategies:

- We found a significant positive effect of TAP on 2012-13 K-8 Assessment Index scores, controlling for previous achievement, percentage of students receiving free or reduced price lunch, school configuration, school size (number of students), and percentage of English language learners. The 66 TAP schools scored 3.7 points higher on average than non-TAP schools statewide.
- We also found that the TAP schools had significantly greater 2012-13 Assessment Index scores than a group of matched control schools. The TAP schools scored 5.5 points higher on average than their matched controls.

The application of propensity score matched-control methodology to this large group of Louisiana schools provides robust evidence of significantly increased achievement growth in TAP schools.

# Introduction

America's education system has transitioned from international leader to former star. For years, students in U.S. schools have, on average, dropped relative to emerging leaders across the globe. Following the most recent performance trend on the Programme for International Student Assessment (PISA) rankings put forth from the Organization for Economic Cooperation and Development (OECD), international rankings continue to place American students near the middle of the pack, slowly losing ground to other leading nations. In fact, the average 15-year-old American student dropped in math (from 25$^{th}$ to 27$^{th}$), reading (from 14$^{th}$ to 17$^{th}$), and science (from 17$^{th}$ to 20$^{th}$) between the 2009 and 2012 PISA administrations (OECD, 2011; OECD, 2012). As this stark reality has settled over the American public education system, researchers, policymakers, and practitioners have sought a host of reforms to reverse this trend.

Many researchers and policymakers posit improving the quality of the nation's teaching workforce is the best intervention for raising student achievement and providing opportunities to all students (e.g. Darling-Hammond, 2006; Goldhaber, 2002; Hanushek, 2013; Rivkin, Hanushek, & Kain, 2005). Notwithstanding the knowledge of the importance of the teacher, wide-scale reforms aimed at teachers are historically few. In a review of education reforms proposed from 1987 through 1997, less than 1 percent of them focused on improving teacher quality (Carpenter, 2000). Since the turn of the century, new efforts have focused on improving teacher quality from alternative pathways for recruiting teachers through myriad professional development reforms. However, despite the near consensus over the impact of the teacher on students' achievement and overall success, sincere concerns with developing, recruiting and retaining effective teachers remain (Clotfelter, Ladd, & Vigdor, 2011; Grissom, 2011; Gordon, Kane, & Staiger, 2006; Ingersoll, 2004; Ingersoll & Merrill, 2012; Johnson, Kraft, & Papay, 2012; Rockoff, 2004; Rowan, Correnti, & Miller, 2002; TNTP, 2012). Given the known need for improved teacher quality (Rowan, Correnti, & Miller, 2002; Haycock, 1998; Ingersoll & Merrill, 2012; Thum, 2003); therefore, continued research investigating efforts to attract, retain, and develop teachers with a clear focus on how they impact student achievement is needed.

One model focused on building comprehensive educator effectiveness is TAP$^{TM}$: The System for Teacher and Student Advancement. The theory of action for the TAP System is that ongoing applied professional development will be created and delivered via local teacher leaders, who serve in TAP schools as master and mentor teachers. TAP schools also utilize a rigorous rubric of evaluation and performance-based compensation, which coupled with the multiple career roles and ongoing applied professional development lead to improved educator effectiveness and improved student achievement. To determine the impact of the TAP System on schools, educators, and students across the participating schools, the authors employ an impact evaluation perspective for measuring and monitoring program outcomes put forward by Rossi, Lipsey, and Freeman (2004). Similarly, the authors follow the framework of other groups investigating the impact of large-scale interventions (Glazerman, Chiang, Wellington, Constantine, & Player, 2011).

Prior research examining the TAP System has demonstrated a consistent pattern of improving the instructional ability of educators and increasing student achievement (Algiers Charter School Association, 2011; Barnett, Rinthapol, & Hudgens, 2014; Buck & Coffelt, 2013; Daley & Kim 2010; Hudson, 2010; Schacter et al., 2004; Schacter & Thum, 2005; Solmon, White, Cohen, & Woo, 2007). A

recent third-party study examined TAP System impact on student achievement and teacher practices in 17 Louisiana TAP schools, demonstrating that their test scores grew more rapidly than those of matched control schools over a four-year period (Mann, Leutscher, and Reardon, 2013). The current study expands the propensity score matched-control methodology to a much larger group of 66 Louisiana TAP schools.

