

Staying Power: English Language Arts Achievement after Installation of an Educator Effectiveness System

Trevor Leutscher, PhD
Senior Research Associate
National Institute of Excellence in Teaching
Scottsdale, AZ

Tanée M. Hudgens, PhD
Vice President of Research and Evaluation
National Institute for Excellence in Teaching
Scottsdale, AZ

Handrea Logis, PhD
Director of Research
National Institute for Excellence in Teaching
Scottsdale, AZ

Marina Serdiouk, PhD
Senior Research Associate
National Institute for Excellence in Teaching
Scottsdale, AZ

Joshua H. Barnett, PhD
Chief Executive Officer
National Institute for Excellence in Teaching
Scottsdale, AZ

Abstract

As long-term effects of educational programs cannot be sufficiently addressed by evaluators during short duration implementation grants, researchers become responsible for investigating sustainability of effects and addressing concerns of policy makers and grant funding agencies. This study examines the impact of the TAP System for Teacher and Student Advancement on the percentage of students passing Indiana's English/language arts achievement test after grant-funded implementation. The study uses a school-level, quasi-experimental design. Propensity score matching generates a comparison group and regression models are then run, controlling for school fixed effects. The results show that schools who sustained the TAP System significantly outperform matched comparison schools in the first post-grant year and the improved performance persists through the third year.

Key Words

evaluation; sustainable implementation; sustaining improved outcomes; English / language arts achievement; TAP System for Teacher and Student Advancement; long-term effects; school reform

School reforms are intended to generate long-term benefits for schools (students, parents, teachers, and administrators). Too often evaluations of reforms lack the resources to measure whether improved outcomes sustain long-term, for 5, 10, or more years (Bigelow et al., 2021). When grants provide funding for initial implementation, program evaluation activities are often limited to the period of grant funding. However, the true test of the success of that initial implementation is whether the reform and improved outcomes are sustained after grant funding ends.

Sustaining the improved outcomes and sustaining the reform usually go together; maintaining the activities that purportedly caused the improved outcomes would be required to sustain those outcomes. This study examines the sustainability of improved English / language arts (ELA) outcomes of an educator effectiveness reform, the TAP System for Teacher and Student Advancement (TAP System), in schools that sustain the TAP System after grant-funded adoption and initial implementation.

The TAP System and Sustainability

Studies and literature reviews examining school improvement and reform initiatives (e.g., Borman et al., 2003; Cohen & Mehta, 2017; Datnow, 2005; Desimone, 2002; Giles & Hargreaves, 2006) have identified characteristics of reforms that influence their sustainability.

Reforms are more likely to be sustained when they have the following five characteristics (Coburn et al., 2012; Cohen & Mehta, 2017; Desimone, 2002; Li, 2017; Savaya & Spiro, 2012; Wiltsey Stirman et al., 2012). First, the reforms solve a problem for the people implementing the reform. Second, people understand how the reforms solve the problem. Third, school, district, community,

and government groups support the reforms. Fourth, existing internal resources or an external organization supports implementing the reforms. Finally, the reforms are consistent with the values of people affected by the reforms.

The TAP System is an educator effectiveness reform designed to attract, develop, motivate, and retain effective educators. Since 1999, hundreds of schools have implemented the TAP System through federal, state, and local funding initiatives.

The TAP System theoretical framework consists of four aligned core elements designed to improve educator effectiveness; thereby, improving students' academic success and opportunities (National Institute for Excellence in Teaching, n.d.).

These elements are described below:

Multiple career paths

In TAP System schools, teachers can serve as teacher leaders, receiving additional compensation for providing high levels of support to their peers. Along with administrators, teacher leaders form a leadership team to deliver school-based professional support and appraise teachers' performance.

Ongoing applied professional growth.

In TAP System schools, teachers participate in weekly professional learning community (PLC) meetings, led by teacher leaders, in which they examine student data, engage in collaborative planning, and learn instructional strategies that have been field-tested in their respective schools. Professional learning continues into each classroom as teacher leaders model lessons, observe classroom instruction, and support classroom teachers in the improvement of their teaching methods.

