WWC Intervention Report

A summary of findings from a systematic review of the evidence

ies INSTITUTE OF EDUCATION SCIENCES

Teacher Training, Evaluation, and Compensation June 2017

TNTP Teaching Fellows

Intervention Description[1]

TNTP Teaching Fellows is a highly selective route to teacher certification that aims to prepare people to teach in high-need public schools. The program recruits professionals seeking to change careers and recent college graduates who are not certified teachers. *TNTP Teaching Fellows* expects its participants to teach for many years, but does not require them to make a minimum time commitment to teaching. Program participants complete online coursework and receive 5–7 weeks of in-person training focused on foundational teaching skills during the summer before they begin teaching. They must demonstrate mastery of these core skills to be eligible to teach. They receive continued professional development and coaching from *TNTP Teaching Fellows* during their first year of teaching, and additional support provided by their schools and districts.[2] As full-time employees of the public schools in which they work, new *TNTP Teaching Fellows* teachers receive the same salary and benefits as other beginning teachers in their school district.

Research[3,4]

The What Works Clearinghouse (WWC) identified one study of teachers trained through *TNTP Teaching Fellows* that falls within the scope of the Teacher Training, Evaluation, and Compensation topic area and meets WWC group design standards.[5] This study meets WWC group design standards without reservations. The study included 4,116 middle and high school students in nine school districts in eight states.[6]

According to the WWC review, the extent of evidence for teachers trained through *TNTP Teaching Fellows* on the academic achievement of middle and high school students was small for one student outcome domain—mathematics achievement. No studies met WWC group design standards in the five other student outcome domains and 11 teacher outcome domains, so this intervention report does not report on the effectiveness of *TNTP Teaching Fellows* teachers for those domains.[7] (See the Effectiveness Summary on p. 6 for more details of effectiveness by domain.)

Effectiveness

TNTP Teaching Fellows teachers had no discernible effects on mathematics achievement for middle and high school students.

Report Contents

This intervention report presents findings from a systematic review of *TNTP Teaching Fellows* conducted using the WWC Procedures and Standards Handbook (version 3.0) and the Teacher Training, Evaluation, and Compensation review protocol (version 3.2).

Table 1. Summary of findings[8]

Outcome domain	Rating of effectiveness	Improvement index *(percentile points)*		Number of studies	Number of students	Extent of evidence
		Average	Range			
Mathematics achievement	No discernible effects	0	na	1	4,116	Small

na = not applicable

Intervention Information

Background

The nonprofit organization TNTP, formerly known as The New Teacher Project, founded *TNTP Teaching Fellows* in 1997. TNTP continues to administer the program. Address: 186 Joralemon St., Suite 300, Brooklyn, NY 11201. Web: https://tntpteachingfellows.org/ and http://tntp.org/. Telephone: (718) 233-2800.

Intervention details

TNTP Teaching Fellows partners with school districts to run local teaching fellows programs that recruit and train professionals and recent college graduates to teach in high-need schools. The programs refer to their teachers as "Teaching Fellows." Eligibility requirements, training, and paths to certification vary across the local programs. Participating teachers must conduct their own job searches for teaching positions but receive support from *TNTP Teaching Fellows* throughout the hiring process.

The program's admission process typically involves two stages: (1) an online application that includes essays and (2) a 25-minute telephone interview. Only applicants who are invited based on their online applications participate in the phone interviews, during which applicants answer questions about their teaching beliefs, react to classroom scenarios, and participate in role-playing activities. For the 2016 cohort, *TNTP Teaching Fellows* admitted 11.5% of applicants.[9]

Before starting pre-service summer training, admitted applicants must pass state-required teacher certification tests for their subject areas and complete *TNTP Teaching Fellows*' online "Enrollment Coursework." The coursework consists of four self-directed modules intended to introduce the fundamentals of effective teaching. Participants typically spend 30–42 hours completing the coursework.

TNTP Teaching Fellows teachers participate in full-time, in-person training during the summer before beginning teaching. This pre-service training typically lasts 5–7 weeks, with participants attending 5 days per week for about 10–12 hours each day. The training aims to help teachers master foundational teaching skills and includes seminar sessions, coaching, and student teaching in real summer school classrooms. To complete the summer training and begin teaching in the fall, program participants must demonstrate mastery of core skills such as managing their classrooms, delivering content, and engaging students.