## Review of Evidence and Core Elements of the TAP System

TAP™: The System for Teacher and Student Advancement is managed and supported by the National Institute for Excellence in Teaching (NIET), a 501(c)(3) non-profit organization. Introduced by Lowell Milken in 1999, the TAP System is a comprehensive, research-driven educator effectiveness model committed to ensuring a highly skilled, strongly motivated, and competitively compensated teacher for every classroom in America (www.niet.org). Since 1999, the TAP System has become the leading comprehensive educator effectiveness model that offers career advancement and leadership opportunities for educators, a fair and transparent evaluation process that is linked to job-embedded professional development, and performance-based compensation. In addition to schools and districts implementing the full TAP System, a number of states are using TAP System elements – in particular, the evaluation rubric and process. As of the 2014-15 school year, NIET initiatives are impacting over 200,000 teachers and more than 2.5 million students, approximately 5 percent of the entire American student population. As such, the TAP System and research examining it is important. Each of the four core elements within the TAP System is discussed below and situated within the literature base supporting each element.

### Multiple career paths

One of the prevailing descriptions of the education system used to be an "egg carton," where each teacher would go into his or her classroom and close the door. That is, each egg, or teacher, would work independently with virtually no need or care for the others around them, and simply be surrounded by other teachers in classrooms nearby (Tyack & Cuban, 1995). The extension of this idea is that some of those eggs around the teachers might be cracked and broken, or whole and intact; regardless of their status, little to no recognition was given to those around.

This type of organizational structure afforded first year teachers the opportunity to work beside 10-, 20-, 40-year veteran teachers, primarily because all teachers held the same position – that is, the profession was "flat" (Elmore, 2000; Kardos, Johnson, Peske, Kauffman, & Liu, 2001; Rowan, 1990). This model of organization also led to high levels of autonomy for teachers; however, relatively low levels of support, encouragement, and collaboration (Bezzina, 2006; Elmore, 2002; Trorey & Cullingford, 2002). In contrast to this former approach, the last two decades have given rise to the implementation of distributed leadership, which Timperley (2005) and others (e.g. Copland, 2003; Gronn & Rawlings-Sanaei, 2003; Lambert, 1998) describe as a replacement for the old model of thinking that required "a single 'heroic' leader standing atop a hierarchy, bending the school community to his or her purpose" (Camburn, Rowan, & Taylor, 2003, p. 348). The movement of distributed leadership in schools focuses on leveraging talent in addition to the role of the principal, where lead teachers emerge to help guide the development of the school as well (i.e. grade-level leaders, content leaders, instructional leaders).

In response to the flat structure found in many schools, TAP schools facilitate skilled teachers serving as

master and mentor teachers, where they receive additional compensation for providing high levels of support to career teachers. Master and mentor teachers form a leadership team, along with the principal, to deliver school-based support and conduct evaluations with a high level of expertise.

This element of TAP leverages the contributions of highly skilled teachers who want to advance their careers and earnings potential without becoming school or district administrators. The TAP System provides such individuals with the opportunity to become master or mentor teachers, largely responsible along with their principals for the evaluation and professional growth elements of TAP in their schools. Simultaneously, this path provides the school with a cadre of trained professionals to carry out classroom observations, mentor teachers using feedback from observations, field-test instructional strategies in the local context of the school, and lead cluster groups in ongoing professional growth using those strategies. This structure allows teachers to work together to increase the local human capital at each school site (Berry, Daughtrey, & Wieder, 2010; Leithwood & Mascall, 2008).

## Ongoing applied professional growth

Too often, educators describe professional development that provides little interest, engagement or relationship to their actual work (Bezzina, 2006). Teacher-led, locally developed professional development was crystalized in 2001 with Michael Fullan's work, *The New Meaning of Educational Change*, where he advocated that effective development must facilitate networking, reculturing, and restructuring the school. While the theory of effective professional development was being articulated by Fullan and others, the TAP System's teacher-led professional development was being forged in schools.

TAP teachers participate in weekly cluster group meetings, led by master and mentor teachers, in which they examine student data, engage in collaborative planning and learn strategies that have been field-tested in their schools. Professional development continues into each classroom as master teachers model lessons, observe classroom instruction, and support other teachers to improve their teaching. Through this approach to development, teachers work continuously with one another to increase application and understanding (Jackson & Bruegmann, 2009; Murray, Ma, & Mazur, 2009). This ongoing, job-embedded, and contextually relevant approach is designed to improve instructional quality as part of the regular school culture. Classroom evaluation data and student growth scores guide leaders in the planning of collaborative professional development directed toward the specific needs of the teachers in the school.

## Instructionally focused accountability

In order to improve the quality of classroom instruction, an assessment of the instruction is vital. Such assessment is also essential if teachers are to be held accountable for their work and for professional improvement. Traditional school systems have not been successful at measuring and assessing classroom instruction. TNTP published a revealing report in 2009 demonstrating most schools fail to evaluate their teachers in any meaningful way. As seen in repeated figures from their study, most teachers in America's schools were rated at the very highest levels, despite the fact that most schools were not educating their students at these highest levels (Weisberg, Sexton, Mulhern, & Keeling, 2009).