Instructionally focused accountability.

In TAP System schools, teachers are observed in classroom instruction several times a year by multiple, trained observers. Student growth analysis complements these classroom observations, rounding out a multi-measure system of teacher appraisals. Observation results guide both formative feedback for one-on-one mentoring sessions and plans for PLC meetings, ensuring relevant professional development for teachers and a consistent vision for instruction.

Performance-based compensation.

Teachers in TAP System schools can earn annual bonuses based on their observed skills, knowledge and responsibilities, the average academic growth of students in their classroom, and the entire school's average growth in achievement. Teacher leaders receive additional compensation in recognition of their additional support roles and responsibilities.

Through these core elements, schools develop a school environment conducive to sustaining the TAP System and the improved outcomes attained from adopting it. The TAP System has most of the characteristics of reforms that support sustainability. First, schools adopt the TAP System to address an identified schoolwide problem.

The TAP System helps school leaders recognize, diagnose, and solve instructional issues across the school or within individual teachers' classrooms (National Institute for Excellence in Teaching, 2021).

This feature of the TAP System increases the likelihood of sustaining improved outcomes beyond initial implementation. Second, the TAP System has a well-defined theory of action: applied professional development delivered via teacher leaders using a rigorous rubric of evaluation complemented by performance-based

compensation will lead to improved teacher effectiveness, which will lead to improved student achievement (Barnett & Hudgens, 2014; National Institute for Excellence in Teaching, 2021). Third, schools implementing the TAP System receive guidance and external support from the National Institute for Excellence in Teaching (NIET).

NIET has available a wealth of documentation about the TAP System and best practices for implementing the core elements of the TAP System (e.g., (National Institute for Excellence in Teaching, 2017, 2020). During initial implementation, schools receive support, sometimes with the assistance of grant funding, from NIET personnel who are trained to help schools effectively implement the TAP System. Post-grant, NIET assistance continues with additional training and support.

Student Performance and the TAP System

Prior research has demonstrated the impact of implementing the TAP System on student outcomes. Barnett and Wills (2016) found that passing percentages on the state achievement test for students in TAP System schools improved over time and narrowed achievement gaps in ELA and mathematics vis-à-vis a matched comparison group.

A study of a grant-funded implementation found that TAP System schools outperformed a matched comparison group over a four-year period and the difference was statistically significant after the second year of implementation (Mann et al., 2013). Schacter and Thum (2005) found that achievement growth of TAP System schools was significantly better than control schools. Springer et al., (2014) found positive effects on fall-to-spring student test score gains that were statistically significant in elementary grades and non-significant in most secondary grades.

These studies all evaluate the TAP System from adoption through a few years of initial implementation. A study, paralleling the current study, investigated the impact of the TAP System on mathematics achievement in Arizona after grant-funded implementation (Leutscher & Barnett, 2020). Schools that sustained the TAP System significantly outperformed comparison schools, selected at the end of the grant, two and three years after grant funding ended.