TNTP Teaching Fellows provides its participants with ongoing training and support during their first year as full-time teachers. The exact nature of the support *TNTP Teaching Fellows* provides its teachers has varied over time and across local programs. Participating teachers typically enroll in TNTP Academy, which can be either an in-person or an online program. Through TNTP Academy, teachers attend seminars focused on advanced instructional techniques and receive coaching. *TNTP Teaching Fellows* teachers also complete certification coursework through either the TNTP Academy or a partner university program during their first year or two of teaching. Teachers certified through TNTP Academy must meet the expectations of the Assessment of Classroom Effectiveness (ACE), a multiple-measures evaluation system for first-year teachers that combines principal feedback, classroom observations, student surveys, and (when available) value-added measures of the teacher's contribution to student achievement. In 2016, 81.4% of *TNTP Teaching Fellows* participants successfully passed the ACE and earned their initial teacher certification.

Cost

TNTP Teaching Fellows teachers typically pay TNTP Academy or university tuition; certification fees (for example, testing and fingerprinting fees); and sometimes pre-service training material costs. As of October 2016, estimated tuition costs for the six currently-operating local programs ranged from $4,440 to $6,200 for the 1-year TNTP Academy and from $8,600 to $11,010 for 2-year master's degree university programs. In addition, school districts pay service fees to *TNTP Teaching Fellows*. Information on school district fees varies by program and is available

from the developer. Program participants apply for open teaching positions in partner school districts. If hired, they become regular full-time employees of their school districts and receive the starting salary and benefits of beginning teachers in the school district.

Research Summary

The WWC identified eight eligible studies that investigated the effects of *TNTP Teaching Fellows* on teacher and student outcomes.[10] An additional 34 studies were identified but do not meet WWC eligibility criteria (see the Glossary of Terms in this document for a definition of this term and other commonly used research terms) for review in this topic area. Citations for all 42 studies are in the References section, which begins on p. 7.

Table 2. Scope of reviewed research

Grades	6–12
Delivery method	Whole class
Intervention type	Teacher level

The WWC reviewed eight eligible studies against group design standards. One study is a randomized controlled trial that meets WWC group design standards without reservations. This report summarizes the study. The remaining seven studies do not meet WWC group design standards.

Summary of study meeting WWC group design standards without reservations

Clark et al. (2013) examined the effectiveness of *TNTP Teaching Fellows* teachers compared to other teachers in their schools using a randomized controlled trial conducted in 44 secondary schools in nine school districts in eight states. In each participating school, the study randomly assigned students to either a math class taught by a *TNTP Teaching Fellows* teacher or a similar math class taught by a teacher in the same grade who did not enter teaching through *TNTP Teaching Fellows*. The mean years of teaching experience was 4.0 for *TNTP Teaching Fellows* teachers and 13.0 for comparison teachers. The authors measured mathematics achievement using state-required end-of-year standardized tests for middle school students and study-administered end-of-course assessments for high school students. The analytic sample included 4,116 students (2,127 taught by *TNTP Teaching Fellows* teachers, 1,989 by comparison group teachers) in grades 6–12, during the 2009–10 or 2010–11 school years. Clark et al. (2013) also reported subgroup findings for school levels, years of teaching experience, and comparison group route to certification (traditional or alternative). Appendix D reports these supplemental findings, which do not factor into the intervention's rating of effectiveness.[11]

Summary of studies meeting WWC group design standards with reservations

No studies of *TNTP Teaching Fellows* met WWC group design standards with reservations.

Effectiveness Summary

The WWC review of studies of teachers trained through *TNTP Teaching Fellows* for the Teacher Training, Evaluation, and Compensation topic area includes both student and teacher outcomes. The review covers six domains for student outcomes and 11 domains for teacher outcomes.[12] The one study of *TNTP Teaching Fellows* teachers that met WWC group design standards reported findings in one of the six domains for student outcomes: mathematics achievement. The study did not report any findings that met WWC group design standards in the 11 domains for teacher outcomes. The following findings present the authors' estimates and WWC-calculated estimates of the size and statistical significance of the effects of *TNTP Teaching Fellows* teachers on students in grades 6–12. Additional comparisons are available as supplemental findings in Appendix D. The supplemental findings do not factor into the intervention's rating of effectiveness. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria on p. 21.

Summary of effectiveness for the mathematics achievement domain

Table 3. Rating of effectiveness and extent of evidence for the mathematics achievement domain

Rating of effectiveness	Criteria met
No discernible effects *No affirmative evidence of effects.*	In the one study that reported findings, the estimated impact of the intervention on outcomes in the *mathematics achievement* domain was neither statistically significant nor large enough to be substantively important.
Extent of evidence	**Criteria met**
Small	One study that included 4,116 students in 44 schools reported evidence of effectiveness in the *mathematics achievement domain.*

One study that met WWC group design standards without reservations reported findings in the mathematics achievement domain.