Five years later, a 2014 report from the National Council on Teacher Quality indicated that while improvements in educator evaluation have occurred since the 2009 report, only approximately half of

all states require every teacher to be evaluated every year. Also, less than one-third of all states have a process in place to provide certification to teacher evaluators. The report further explains:

- 28 states require all teachers to be evaluated
- 25 states require multiple observations of teachers
- 22 states require evaluation feedback be given to teachers
- 15 states require unannounced observations
- 13 states have a process in place for certifying evaluation trainers

Given that differences in teacher effectiveness represent the single most important school-related factor affecting student learning, accurately measuring differences in teacher performance is critical to the improvement of teaching and learning (NCTQ, 2014).

In contrast to traditional evaluation methods noted above, TAP has developed a comprehensive approach to teacher evaluation and incentives that depend on multiple measures of both teaching practice and teaching outcomes. The TAP System provides differentiated feedback for teacher improvement, in contrast to the consistently high ratings found in many status quo evaluation systems (Weisberg et al., 2009). TAP teachers are observed in classroom instruction multiple times a year by multiple trained observers, including principals and master and mentor teachers, using research-based rubrics covering multiple dimensions of instructional quality. Evaluators are trained and certified, and leadership teams monitor the reliability and consistency of evaluations in their schools. This evaluation process generates rapid-feedback data that school personnel can use in decision-making during the school year rather than only summative feedback reporting on the prior year.

In addition, student achievement growth results provide outcome-based measurements of school and teacher performance. These indicators are available on an annual cycle rather than a rapid-feedback cycle, providing a complement to classroom observations in a multiple-measure evaluation system. NIET provides support to school personnel in how to monitor and utilize both kinds of data to inform school goals and individual professional growth planning. The rigor and balance of the TAP System evaluation is predicated on trained individuals using high-quality, research-based rubrics multiple times per year. The TAP System evaluation also incorporates the use of announced and unannounced evaluations, where announced evaluations include a pre-conference discussion. All observations include a post-conference meeting.

## Performance-based compensation

Approximately 95 percent of K-12 teachers in the U.S. work in a school or district with a salary schedule that provides pay largely on years of experience and number of degrees attained (Podgursky & Springer, 2007). This "single-salary schedule" was developed explicitly to enhance equity for teachers in the 1920's. Prior to that time, teachers were paid differentially based on their positions in the schools, which resulted in elementary teachers being paid less than secondary teachers and often women and minority teachers being paid less than their male peers. To rectify this issue, the single-salary schedule was designed to pay the same salary to teachers with the same qualifications regardless of race, gender, or grade level taught. In this way, the justification for paying differential salary amounts were objective, measurable, and not subject to administrative whim (Ritter & Barnett, 2013). The usage of the single-salary schedule persists despite research showing relatively little impact for qualifications such as

traditional certification, long-term experience, and advanced degrees (Goldhaber & Brewer, 2000; Gordon, Kane, & Staiger, 2006; Ritter & Barnett, 2013).

Teachers in TAP schools have the opportunity to earn bonuses each year based on their observed skills, knowledge, and responsibilities (SKR scores in TAP), their students' average growth in achievement, and the entire school's average growth in achievement. Master and mentor teachers also receive additional compensation based on their added roles and responsibilities. Combining these sources, performance pay for a teacher in a TAP school can be a significant award (i.e. $15,000) in any one year. This amount is not a permanent increase, but a one-year-at-a-time reward for effectiveness. The award is based on measured performance on multiple dimensions and is integrated and aligned with support for professional growth.

In TAP System schools, there are multiple measures of performance during the school year which provide feedback for teachers over the course of the year as well as a summative rating at the end of the year. Performance-based compensation results from a weighted value of two or three categories of measures. Teachers for whom a classroom level growth, or value-added, score can be calculated are generally compensated with 50 percent based on classroom observation scores (averaged score of multiple observations of their practice), 30 percent classroom-based student growth, and 20 percent on schoolwide student growth. Teachers who do not have a classroom based value are compensated with 50 percent based on SKR scores and 50 percent based on schoolwide growth scores. Each of these score components is discussed below.

**Skills, Knowledge, and Responsibilities Scores.** Teachers are observed multiple times a year (3-4 times) by multiple certified raters (i.e. school principal, master teacher, mentor teacher) on 19 indicators of instructional practice. Observation scores are combined through a weighted average with additional responsibility indicators to create an overall Skills, Knowledge, and Responsibilities (SKR) score for each teacher. "Skills" and "Knowledge" scores are comprised of a multi-tiered percentage breakdown of various research-based pedagogical best practice components related to instructional delivery. Each of these are clearly defined and outlined in the TAP rubric domains (Instruction, Designing and Planning Instruction, and the Learning Environment). "Responsibility" scores are comprised of a teacher's response and impact on staff development, instructional supervision, and mentoring if the teacher is a master teacher or mentor teacher. "Responsibility" scores also include community involvement, school responsibilities, growing and developing professionally, and level of reflection on teaching for all. The weighted scores from each indicator are combined to create an overall TAP SKR score, which ranges from a 1.0 (unsatisfactory performance) to 5.0 (exemplary performance) in half-point increments.