Current Study

The study investigates the impact of the TAP System after initial, grant-funded

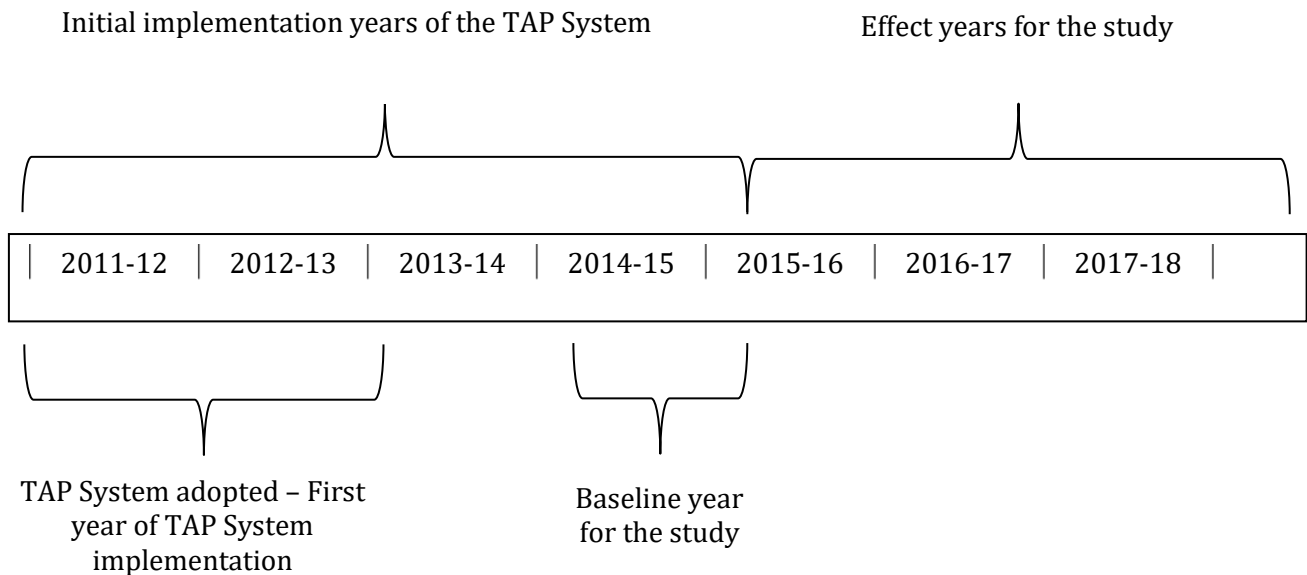
implementation (post-grant implementation) on ELA achievement.

Data on schools with a post-grant implementation of the TAP System is seen below in Figure 1:

1. adopt the TAP System at least two years before the baseline year for the study,
2. implement the TAP System for three consecutive years though the baseline year (i.e., the baseline year may count as one of the initial implementation years), and
3. maintain the TAP System for at least three years after the baseline year.

Figure 1

Timeline for Grant and Post-grant Implementation of the TAP System



Background

A Teacher Incentive Fund Cohort 3 (TIF3) grant assisted with funding the adoption and initial years of TAP System implementation for 44 schools. The TIF3 grant required (a) implementing a performance-based compensation system (PBCS), (b) guarantees of fiscal sustainability of the PBCS and (c) alignment of the PBCS with strategies for strengthening the workforce.

All schools adopted and began implementation during the 2010-11 school year. The final evaluation report (Mann & Leutscher, 2016) finds no statistically significant difference on state achievement tests between the TAP System schools and matched comparison schools at any point during initial implementation. On average across all schools in the grant, the percentage passing the state ELA assessment from 2009-10 to 2013-14 (the state test changed in 2014-15) increased 6.5 points. Comparing schools that sustained the TAP System after the grant to those that did not, the sustaining group increased the percentage passing 7.5 points in ELA, while the non-sustaining group increased 5.6. In 2013-14, the mean difference between the sustaining group and the non-sustaining group was not significant.

The TIF3 grant ended in school year 2014-15, but with an optional one-year, no-cost extension through 2015-16. After 2014-15, some schools sustained the TAP System over consecutive years through the 2017-18 school year; many other grant schools maintained a relationship with NIET but did not maintain the full TAP System. During the 2014-15 to 2017-18 post-grant implementation, the TAP System schools continued with some NIET services and could request additional support or training. The continued use of its services allowed NIET to monitor whether schools sustained the TAP System with fidelity.

Methods

Data

The study uses school-level, public-use data files available on the Indiana Department of Education website (<https://www.in.gov/doe/it/data-center-and-reports/data-reports-archive/>).

Research (Jacob et al., 2014) has shown that under certain conditions (i.e., sample size greater than thirty and low variation in school size) aggregate school-level data are sufficient for assessing impacts of school-based programs. For privacy reasons (e.g., small group size), some data may be masked in public-use data sets.

However, in Indiana, data masking is minimal. The Indiana Department of Education reports school-level aggregate achievement results as the percentage of students passing the test. To protect students' anonymity, the state masks data when the number of test takers is less than ten students. For enrollment data, Indiana does not mask any data, which allows for accurate calculations for percentage of students eligible for free and reduced-price meals and for the percentage of students in racial/ethnic groups.