Clark et al. (2013) examined one outcome in the mathematics achievement domain: the authors created an aggregated standardized achievement measure (reported as a *z*-score) based on state-required assessments for students in grades 6–8 and study-administered Northwest Evaluation Association (NWEA) end-of-course assessments for students in grades 9–12. The authors did not find a statistically significant effect of *TNTP Teaching Fellows* teachers on mathematics achievement. The WWC-calculated average effect size was not large enough to be considered substantively important. The WWC characterizes this study finding as an indeterminate effect. Supplemental findings presented in Appendix D do not factor into the intervention's rating of effectiveness. As part of these supplemental findings, Clark et al. (2013) found, and the WWC confirmed, two statistically significant positive effects: (1) students of *TNTP Teaching Fellows* teachers had higher mathematics achievement than students of teachers from less selective alternative routes to certification; and (2) students of novice *TNTP Teaching Fellows* teachers (that is, those in their first 3 years of teaching) outperformed students of novice comparison teachers. The authors also reported, and the WWC confirmed, one statistically significant negative effect: students of novice *TNTP Teaching Fellows* teachers had lower mathematics achievement than students of experienced comparison teachers (that is, those with more than 3 years of experience).

Thus, for the mathematics achievement domain, one study showed an indeterminate effect. This results in a rating of no discernible effects, with a small extent of evidence.

References

Study that meets WWC group design standards without reservations

Clark, M. A., Chiang, H. S., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). *The effectiveness of secondary math teachers from Teach For America and the Teaching Fellows programs* (NCEE 2013-4015). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from https://eric.ed.gov/?&id=ED544171

Studies that meet WWC group design standards with reservations

None.

Studies that do not meet WWC group design standards

Boyd, D., Dunlop, E., Lankford, H., Loeb, S., Mahler, P., O'Brien, R., & Wyckoff, J. (2012, March). *Alternative certification in the long run: A decade of evidence on the effects of alternative certification in New York City*. Paper presented at the 37th Annual Conference of the Association for Education Finance and Policy, Boston, MA. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wycoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy, 1*(2), 176–216. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Additional source:

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2005). *How changes in entry requirements alter the teacher workforce and affect student achievement* (NBER Working Paper 11844). Cambridge, MA: National Bureau of Economic Research.

Boyd, D., Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2007). *The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools* (Working Paper No. 10). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review, 27*(6), 615–631. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Additional sources:

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2006). *What does certification tell us about teacher effectiveness? Evidence from New York City* (NBER Working Paper 12155). Cambridge, MA: National Bureau of Economic Research.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2007). Photo finish: Teacher certification doesn't guarantee a winner. *Education Next, 7*(1).

Ogundimu, C. O. (2014). *Does the mode of entry into teaching matter in teacher retention? A discrete-time survival analysis modeling of New York City public school teachers* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 1534365411) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education Finance and Policy, 6*(1), 43–74. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Additional source:

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2008). *Can you recognize an effective teacher when you recruit one?* (NBER Working Paper 14485). Cambridge, MA: National Bureau of Economics Research.

Tennessee State Board of Education. (2015). *2015 report card on the effectiveness of teacher training programs.* Nashville, TN: Author. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Additional sources:

Tennessee State Board of Education. (2011). *2011 report card on the effectiveness of teacher training programs.* Nashville, TN: Author. Retrieved from ERIC: https://eric.ed.gov/?&id=ED530920

Tennessee State Board of Education. (2012). *2012 report card on the effectiveness of teacher training programs*. Nashville, TN: Author.

Tennessee State Board of Education. (2013). *2013 report card on the effectiveness of teacher training programs.* Nashville, TN: Author.

Tennessee State Board of Education. (2014). *2014 report card on the effectiveness of teacher training programs.* Nashville, TN: Author.

Studies that are ineligible for review using the Teacher Training, Evaluation, and Compensation Evidence Review Protocol

Bisland, B. M., Malow, M. S., & O'Connor, E. A. (2010). Perspectives, constraints and practices in social studies instruction: A case study of alternatively certified elementary teachers. *International Journal of Social Education, 24*(2), 117–144. The study is ineligible for review because it does not use an eligible design.

Bisland, B. M., Malow-Iroff, M. S., & O'Connor, E. A. (2005, November). *Social studies instructional practices among alternatively certified elementary teachers: The New York City Teaching Fellows.* Paper presented at the College and University Faculty Assembly of the National Council for Social Studies, Kansas City, MO. Retrieved from https://eric.ed.gov/?&id=ED490673 The study is ineligible for review because it does not use an eligible design.

Boyd, D., Goldhaber, D., Lankford, H., & Wyckoff, J. (2007). The effect of certification and preparation on teacher quality. *Future of Children, 17*(1), 45–68. Retrieved from https://eric.ed.gov/?&id=EJ795877 The study is ineligible for review because it does not use an eligible design.