**Classroom Growth/Value-Added Scores.** Classroom value-added scores are the achievement growth of a teacher's or school's students during a school year. A student's test scores are matched to his or her own prior scores and compared to similar performing students to measure the student's progress (or growth) during the year. Scores are calculated and converted from the original test metric by a third-party vendor (TAP schools in different states work with different agencies using their local measures to obtain teacher and school growth scores) into a 1 to 5 scale (in whole numbers) indicating how the teacher's average student growth compares to the average student growth for teachers of the same subject with similar students. Scores on the scale range from 1 (much less than a year's growth) to 5 (much more than a year's growth), with a 3 representing one year's growth and a 2 and 4 representing

less and more than a year's growth respectively.

**School Growth/Value-Added Scores.** School value-added is based on the achievement growth of all students in a school. Similar to classroom value-added, this growth is converted into a scale ranging from 1 to 5 reflecting how a school's average student growth compares to the average student growth of schools with similar students.

To ensure the balance of this multiple-measure system, TAP recommends that 50 percent be allocated for bonuses based on each teacher's classroom observation score and 50 percent of the bonus be based on student growth. This balanced approach ensures the efforts to improve instructional practices by the teacher are as rewarded as the outcomes, measured by classroom growth measures. Further, this approach ensures that no single factor determines a teacher's compensation or career status. While each of these components operates individually, each also requires teachers to meet a minimum threshold to qualify for that component's associated bonus amount. For example, a teacher could be eligible for the observation (SKR) bonus; however, due to a low value-added score not qualify for the classroom and schoolwide component bonus. These thresholds are articulated in Table 1 below.

Table 1: Qualifying Levels for Bonus Amounts in TAP System Schools

| Component | Career Teachers | Mentor Teachers | Master Teachers |
|---|---|---|---|
| SKR scores (observations) | 2.5 | 3.5 | 4* |
| Classroom growth | 3 | 3 | 3 |
| School growth | 3 | 3 | 3 |

*Based on individual districts, this level may be changed to 3.5.

### Integration, alignment, and fidelity of the TAP elements

As noted throughout this section, the integration and mutual alignment of these four elements are central to understanding TAP (Jerald & Van Hook, 2011). The TAP System's structure of evaluation provides feedback for professional growth, and serves as the basis for determining performance-pay awards. Simultaneously, this evaluation structure relies on master and mentor teachers as well as principals to carry out the observational assessments and provide personalized feedback, mentoring, training, and other support for improvement.

# Design

## Sample

The primary research question for this study is to examine the academic performance of students in Louisiana TAP schools relative to students in matched schools. The Louisiana Department of Education received a TIF-3 grant in fall 2010 to implement TAP in 68 public schools. Seventeen other Louisiana public schools participated in the TAP System either through separate TIF grants or their own funding initiatives. This study examines the impact of the TAP System on student achievement in the 66 schools across the state that participated in the program in 2010-11 through at least 2012-13 and for which the state reported K-8 student achievement data. The 66 schools included 52 schools participating in the TIF Cycle 3 grant and 14 independently utilizing TAP.

## Description of Data Sources

Louisiana public school students participate in the Louisiana LEAP and iLEAP testing program. Sub-tests are administered each spring in English language arts (ELA), mathematics, science, and social studies in grades 3 through 8. The iLEAP integrates norm-referenced and criterion-referenced components. LEAP is criterion-referenced only and is used in grades 4 and 8 whereas iLEAP is given in grades 2, 5, 6, and 7. High school students participate in end-of-course (EOC) exams. Rather than using LEAP, iLEAP, or EOC scores alone, the analysis utilizes the continuous Assessment Index measure to include as many grades and subject areas as possible.

For schools serving students in any of the grades 3 through 8, the K-8 Assessment Index (AI) is based on a formula weighted by performance category and subject. The five possible performance categories of LEAP and iLEAP and their assigned weights are Advanced (150 points), Mastery (125 points), Basic (100 points), Approaching Basic (0 points), and Unsatisfactory (0 points). ELA and math sub-tests are given a weight of 2 each, while science and social studies tests are weighted 1 each. The weighted scores are summed then divided by the number of test units weighted. As an example, a 6th grade student who scored Advanced in math and Basic on the other three subject area sub-tests would contribute to the school total as follows:

(100*2) for ELA + (150*2) for math + (100*1) for science + (100*1) for social studies = 700 points / 6 (# of test units weighted) = 116.7 points

Since 2013, the maximum possible score for a student or school (average of all students tested) is 150. Prior to 2013, the maximum score was 200, and the point values by scoring category were Advanced=200, Mastery=150, Basic =100, Approaching Basic=50, and Unsatisfactory=0.