Outcome measure

The outcome measure is the percentage of students passing the Indiana state assessment (Indiana Statewide Testing for Educational Progress-Plus, ISTEP+) for ELA. The ISTEP+ assessment measures student achievement according to the Indiana Academic Standards (<https://www.in.gov/doe/students/indiana-academic-standards/>). For the 2014-15 school year, Indiana implemented new academic standards and new ISTEP+ tests to assess achievement of the standards. As expected, the pass rates for students decreased about 20% statewide from 2013-14 to 2014-15. While pass

rates often recover in subsequent years from introduction of a new standardized test, such a recovery did not happen in Indiana. Pass rates continued to decline statewide through 2017-18 (the last year of ISTEP+ testing in K-8 grades). During the study period (2014-15 to 2017-18), statewide, TAP System group, and matched comparison group averages trend downward. On average, Indiana schools drop 4.5 percentage points in ELA from 2014-15 to 2017-18. The average TAP System school drops 3.1 points.

Study Sample

All study schools are in Indiana. In total, Indiana has 1,883 schools with ELA test data for at least one year during the study period. All study schools (TAP System and the pool of comparison schools) must have all relevant data (percentage passing and demographics) publicly available across the four study years, 2014-15 (baseline) through 2017-18. For the pool of comparison schools, 1,406 schools satisfied this criterion. The first step in the matching process required matching TAP System schools to schools with the same tested grades. Filtering the pool of comparison

schools provides 646 potential comparison schools.

Table 1 provides baseline demographic information about the TAP System and potential comparison groups prior to matching. The Free and Reduced Meals (FRM) percentage for the comparison group is 22 points less than for the TAP System group. The comparison group, on average, is 26 points below the TAP System group in percentage of English language learners. Average enrollment is 93 students higher in the TAP System group.

The TAP System group has a higher percentage of Hispanic students than the comparison group, while the comparison group has a higher percentage of White students. The percentage minority (non-white) differs by 35 points with the higher percentage in the TAP System schools. The percentage of students passing the state assessment shows a ten-point gap in favor of the comparison group. Due to the observed gaps in percentage passing ELA and percentage minority, the study employs a one-to-two matching procedure to select 26 comparison schools.

Table 1

Schools Characteristics of TAP and All Potential Comparison Schools at baseline (2014-15)

School Characteristic	TAP System Schools			Potential Comparison Schools		
	Mean	St. Dev.	Range	Mean	St. Dev.	Range
Free and Reduced Meals	73%	8%	59% - 86%	51%	22%	1% - 100%
English Language Learners	31%	13%	5% - 47%	5%	8%	0% - 50%
Students with Disabilities	14%	4%	9% - 25%	16%	5%	2% - 36%
School Enrollment	605	305	366 - 1,485	512	220	78 - 1,579
Asian	6%	13%	0% - 43%	2%	4%	0% - 31%
Native American	0%	0%	0% - 1%	0%	0%	0% - 2%
Black/African American	13%	26%	1% - 95%	10%	15%	0% - 95%
Hispanic/Latino	42%	19%	8% - 64%	10%	12%	0% - 79%
White	39%	16%	1% - 73%	74%	24%	0% - 99%
Percent Minority (Non-White)	61%	16%	27% - 99%	26%	24%	1% - 100%
Percent Passing ELA	59%	7%	46% - 71%	69%	12%	24% - 96%

Propensity Score Matching

The study selects comparison schools using a two-phase matching process. The first phase separates the schools into groups by grade configurations that contain the same tested grades at baseline (e.g., group schools with only tested grades 6 through 8). This filtering process ensures that differences in test difficulty by grade do not introduce bias. The second phase performs one-to-two, nearest-neighbor propensity score matching (PSM) on each group. As covariates, the PSM uses baseline (2014-15) (a) school-wide percentage passing the state ELA assessment, (b) percentage of minority (non-white) students and (c) state computed letter grade for the school. Table 2 presents results from the propensity score matching. As shown, the

balance between TAP System and the comparison groups before and after matching improves for all the covariates used in the matching process. Table 2 also shows other variables used during the matching process but not included in the final matching model. Among these variables, the percentage of students eligible for free or reduced-price meals has a substantial improvement in balance; enrollment and the percentage of Hispanic/Latino students reduce the difference by about half; and the percentage of Black/African American students does not change. The models using these additional variables were usually rejected due to lack of baseline equivalence for the percent passing variable.