Brantlinger, A., & Smith, B. (2013). Alternative teacher certification and the new professionalism: The pre-service preparation of mathematics teachers in the New York City Teaching Fellows program. *Teachers College Record, 115*(7), 1–44. The study is ineligible for review because it does not use an eligible design.

Additional sources:

Meagher, M., Brantlinger, A., and Cooley, L. (2007, October). *Alternative certification in urban school districts: The case of the NYC Teaching Fellows.* Paper presented at the annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Reno, NV. Retrieved from http://www.pmena.org/

Smith, B., Brantlinger, A., & Cooley, L. (2009, September). *Preservice preparation of mathematics candidates in the largest alternative certification program in the United States*. Paper presented at the annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Atlanta, GA.

Cochran-Smith, M., Cannady, M., McEachern, K. P., Mitchell, K., Piazza, P., Power, C., & Ryan, A. (2012). Teachers' education and outcomes: Mapping the research terrain. *Teachers College Record, 114*(10), 1–49. The study is ineligible for review because it does not use an eligible design.

Additional source:

Cochran-Smith, M., Cannady, M., McEachern, K. P., Piazza, P., Power, C., & Ryan, A. (2011). Teachers' education, teaching practice, and retention: A cross-genre review of recent research. *Journal of Education, 191*(2), 19–31.

Cochran-Smith, M., & Villegas, A. M. (2015). Framing teacher preparation research: An overview of the field, part 1. *Journal of Teacher Education, 66*(1), 7–20. The study is ineligible for review because it does not use an eligible design.

Cooley, L. A. (2009, September). *An alternate route to urban mathematics teaching: The NYC Teaching Fellows Program.* Paper presented at the tenth international conference for Mathematics Into the 21st Century, Dresden University of Applied Sciences, Germany. The study is ineligible for review because it does not use an eligible design.

Costigan, A. T. (2004). Finding a name for what they want: A study of New York City's Teaching Fellows. *Teaching and Teacher Education, 20*(2), 129–143. The study is ineligible for review because it does not use an eligible design.

Costigan, A. T. (2005). Choosing to stay, choosing to leave: New York City Teaching Fellows after two years. *Teacher Education Quarterly, 32*(2), 125–142. Retrieved from ERIC: https://eric.ed.gov/?&id=EJ795315 The study is ineligible for review because it does not use an eligible design.

Daly, T., & Keeling, D. (2009). Hiring for teacher quality at the district level: Lessons from the New Teacher Project. In D. H. Gitomer (Ed), *Measurement issues and assessment for teaching quality* (pp. 30–46). Thousand Oaks, CA: Sage. The study is ineligible for review because it does not use an eligible design.

Edgar, B. (2014). *Blackboard bodhisattvas: Narratives of New York City Teaching Fellows* (Unpublished doctoral dissertation). Stanford University, CA. The study is ineligible for review because it does not use an eligible design.

Engelberg, E. S. (2011). *Second year New York City Teaching Fellows: Navigating the gap between vision and practice* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3484381) The study is ineligible for review because it does not use an eligible design.

Evans, B. R. (2010). Secondary mathematics teacher differences: Teacher quality and preparation in a New York City alternative certification program. *Mathematics Educator, 20*(2), 24–32. The study is ineligible for review because it does not use an eligible design.

Evans, B. R. (2011). Content knowledge, attitudes, and self-efficacy in the mathematics New York City Teaching Fellows (NYCTF) program. *School Science and Mathematics, 111*(5), 225–235. The study is ineligible for review because it does not use an eligible design.

Foote, M. Q., Brantlinger, A., Haydar, H. N., Smith, B., & Gonzalez, L. (2011). Are we supporting teacher success: Insights from an alternative route mathematics teacher certification program for urban public schools. *Education and Urban Society, 43*(3), 396–425. The study is ineligible for review because it does not use an eligible design.

Gansle, K. A., Burns, J. M., & Noell, G. (2011). *Value added assessment of teacher preparation programs in Louisiana: 2007-2008 to 2009-2010 overview of 2010-2011 results No. 8*. Baton Rouge: Louisiana Board of Regents. The study is ineligible for review because it does not use an eligible design.

Additional sources:

Burns, J. M., Noell, G. H., & Gansle, K. A. (2009). *Value added teacher preparation assessment model: A bold step forward in preparing, inducting, and supporting new teachers* (Qualitative research study). Baton Rouge: Louisiana Board of Regents.

Gansle, K. A., Burns, J. M., & Noell, G. (2010). *Value added assessment of teacher preparation in Louisiana: 2005-2006 to 2008-2009: Overview of performance bands.* Baton Rouge: Louisiana Board of Regents.