School demographic data, including enrollment, student ethnicity, free and reduced lunch status, and Limited English Proficiency (LEP) classification, were taken from public files available from the Louisiana Department of Education for the 2011-2012 school year.

## Analytic Strategy

We used two strategies to assess the impact of the TAP System on student achievement, as measured by the 2012-13 K-8 Assessment Index (AI). Using a linear regression, we first compared 2012-13 AI performance of the 66 TAP schools and non-TAP schools statewide, controlling for prior (2010-11) achievement, percentage of students receiving free or reduced price lunch, school configuration, school size (number of students), and percentage of English language learners.

Second, we compared the TAP schools with a propensity score matched group of non-TAP schools (Rosenbaum & Rubin, 1983; 1985). We matched on baseline (2010-11) student achievement, school configuration (grades taught), percentage of students receiving free or reduced price lunch, and school size (number of students), using the MatchIt package in R (Ho, Imai, King, & Stuart, 2011). A matching algorithm was employed to achieve optimal balance on the covariates after matching (Sekhon, 2011). The matched groups were compared on 2012-13 K-8 AI as the dependent variable.

# Results

*Comparison of TAP with non-TAP schools statewide*. We performed a linear regression of the school 2012-13 Assessment Index (AI) on the following covariates:

- 2010-11 AI
- Percentage of students receiving free or reduced price lunch, 2011-12
- School configuration (elementary, middle/junior high, or combination school)
- School size (number of students), 2011-12
- Percentage of English language learners, 2011-12

Regression coefficients for all covariates were significant at $p < .05$ (n = 1003 schools with all covariate information), as shown in Table 2. Controlling for the covariates, implementation of the TAP System showed a significant positive effect on 2012-13 achievement: the 66 TAP schools scored 3.7 points higher on average than non-TAP schools ($p < .01$).

Table 2: Coefficient Estimates for Regression with Outcome 2012-13 AI

| | Estimate | Standard error | t | p |
|---|---|---|---|---|
| Intercept | 19.252 | 3.558 | 5.41 | .000*** |
| 2010-11 AI | 0.786 | 0.022 | 36.16 | .000*** |
| % FRL | -0.183 | 0.022 | -8.27 | .000*** |
| School configuration (baseline is elementary): | | | | |
| Middle/junior high school | -1.669 | 0.786 | -2.12 | .034* |
| High school | -4.009 | 0.984 | -4.07 | .000*** |
| School size | -0.003 | 0.001 | -2.37 | .018* |
| % LEP | 0.334 | 0.076 | 4.41 | .000*** |
| TAP school | 3.699 | 1.221 | 3.03 | .003** |

* $p < .05$. ** $p < .01$. *** $p < .001$.

*Comparison of TAP schools with matched controls*. We selected one propensity score matched control for each of the 66 TAP schools. The best fitting model for the propensity score logistic regression was determined by minimum Akaike Information Criterion (AIC) (Findley, 1999). The covariates in this model were 2010-11 AI, school configuration, percentage of students receiving free or reduced price lunch, and school size (number of students). To verify the quality of matching, we checked the standardized differences in means of the covariates between the TAP schools and the matches. Table 3 shows that all differences were well below the maximum acceptable value of 0.25 (Stuart, 2010). Like the TAP schools, the matched group consists of 47 elementary schools, 15 middle schools, and 4 combination schools that serve students in the K-8 range.

Table 3: Means of Covariates for Louisiana TAP Schools and Matched Schools

| | TAP Mean | Matched Controls Mean | Absolute Mean Difference | SD Treated Group (TAP) | Standardized Difference |
|---|---|---|---|---|---|
| 2010-11 AI | 74.22 | 74.26 | 0.04 | 13.09 | 0.003 |
| % FRL | 86.40 | 86.39 | 0.01 | 12.62 | 0.001 |
| School size | 495.30 | 495.15 | 0.15 | 201.34 | 0.001 |

The average 2012-13 AI for TAP schools (64.45) was 5.47 points greater than the average for the matched controls (58.98). To illustrate the gain in score for the TAP schools, Figure 1 shows the equivalent starting point for both TAP and control schools as a function of the maximum score on the 2010-11 AI and then shows the change in score relative to the maximum score on the 2012-13 AI. The propensity matching literature debates whether to use an independent t-test or a dependent (matched pairs) t-test to compare the treatment group with the matched controls (Austin, 2011; Stuart, 2010). We initially conducted the more conservative analysis – treating TAP schools and controls as independent groups – resulting in a p-value of .05 ($t$ = 2.00). In addition, we then treated the matches as dependent pairs, which, as expected, resulted in a more significant finding, $p < .01$ ($t$ = 3.10). Both of our analyses result in a significant finding at the $p < .05$ level. The effect size for the mean difference is $d$ = 0.35.