Table 2

Propensity Score Matching. Before and After Results for Covariates

PSM Covariate		TAP	Comparison		Balance Improvement	T-Test p-value	Effect Size
		System (n = 13)	Before (n = 646)	After (n = 26)			
Letter Grade	A	31%	61%	31%	100%		
	B	31%	18%	31%	100%		
	C	38%	15%	38%	100%		
	D	0%	5%	0%	100%		
	F	0%	2%	0%	100%		
Percent Minority		61.0%	25.6%	59.8%	96%	0.856	0.05
Percent Passing		59.3%	69.0%	58.6%	93%	0.799	0.07
Other Variables							
Free/Reduced Meals		73.3%	50.8%	74.4%	95%	0.794	0.07
School Enrollment		605	512	567	59%	0.675	0.18
Percent Black		12.8%	23.4%	23.4%	-1%	0.197	0.52
Percent Hispanic		38.4%	10.1%	25.1%	53%	0.074	0.65

Four public school districts and one charter school association operate the TAP System schools in this study. Eight of the TAP System schools belong to one school district. The comparison schools come from fifteen school districts. One TAP System school is a charter school. The matching process did not filter charter schools, but no charter schools were selected as comparison schools. The schools in both groups reside in a mix of urban, suburban, and rural settings.

Analytic Approach

Linear regression models estimate the impact of the TAP System on the three outcome variables (e.g., 1st Year, 2015-16, percentage passing ELA). The regression controls for baseline (2014-15) percentage passing ELA, percentage of minority (non-white) students, and school letter grade as dummy coded variables (Grade A coded zero and variables for Grades B and C). The study computes Hedge's *g* effect sizes and t-test statistical significances

for the unstandardized regression coefficient on the TAP System/comparison indicator variable.

Results

As is shown in Figure 2, the TAP System group has a small advantage at baseline (0.7 points). Controlling for this advantage, removes a small portion of the unadjusted mean difference (grey bars) compared to the regression estimates, adjusted mean difference (blue bars). As presented in Table 3, the first year after baseline the TAP System group performs 4.3 points better than the comparison group after controlling for baseline differences, a statistically significant result ($p = .017$) with a small effect size of 0.39. Two years from baseline the TAP System group obtains a medium effect size ($g = 0.60$) performing at a statistically significant ($p = .002$) 6.1 points better than the comparison group. Three years from baseline the impact continues to be statistically significant ($p = .002$) with a medium effect ($g = 0.64$) and the TAP System group outperforming the comparison group by 7.8 points.