Louisiana Board of Regents. (2013). *Overview of 2011-2012 annual report for teacher preparation.* Baton Rouge: Author.

Noell, G. H., & Burns, J. M. (2007). *Value added teacher preparation assessment: Overview of 2006-07 study*. Baton Rouge: Louisiana Board of Regents.

Noell, G. H., & Burns, J. M. (2008). *Value added teacher preparation assessment: Overview of 2007-08 study*. Baton Rouge: Louisiana Board of Regents.

Noell, G. H., Burns, J. M., & Gansle, K. A. (2009). *Value added teacher preparation in Louisiana: 2005-2006 to 2007-2008: Background & new results.* Baton Rouge: Louisiana Board of Regents.

Grossman, P., & Loeb, S. (2010). Learning from multiple routes. *Educational Leadership, 67*(8), 22–27. The study is ineligible for review because it does not use an eligible design.

Haycock, K., & Crawford, C. (2008). Closing the teacher quality gap. *Educational Leadership, 65*(7), 14–19. The study is ineligible for review because it does not use an eligible design.

Haydar, H., Vatuk, S., & Angulo, N. (2009, September). *Any right to get it wrong? Beginning urban teachers and student mathematical errors.* Paper presented at the annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Atlanta, GA. Retrieved from http://www.pmena.org/ The study is ineligible for review because it does not use an eligible design.

Lincove, J., Osborne, C., Mills, N., & Bellows, L. (2014). Training teachers for profit or prestige: Analysis of a diverse market for teacher preparation. *Journal of Teacher Education, 66*(5), 415–434. The study is ineligible because it does not use an eligible design.

Loeb, S., Miller, L. C., & Wyckoff, J. (2015). Performance screens for school improvement: The case of teacher tenure reform in New York City. *Educational Researcher, 44*(4), 199–212. The study is ineligible for review because it is out of scope of the protocol.

Additional source:

Loeb, S., Miller, L. C., & Wyckoff, J. (2014). *Performance screens for school improvement: The case of teacher tenure reform in New York City* (Working Paper No. 115). Washington, DC: American Institutes of Research, National Center for Analysis of Longitudinal Data in Education Research (CALDER).

Mac Iver, M. A., Katz, G., & Vaughn, E. S. (2006). *A teacher for every classroom: New teachers in the Baltimore City Public Schools, 1999-2005. Brief report update, May 2006.* Baltimore, MD: The Abell Foundation. The study is ineligible for review because it is out of scope of the protocol.

Malow-Iroff, M., O'Connor, E. A., & Bisland, B. M. (2007). Intention to return: Alternatively certified teachers' support, ideology and efficacy beliefs. *Teacher Development, 11*(3), 263–275. The study is ineligible for review because it does not use an eligible design.

National Council on Teacher Quality. (2013). Findings on secondary alternative certification programs. In J. Greenberg, K. Walsh, & A. McKee (Eds.), *2014 teacher prep review.* Washington, DC: Author. The study is ineligible for review because it does not use an eligible design.

O'Connor, E. A., Malow, M. S., & Bisland, B. M. (2011). Mentorship and instruction received during training: Views of alternatively certified teachers. *Educational Review, 63*(2), 219–232. The study is ineligible for review because it does not use an eligible design.

Rivera Maulucci, M. S. (2010). Navigating role forces and the aesthetic|authentic caring dialectic: A novice urban science teacher's developmental trajectory. *Cultural Studies of Science Education, 5*(3), 625–647. The study is ineligible for review because it does not use an eligible design.

Rochkind, J., Ott, A., Immerwahr, J., Doble, J., & Johnson, J. (2007). *Lessons learned: New teachers talk about their jobs, challenges and long-range plans. Issue no. 2. Working without a net: How new teachers from three prominent alternate route programs describe their first year on the job*. New York, NY: National Comprehensive Center for Teacher Quality and Public Agenda. Retrieved from https://eric.ed.gov/?&id=ED501569 The study is ineligible for review because it is out of scope of the protocol.

Schonfeld, I. S., & Feinman, S. J. (2012). Difficulties of alternatively certified teachers. *Education and Urban Society, 44*(3), 215–246. The study is ineligible for review because it is out of scope of the protocol.

Shuls, J. V., & Trivitt, J. R. (2015). Teacher effectiveness: An analysis of licensure screens. *Educational Policy, 29*(4), 645–675. The study is ineligible for review because it is out of scope of the protocol.