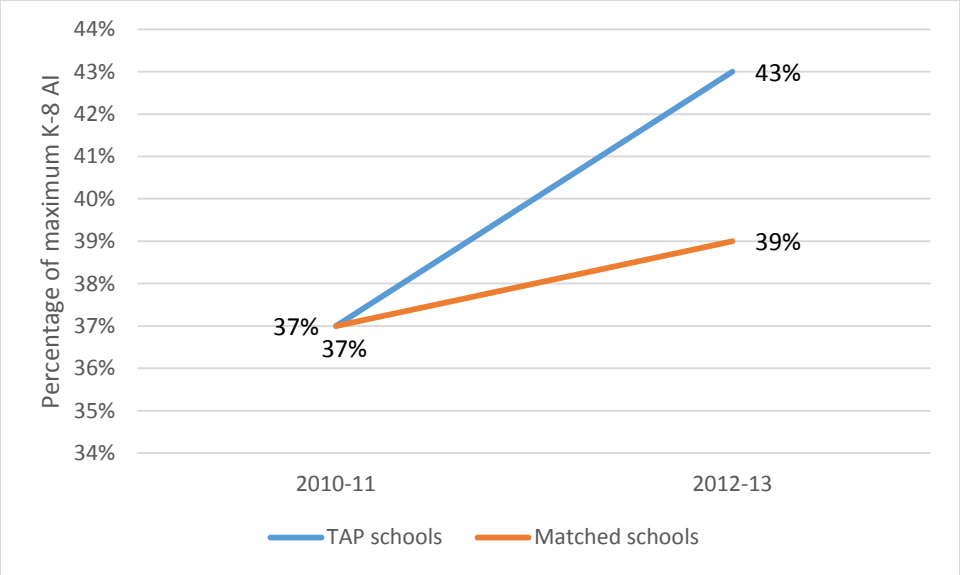


Figure 1. 2012-13 K-8 Assessment Index averages (as percentages of maximum possible), TAP schools and matched controls. (The maximum K-8 AI in 2010-11 was 200, while the maximum in 2012-13 was 150.)

# Discussion

The push for models of improvement in education is ushered on by the reality that America's education system is falling behind other nations. International reports and public perception continue to establish a clear pattern of mediocrity in America's schools, where students are not performing nearly as well as previously thought. For example, the most recent administration of the PISA (OECD, 2012) indicates only 2 percent of American students reach the highest level of mathematics performance, a figure below the average of 3 percent for all nations and far below the maximum of 31 percent of students, achieved by students in Shanghai-China. Further highlighting this trend, over 25 percent of America's students do not reach the PISA proficiency baseline (level 2) in mathematics.

The theory of action for the TAP System is that creating multiple career paths for teachers, providing ongoing applied professional development using a rigorous rubric of evaluation, and providing performance-based compensation will lead to improved educator effectiveness and result in improved student achievement. Prior research examining the TAP System has supported this theory by demonstrating a consistent pattern of increased student achievement. The research base on the achievement impact of the TAP System has applied a variety of sophisticated and robust statistical strategies for analyzing test score data (Hudson, 2010; Schacter & Thum, 2005; Solmon et al., 2007). A recent third-party study of Louisiana used propensity score matching of 14 TAP and non-TAP schools (Mann, Leutscher, & Reardon, 2013). The current study expands the propensity score matched-control methodology to a much larger group of 66 Louisiana TAP schools, providing additional robust evidence of significantly increased achievement growth. To further interpret the achievement growth of schools within this study, the effect size for the improvement is above that found for other interventions, including reducing class sizes down to 15 students (Coe, 2002).

The improvement experienced in TAP schools does not happen by chance. The results for TAP System schools are continually crystallized in action by the leadership teams at each school site. These individuals work tirelessly to ensure their practices are improving and recognize that with improved teacher and principal practice, student achievement gains will follow. The present study demonstrates student achievement improvement in TAP System schools — providing evidence that the TAP System can be successfully implemented across a variety of school and community contexts, and is indeed associated with increased achievement growth in high-need schools.

# References


Algiers Charter School Association. (2011). Annual report 2011. Retrieved from http://www.tapsystem.org/publications/algiers_charter_schools_association_annual_report_2011.pdf

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46,* 399-424.

Barnett, J. H., Rinthapol, N., & Hudgens, T. (2014). TAP Research Summary: Examining the Evidence and Impact of TAP: The System for Teacher and Student Advancement. National Institute for Excellence in Teaching.

Berry, B., Daughtrey, A., & Wieder, A. (2010). Teacher effectiveness: The conditions that matter most and a look to the future. *Center for Teaching Quality*.

Bezzina, C. (2006). Views from teachers: Beginning teachers' perceptions about their professional development. *Journal of In-Service Education, 32*(4), 411-430.