Table 3

Percentage Passing English / Language Arts State Assessments

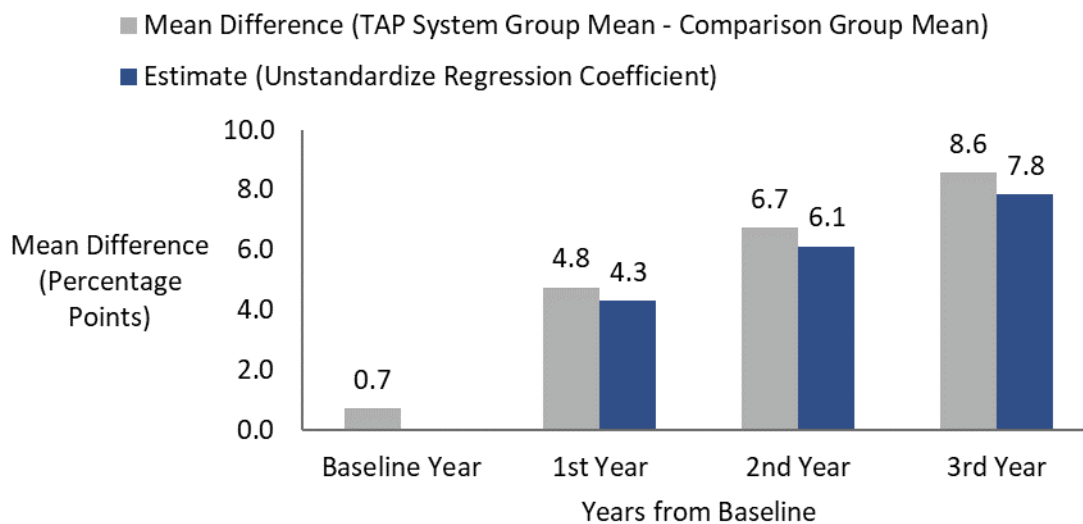
Outcome measures (% Passing ELA)	TAP System group			Comparison group			Estimated effect ¹		
	Sample size	Mean	SD	Sample size	Mean	SD	Estimate	p-value	Effect size ²
3 rd Year	13	56.2	9.0	26	47.6	13.4	7.8	0.002	0.64
2 nd Year	13	58.1	7.8	26	51.4	11.1	6.1	0.002	0.60
1 st Year	13	58.3	7.8	26	53.6	12.2	4.3	0.017	0.39
Baseline measures							Baseline effect		
% Passing	13	59.3	7.1	26	58.6	9.0	0.7	0.399	0.08
% Minority	13	61.0	16.3	26	59.8	26.2	1.2	0.856	0.05

¹Estimates for the outcome measures are the unstandardized regression coefficients on a dummy coded variable for the TAP System/comparison school condition. The baseline measures and dummy-coded variable for the school letter grades are covariates in the regression equation.

²Effect sizes (Hedge's *g*) for the outcome measures are computed using the `esc_B` function of the `esc` package in R (Lüdtke, 2019).

Figure 2

Percentage Passing Indiana's State ELA Assessments, TAP System Schools Compared to Comparison Schools



Discussion

Study Characteristics

This study uses publicly available data obtained from the Indiana Department of Education website. The outcome measure is the

percentage passing the Indiana statewide ELA assessment. A difficulty using percentage passing is that students' scale scores may grow

but not sufficiently to reach the passing threshold. That is, despite growth in scale scores, schools may show only small increases in percentage passing. Since most intervention programs tend to focus on low achieving students—those far below the passing threshold—percentage passing data are not ideal for assessing the performance of such programs.

The TAP System, however, is a school-wide comprehensive (all grades and all subjects) program that aims to help all students grow their performance. As a result, if the TAP System has the intended impact, a larger number of students should grow their achievement above the passing threshold in a single year and over multiple years than programs focused solely on low achieving students. One of the implications of this study is that percentage-passing data can reveal the impact of school-wide reform programs.

The study evaluates the sustained impact of the TAP System on ELA achievement. By the time of the study, all TAP System schools had been implementing for at least three years. When these schools adopted the TAP System, they were among the highest-needs schools in Indiana serving primarily high-needs students, and they were among the lowest performing schools in Indiana. The schools chose to maintain the TAP System after grant funding, which indicates they derived some success during the initial implementation. Given improvement during the initial years of implementation, this study “resets the bar” for the TAP System to show an impact. That is, the study matches TAP System schools to comparison schools using data from at least three years after the adoption and at the end of initial grant funding. The TAP System schools have not only had to continue to perform better than other schools but also had to perform at a higher level than during the initial implementation.

Findings

The study shows that the TAP System not only sustains but improves outcomes during post-grant implementation. That is, the effects of the TAP System do not fade over time. During post-grant implementation of the TAP System, after one year, results are statistically significant with a small effect size. During this first year, all TAP System schools participated in a no cost extension year for the grant, which may have influenced the results. However, two and three years after baseline results continue to be statistically significant with medium effect sizes. At least six years after adoption and grant-funded implementation, the TAP System continues to improve student performance for schools that sustain the system.

Limitations and next steps

As noted, the study uses percentage passing as the outcome variable, which means only students near the passing threshold influence the percentage passing (students far below or above the threshold have little or no effect). Obtaining the average scale score for each school makes every test-taking student in a school count equally toward that outcome measure. Further, obtaining student level outcome and demographic data from Indiana would allow for the development and testing of multilevel models of TAP System performance.