Silin, J. G., Snyder, J., Barry, M., Samuels, S., Sacks, A., & Ellenzweig, A. (2008). *Alternative routes to teacher certification* (Occasional Paper Series 20). New York, NY: Bank Street College of Education. The study is ineligible for review because it does not use an eligible design.

TNTP. (2009). *The New Teacher Project.* New York, NY: Author. The study is ineligible for review because it does not use an eligible design.

TNTP. (2014). *School leadership practices across a high-performing charter sector TNTP.* New York, NY: Author. Retrieved from http://tntp.org/ The study is ineligible for review because it does not use an eligible design.

Tyler, J., Jacob, B., Dougherty, S., Hanson, H., Fullerton, J., & Herlihy, C. (2012). *Are practice-based teacher evaluations and teacher effectiveness linked in TNTP's Performance Assessment System?* Cambridge, MA: Center for Education Policy Research, Harvard University. Retrieved from http://cepr.harvard.edu/ The study is ineligible for review because it does not use an eligible design.

Vatuk, S., & Meagher, M. (2009). The Mathematical Knowledge for Teaching (MKT) of New York City Teaching Fellows. In S. L. Swars, D. W. Stinson, & S. Lemons-Smith (Eds.), *Proceedings from the 31st Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education.* Atlanta: Georgia State University. The study is ineligible for review because it does not use an eligible design.

Appendix A: Research details for Clark et al. (2013)

Clark, M. A., Chiang, H. S., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). *The effectiveness of secondary math teachers from Teach For America and the Teaching Fellows programs* (NCEE 2013-4015). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from https://eric.ed.gov/?&id=ED544171

Table A. Summary of findings

Meets WWC group design standards without reservations

Outcome domain	Sample size	Study findings	
		Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	153 teachers/4,116 students	0	No

Setting

The study was conducted in 44 secondary schools in nine school districts in eight states.[13]

Study sample

The study included two cohorts of students in grades 6–12: one that participated in the 2009–10 school year, and one that participated in the 2010–11 school year. In each participating school, students were randomly assigned within "classroom matches" to either a class taught by a *TNTP Teaching Fellows* teacher or a class taught by a comparison teacher. A classroom match consisted of two or more classes covering the same eligible middle or high school math courses that were deemed comparable by the study authors based on factors such as level (for example, honors or regular), length (one or two semesters), and arrangements made for the inclusion of English learners and special education students.[14] After 7,288 students (3,659 *TNTP Teaching Fellows*, 3,629 comparison) were randomly assigned, attrition occurred due to students leaving the school prior to the start of the school year, lack of parental consent, or students not having valid end-of-year mathematics achievement scores. The analytic sample included 4,116 students (2,127 *TNTP Teaching Fellows*, 1,989 comparison) taught by 153 teachers (69 *TNTP Teaching Fellows*, 84 comparison) in 44 schools. The mean age of the students was 14.3 years.[15] Among the sample, 60% of students were in grades 9–12, 54% were female, 75% were eligible for free or reduced-price lunch, 7% were limited English proficient, and 6% had an individualized education plan. The racial/ethnic demographics were as follows: 50% were Black, 36% were Hispanic, 9% were Asian, 5% were White, and 1% were another race/ethnicity.

In addition, the authors present subgroup findings for school levels (middle or high school), years of teaching experience, and comparison group teachers' route to certification (traditional or less selective alternative). The years of teaching experience comparisons include: (a) *TNTP Teaching Fellows* teachers in their first 3 years of teaching vs. non-*TNTP Teaching Fellows* teachers in their first 3 years of teaching, (b) *TNTP Teaching Fellows* teachers in their first 3 years of teaching vs. non-*TNTP Teaching Fellows* teachers with more than 3 years of experience, (c) *TNTP Teaching Fellows* teachers with more than 3 years of experience vs. non-*TNTP Teaching Fellows* teachers with more than 3 years of experience, and (d) *TNTP Teaching Fellows* teachers vs. non-*TNTP Teaching Fellows* teachers whose levels of teaching experience differ by no more than 2 years.[16] The subgroup findings are reported in Appendix D. The supplemental findings do not factor into the intervention's rating of effectiveness.

Intervention group

Students were taught by *TNTP Teaching Fellows* teachers. The mean years of teaching experience at the end of the study year was 4.0. Among *TNTP Teaching Fellows* teachers, 72% had a bachelor's degree from a most, highly, or very competitive college or university; 25% majored in math, none majored in secondary math education, and 33% majored in other math-related subjects.[17] Regarding math content knowledge, the mean score was 158 among teachers who took the Praxis II Mathematics Content Knowledge Test (0.80 standard deviations higher than comparison teachers) and 187 among teachers who took the Praxis II Middle School Mathematics Test (0.92 standard deviations higher than comparison teachers). The mean age of *TNTP Teaching Fellows* teachers at the time of the study was 33.3 years, and 54% of *TNTP Teaching Fellows* teachers were female, 71% were White, 17% were Black, 9% were Hispanic, and 9% were Asian. The authors did not report any deviations from the *TNTP Teaching Fellows* model.