Buck, B., & Coffelt, C. (2013). Leadership program good for all teachers: TAP leaders provide support for colleagues, create collaborative team. *Des Moines Register*. Retrieved from http://altoonaherald.desmoinesregister.com/article/20130814/OPINION01/308140015/Iowa-View-Leadership-program-good-all-teachers

Camburn, E., Rowan, B., & Taylor, J. E. (2003). Distributed leadership in schools: The case of elementary schools adopting comprehensive school reform models. *Educational Evaluation and Policy Analysis, 25*(4), 347-373.

Carpenter, W. A. (2000). Ten years of silver bullets: Dissenting thoughts on education reform. *Phi Delta Kappan, 81*, 383-89.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2011). Teacher mobility, school segregation, and pay-based policies to level the playing field. *Education Finance and Policy, 6*(3), 399–438.

Coe, R. (2002). It's the effect size, stupid. Paper presented at the British Educational Research Association annual conference. September 12-14, 2002. Retrieved from http://www.cem.org/attachments/ebe/ESguide.pdf

Copland, M. A. (2003). Leadership of inquiry: Building and sustaining capacity for school improvement. *Educational Evaluation and Policy Analysis, 25*(4), 375-395.

Daley, G., & Kim, L. (2010). A teacher evaluation system that works (NIET Working Paper). Santa Monica, CA: National Institute for Excellence in Teaching.

Darling-Hammond, L. (2006). Assessing teacher education: The usefulness of multiple measures for assessing program outcomes. *Journal of Teacher Education, 57*(2), 120-138. doi:10.1177/0022487105283796

Eckert, J. (2013). Increasing educator effectiveness: Lessons learned from Teacher Incentive Fund sites.

National Institute for Excellence in Teaching. Retrieved from http://www.niet.org/assets/increasing-educator-effectiveness-lessons-learned-from-teacher-incentive-fund-sites.pdf

Elmore, R. F. (2000). *Building a structure for school leadership.* Washington, DC: The Albert Shanker Institute.

Elmore, R. F. (2002). The limits of change. *Harvard Education Letter, 18,* 6-8.

Findley, D. F. (1999). AICII. In S. Kotz, C. R. Read, & D. L. Banks (Eds.), *Encyclopedia of Statistical Sciences.* Wiley-Interscience, New York.

Fullan, M. (2001). *The new meaning of educational change*. London: Routledge Falmer.

Glazerman, S., Chiang, H., Wellington, A., Constantine, J., & Player, D. (2011). *Impacts of performance pay under the Teacher Incentive Fund: Study design report.* Mathematica Policy Research. Retrieved from http://www.mathematica-mpr.com/publications/PDFs/education/performpay_TIF.pdf

Goldhaber, D. (2002). The mystery of good teaching. Education Next, 1, 50-55.

Goldhaber, D. D. & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation & Policy Analysis*, *22*(2), 129-45.

Gordon, R., Kane, T. J., & Staiger, D. O. (2006). *Identifying effective teachers using performance on the job*. Washington, D.C.: The Brookings Institution.

Grissom, J. A. (2011). Can good principals keep teachers in disadvantaged schools? Linking principal effectiveness to teacher satisfaction and turnover in hard-to-staff environments. *Teachers College Record, 113*(11), 2552–2585.

Gronn, P., & Rawlings-Sanaei, F. (2003). Recruiting principals in a climate of disagreement. *Australian Journal of Education, 47*(2), 172-184.

Hanushek, E. (2013). Why educators' wages must be revamped now. Education Week. Retrieved from http://www.edweek.org/ew/articles/2013/02/06/20hanushek_ep.h32.html

Haycock, K. (1998). *Good teaching matters: How well-qualified teachers can close the gap.* Washtington DC: Education Trust.

Ho, D., Imai, K., King, G., & Stuart, E. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, *42*(8), 1-28. http://www.jstatsoft.org/v42/i08/.

Hudson, S. (2010). The effects of performance-based teacher pay on student achievement. Discussion Paper for the Stanford Institute for Economic Policy Research, Stanford University. Retrieved from http://www.stanford.edu/group/siepr/cgi-bin/siepr/?q=system/files/shared/pubs/papers/09023_Paper_Hudson.pdf

Ingersoll, R. M. (2004). Why do high-poverty schools have difficulty staffing their classrooms with

qualified teachers? Washington, DC: Center for American Progress and the Institute for America's Future. Retrieved from http://www.americanprogress.org/kf/ingersoll-final.pdf

Ingersoll, R. M., & Merrill, L. (2012). Seven trends: The transformation of the teaching force. The Consortium for Policy Research in Education.

Jackson, C. K. & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. NBER Working Paper 15202. Washington, DC: National Bureau of Economic Research.