The current study examines sustainability of improved outcomes for three years after grant-funded adoption and implementation of the TAP System. Of the schools that began implementation of the TAP System as part of the TIF3 grant, less than half sustained the system through the end of the study period.

An area for further research is the characteristics of schools and districts that sustain the TAP System beyond initial implementation versus schools that do not. In

fact, investigating why any federal or state funded program fails to be sustained by schools and districts is important and fiscally responsible. While there is no reason to continue a reform that is not working effectively, federal and state grants use tax dollars to support implementation. When schools and districts do not maintain or sustain these reforms, it is important to understand the reasons for future planning of grant funding opportunities.

Conclusions

Policy makers and funding agencies seek reforms that can be sustained and remain effective beyond initial funding. Ultimately, school and district leaders make decisions about the reforms that will benefit their students and teachers. Currently, many schools across the U.S. invest in educator effectiveness programs with elements like the TAP System (teacher leaders, job-embedded professional

development, teacher observation, and financial incentives).

This study is one of the first to examine the TAP System analytically beyond the adoption and initial implementation period. With a post-grant implementation of the TAP System, this study in Indiana examining ELA achievement finds a statistically significant difference between the TAP System schools and matched comparison schools in the first post-grant year and the improved performance persists through the third post-grant year.

This study informs policymakers on the return on investing in such educator effectiveness models and sustaining them for increasing benefits. Finally, the study provides evidence to local school leaders that committing to implementation of the TAP System results in improved student academic performance long-term.

Author Biographies

Trevor Leutscher is a senior research associate at the National Institute for Excellence in Teaching. His research interests include educational evaluation, school and teacher effectiveness, measurement of teaching quality, and educational reform. e-mail: tleutscher@niet.org

Tanée Hudgens is Research and Evaluation Vice President at the National Institute for Excellence in Teaching. Her research interests include policy and practice related to educator professional development, effectiveness, and evaluation. e-mail: thudgens@niet.org

Handrea Logis is a senior researcher at the National Institute for Excellence in Teaching. Her research interests include educational equity, teacher effectiveness, and teacher preparation programs. e-mail: hlogis@niet.org

Marina Serdiouk is a senior research associate at the National Institute for Excellence in Teaching. Her research interests include teacher and principal professional development, teaching practices as they relate to student outcomes, and program evaluation. e-mail: mserdiouk@niet.org

Joshua Barnett is chief executive officer at the National Institute for Excellence in Teaching. His work aims to improve educator quality by examining teacher and principal support and resource distribution and use within schools. He has worked to apply evidence-based best practices across the nation. e-mail: jbarnett@niet.org

References

- Barnett, J. H., & Hudgens, T. M. (2014). *Staying power: The impact of the TAP System on retaining teachers nationwide* (ED556342). ERIC. <https://files.eric.ed.gov/fulltext/ED556342.pdf>
- Barnett, J. H., & Wills, K. C. (2016, April 8). *Narrowing the gap: Achievement gaps in Indiana TAP schools* [Paper presentation]. American Educational Research Association Annual Meeting, Washington, DC.
- Bigelow, J., Pennington, A., Schaberg, K., & Jones, D. (2021). *A guide for using administrative data to examine long-term outcomes in program evaluation* (OPRE Report No. 2021–145). Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. <https://www.acf.hhs.gov/opre/report/guide-using-administrative-data-examine-long-term-outcomes-program-evaluation>
- Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, 73(2), 125–230. <https://doi.org/10.3102/00346543073002125>
- Coburn, C. E., Russell, J. L., Kaufman, J. H., & Stein, M. K. (2012). Supporting sustainability: Teachers' advice networks and ambitious instructional reform. *American Journal of Education*, 119(1), 137–182. <https://doi.org/10.1086/667699>
- Cohen, D. K., & Mehta, J. D. (2017). Why reform sometimes succeeds: Understanding the conditions that produce reforms that last. *American Educational Research Journal*, 54(4), 644–690. <https://doi.org/10.3102/0002831217700078>
- Datnow, A. (2005). The sustainability of comprehensive school reform models in changing district and state contexts. *Educational Administration Quarterly*, 41(1), 121–153. <https://doi.org/10.1177/0013161X04269578>
- Datnow, A., Lockton, M., & Weddle, H. (2020). Redefining or reinforcing accountability? An examination of meeting routines in schools. *Journal of Educational Change*, 21(1), 109–134. <https://doi.org/10.1007/s10833-019-09349-z>
- Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research*, 72(3), 433–479. <https://doi.org/10.3102/00346543072003433>
- Giles, C., & Hargreaves, A. (2006). The sustainability of innovative schools as learning organizations and professional learning communities during standardized reform. *Educational Administration Quarterly*, 42(1), 124–156. <https://doi.org/10.1177/0013161X05278189>
- Jacob, R. T., Goddard, R. D., & Kim, E. S. (2014). Assessing the use of aggregate data in the evaluation of school-based interventions: Implications for evaluation research and state policy regarding public-use data. *Educational Evaluation and Policy Analysis*, 36(1), 44–66. <https://doi.org/10.3102/0162373713485814>