Comparison group

Students in the comparison group were taught by teachers who did not enter teaching through *TNTP Teaching Fellows*, *Teach For America*, or other highly selective alternative routes to certification. The majority (73%) of comparison teachers entered teaching through a traditional route to certification (that is, they became certified teachers after completing a standard postsecondary program for teaching and related certification requirements), with the remainder entering through a less selective alternative route. The mean years of teaching experience at the end of the study year was 13.0. Among comparison teachers, 34% had a bachelor's degree from a most, highly, or very competitive college or university; 43% majored in math, 13% majored in secondary math education, and 23% majored in other math-related subjects. Regarding math content knowledge, the mean score was 139 among teachers who took the Praxis II Mathematics Content Knowledge Test and 170 among teachers who took the Praxis II Middle School Mathematics Test. The mean age of comparison teachers at the time of the study was 41.0 years, and 57% of comparison teachers were female, 43% were White, 36% were Black, 19% were Asian, and 13% were Hispanic.

Outcomes and measurement

An outcome in the mathematics achievement domain was reported. All assessment scores were converted into *z*-scores, thus providing a single outcome for the analysis that expressed mathematics achievement in standard deviation units. For students in grades 6–8, study authors obtained scores from state-required assessments administered in the spring semester of the school year in which the students were randomly assigned. For students in grades 9–12, study authors administered end-of-course math assessments. For a more detailed description of these outcome measures, see Appendix B. The study also examined measures of student absences and teacher job satisfaction; these outcomes are ineligible for review because they are not within a domain specified in the Teacher Training, Evaluation, and Compensation protocol.

Support for implementation

Training provided to *TNTP Teaching Fellows* participants prior to their becoming classroom teachers consists of about 25 hours of independent study and a 4-hour orientation followed by an intensive 5- to 7-week summer institute that includes practice teaching in public summer school classrooms, coursework led by program and district staff, and program staff providing feedback after evaluating participants' teaching performance. Of the eight *TNTP Teaching Fellows* programs in the study, three also provided a review of mathematical concepts in intensive summer "math immersion" programs for participants who otherwise might be ineligible to teach secondary math (for example, participants who lacked sufficient college math credits). After program participants begin teaching, *TNTP Teaching Fellows* staff provide about 10 hours of professional development in group sessions on topics such as classroom management, using data to inform instruction, and tailoring instruction for different students; conduct at least two formal classroom observations of each new teacher; hold at least two one-on-one meetings with each new teacher; and engage in informal check-in discussions or offer other support as needed. *TNTP Teaching Fellows* teachers also enrolled in local, state-authorized programs to complete the coursework required for certification.

Appendix B: Outcome measure for the mathematics achievement domain

Mathematics achievement	
Mathematics assessments	Clark et al. (2013) used state-required math assessments for students in grades 6–8 and study-administered Northwest Evaluation Association (NWEA) end-of-course math assessments for students in grades 9–12. The state-required assessments were criterion-referenced tests. The tests differed across states, with each test being part of the state's accountability system. Each score was converted to a *z*-score using as a reference population the full population of students in the same state, year, and grade who took the same assessment. The NWEA assessments were computer-adaptive tests administered in general high school math, Algebra I, Geometry, or Algebra II, depending on the content of the student's math course. The administration and scoring of the tests for the study differed from standard NWEA procedures, in that the study authors imposed a 35-minute time limit and obtained scores for incomplete tests. The study authors reported marginal reliability coefficients of .927 or greater for the analytic sample. Each score was converted to a *z*-score using the NWEA's nationwide norming sample for the assessment as the reference population (as cited in Clark et al., 2013).

Appendix C: Findings included in the rating for the mathematics achievement domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Clark et al. (2013)[a]								
Mathematics assessments	All teachers	153 teachers/ 4,116 students	–0.39 (1.12)	–0.39 (1.02)	0.00	0.00	0	.96
Domain average for mathematics achievement (Clark et al., 2013)						**0.00**	**0**	**Not statistically significant**
Domain average for mathematics achievement across all studies						**0.00**	**0**	**na**

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. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. The statistical significance of the study's domain average was determined by the WWC. Some statistics may not sum as expected due to rounding. na = not applicable.