Jerald, C. D., & Van Hook, K. (2011). More than measurement: The TAP System's lessons learned for designing better teacher evaluation systems. National Institute for Excellence in Teaching. Retrieved from http://www.tapsystem.org/publications/eval_lessons.pdf

Johnson, S. M., Kraft, M., & Papay, J. P. (2012). How context matters in high-need schools: The effects of teachers' working conditions on their professional satisfaction and their students' achievement. Teachers College Record, 114(10), 1–39.

Kardos, S. M., Johnson, S. M., Pekse, H. G., Kauffman, D., & Liu, E. (2001). Counting on colleagues: New teachers encounter the professional cultures of their schools. *Educational Administration Quarterly, 37,* 250-290.

Lambert, L. (1998). How to build leadership capacity. *Educational Leadership, 55*(7), 17-19.

Leithwood, K., & Mascall, B. (2008). Collective leadership effects on student achievement. *Educational Administration Quarterly, 44*(4), 529–561.

Mann, D., Leutscher, T., Reardon, R. M. (2013). The system for teacher and student advancement: An evaluation of achievement and engagement in Louisiana. National Institute for Excellence in Teaching. Retrieved from http://www.niet.org/assets/PDFs/interactive-louisiana-student-achievement.pdf

Murray, S., Ma, X., & Mazur, J. (2009). Effects of peer coaching on teachers' collaborative interactions and students' mathematics achievement. *Journal of Educational Research, 103*(2) 203–212.

National Center on Teacher Quality. (2014). *2013 State Teacher Policy Yearbook: National Summary*. NCTQ. Washington DC. Retrieved from http://www.nctq.org/dmsView/2013_State_Teacher_Policy_Yearbook_National_Summary_NCTQ_Report

OECD. (2011). *Lessons from PISA for the United States, Strong performers and successful reformers in education*. OECD Publishing. Retrieved from http://www.oecd.org/pisa/46623978.pdf

OECD. (2012). *Programme for International Student Assessment (PISA) results from PISA 2012: United States Key Findings*. OECD Publishing. Retrieved from http://www.oecd.org/pisa/keyfindings/PISA-2012-results-US.pdf

Rivkin, S.G., Hanushek, E.A., & Kain, J.F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417-458.

Ritter, G. W., & Barnett, J. H. (2013). *A straightforward guide to teacher merit pay: Encouraging and rewarding schoolwide improvement.* Thousand Oaks, CA: Sage.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review, 94*(2), 247-252.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*(1): 41–55. doi:10.1093/biomet/70.1.41

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician, 39*, 33–38.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach.* (7th ed.). Thousand Oaks, CA: Sage.

Rowan, B. (1990). Commitment and control: Alternative strategies for the organizational design in schools. In C. Cazden's (Ed.), *Review of Research in Education* (pp. 353-389). Washington DC: American Educational Research Organization.

Rowan, B., Correnti, R. & Miller, R. J. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the Prospects study of elementary schools. *Teachers College Record, 104*, 1525-1567.

Schacter, J., Thum, Y. M., Reifsneider, D., & Schiff, T. (2004). The Teacher Advancement Program report two: Year three results from Arizona and year one results from South Carolina TAP schools. Santa Monica, CA: Milken Family Foundation. Retrieved from http://www.tapsystem.org/pubs/tap_results_azsc2004.pdf

Schacter, J., & Thum, Y. M. (2005). TAPping into high quality teachers: Preliminary results from the Teacher Advancement Program comprehensive school reform. *School Effectiveness & School Improvement, 16*(3), 327–353.

Sekhon, J. S. (2011). Multivariate and propensity score matching software with automated balance optimization: The Matching package for R. *Journal of Statistical Software, 42*(7), 1-52. http://www.jstatsoft.org/v42/i07/.

Solmon, L. C., White, J. T., Cohen. D., & Woo, D. (2007). The effectiveness of the teacher advancement program. National Institute for Excellence in Teaching. Retrieved from http://www.tapsystem.org/pubs/effective_tap07_full.pdf.

Timperley, H. S. (2005). Distributed leadership: Developing theory from practice. *Journal of Curriculum Studies, 00*, 1-26.

TNTP (The New Teacher Project). (2012). The Irreplaceables: Understanding the real retention crisis in America's urban schools. Retrieved from http://tntp.org/assets/documents/TNTP_Irreplaceables_2012.pdf

Trorey, G., & Cullingford, C. (2002). *Professional development and institutional needs.* Aldershot Hampshire, England: Ashgate.

Tyack, D., & Cuban, L. (1995). *Tinkering toward Utopia*. Cambridge, MA: Harvard University Press.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The Widget Effect: Our national failure to acknowledge and act on differences in teacher effectiveness. Retrieved from http://tntp.org/publications/view/evaluation-and-development/the-widget-effect-failure-to-act-on-differences-in-teacher-effectiveness