Leutscher, T., & Barnett, J. H. (2020, April 18). *Sustaining success: The TAP System's long-term effects on math achievement in Arizona* [Roundtable presentation]. American Educational Research Association Annual Meeting, San Francisco, CA.

Li, Y. Y. (2017). Processes and dynamics behind whole-school reform: Nine-year journeys of four primary schools. *American Educational Research Journal*, 54(2), 279–324. <https://doi.org/10.3102/0002831216689591>

Lüdecke, D. (2019). *esc: Effect size computation for meta-analysis* (Version 0.5.1) [Computer software]. <https://CRAN.R-project.org/package=esc>

Mann, D., & Leutscher, T. (2016). *Indiana's TIF/TAP program: Report of the year four evaluation: 2014-15*. Interactive, Inc.

Mann, D., Leutscher, T., & Reardon, R. M. (2013). *Findings from a two-year examination of teacher engagement in TAP schools across Louisiana*. Interactive, Inc.

<https://www.niet.org/assets/ResearchAndPolicyResources/fedfc4c05c/interactive-louisiana-student-achievement.pdf>

National Institute for Excellence in Teaching. (n.d.). *The TAP system for teacher and student advancement*. <https://www.niet.org/our-work/our-services/show/the-tap-system-for-teacher-and-student-advancement>

National Institute for Excellence in Teaching. (2017). *TAP evaluation and compensation guide*. <https://www.niet.org/assets/Resources/3cc467ea67/tap-evaluation-and-compensation-tec-guide.pdf>

National Institute for Excellence in Teaching. (2020). *Annual report 2020*. <https://www.niet.org/assets/NIETInFocus/67a147925f/niet-now-2020-annual-report-full-publication.pdf>

National Institute for Excellence in Teaching. (2021). *Elevating teacher effectiveness in a rural community: Findings from a Teacher Incentive Fund grant* [Research brief]. <https://www.niet.org/research-and-policy/show/research/cross-county-arkansas-spring-2021>

Savaya, R., & Spiro, S. E. (2012). Predictors of sustainability of social programs. *American Journal of Evaluation*, 33(1), 26–43. <https://doi.org/10.1177/1098214011408066>

Schacter, J., & Thum, Y. M. (2005). TAPping into high quality teachers: Preliminary results from the Teacher Advancement Program comprehensive school reform. *School Effectiveness and School Improvement*, 16(3), 327–353. <https://doi.org/10.1080/13803610500146202>

Springer, M. G., Ballou, D., & Peng, A. X. (2014). Estimated effect of the Teacher Advancement Program on student test score gains. *Education Finance and Policy*, 9(2), 193–230. https://doi.org/10.1162/EDFP_a_00129

Wiltsey Stirman, S., Kimberly, J., Cook, N., Calloway, A., Castro, F., & Charns, M. (2012). The sustainability of new programs and innovations: A review of the empirical literature and recommendations for future research. *Implementation Science*, 7, Article 17.
<https://doi.org/10.1186/1748-5908-7-17>