[a] For Clark et al. (2013), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The *p*-value presented here was reported in the original study. The study authors calculated the intervention group mean by adding the impact of the intervention (the regression-adjusted difference between the intervention and comparison groups) to the unadjusted comparison group mean. The unadjusted standard deviations were provided by the study authors at the WWC's request. This study is characterized as having an indeterminate effect because the estimated effect for the one measure in this domain is neither statistically significant nor substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

Appendix D.1: Supplemental school level findings for the mathematics achievement domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Clark et al. (2013)[a]								
Mathematics assessments	Middle school teachers	53 teachers/ 1,610 students	–0.35 (0.88)	–0.39 (0.80)	0.04	0.05	+2	.38
Mathematics assessments	High school teachers	101 teachers/ 2,506 students	–0.41 (1.22)	–0.39 (1.12)	–0.02	–0.02	–1	.47

Table Notes: The supplemental findings presented in this table are additional findings from studies in this report that meet WWC design standards with or without reservations, but do not factor into the determination of the intervention rating. 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. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. Some statistics may not sum as expected due to rounding.

[a] For Clark et al. (2013), a correction for multiple comparisons was needed but did not affect whether any of the contrasts were found to be statistically significant. The *p*-values presented here were reported in the original study. The study authors calculated the intervention group mean by adding the impact of the intervention (the regression-adjusted difference between the intervention and comparison groups) to the unadjusted comparison group mean. The unadjusted standard deviations were provided by the study authors at the WWC's request. A single study review of Clark et al. (2013) was released in May 2014 and modified in September 2015. Some of the effect sizes reported in the single study review differ from the effect sizes reported in this table because the WWC calculated the effect sizes in this intervention report using unadjusted standard deviations provided by the study authors at the WWC's request. The single study review and intervention report effect sizes differ by no more than 0.01 standard deviations. In addition, the single study review incorrectly reported the intervention group mean as –0.47 for the high school subgroup.

Appendix D.2: Supplemental teacher subgroup findings for the mathematics achievement domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Clark et al. (2013)[a]								
Mathematics assessments	*TNTP Teaching Fellows* teachers and comparison teachers from traditional certification routes	113 teachers/ 3,268 students	–0.36 (1.15)	–0.32 (0.99)	–0.03	–0.03	–1	.25
Mathematics assessments	*TNTP Teaching Fellows* teachers and comparison teachers from less selective alternative certification routes	46 teachers/ 902 students	–0.50 (0.98)	–0.63 (1.10)	0.13	0.12	+5	.01
Mathematics assessments	Novice teachers	17 teachers/ 354 students	–0.40 (0.81)	–0.53 (0.83)	0.13	0.16	+6	< .01
Mathematics assessments	Novice *TNTP Teaching Fellows* teachers and experienced comparison teachers	53 teachers/ 1,153 students	–0.63 (1.24)	–0.53 (1.08)	–0.10	–0.09	–3	< .01
Mathematics assessments	Experienced teachers	80 teachers/ 2,408 students	–0.27 (1.10)	–0.30 (1.00)	0.03	0.03	+1	.45
Mathematics assessments	Teachers with similar years of experience	46 teachers/ 1,283 students	–0.17 (0.96)	–0.20 (0.91)	0.03	0.03	+1	.40

Table Notes: The supplemental findings presented in this table are additional findings from studies in this report that meet WWC design standards with or without reservations, but do not factor into the determination of the intervention rating. 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. The effect size is a standardized measure of the effect of an intervention on outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. Some statistics may not sum as expected due to rounding.

[a] For Clark et al. (2013), a correction for multiple comparisons was needed but did not affect whether any of the contrasts were found to be statistically significant. The *p*-values presented here were reported in the original study. The study authors calculated the intervention group mean by adding the impact of the intervention (the regression-adjusted difference between the intervention and comparison groups) to the unadjusted comparison group mean. The study defines novice teachers as those in their first 3 years of teaching. Experienced teachers are those with more than 3 years of teaching experience. The study authors categorized *TNTP Teaching Fellows* and comparison teachers as having similar years of experience if their levels of teaching experience differed by no more than 2 years. The unadjusted standard deviations were provided by the study authors at the WWC's request. A single study review of Clark et al. (2013) was released in May 2014 and modified in September 2015. Some of the effect sizes and two improvement index values reported in the single study review differ from the effect sizes reported in this table because the WWC calculated the effect sizes in this intervention report using unadjusted standard deviations provided by the study authors at the WWC's request. The single study review and intervention report effect sizes differ by no more than 0.03 standard deviations. The improvement index for the subgroup analysis comparing novice teachers changed from +5 to +6. The improvement index for the subgroup analysis comparing novice *TNTP Teaching Fellows* teachers to experienced comparison teachers changed from –4 to –3.

Endnotes

[1] The descriptive information for this intervention comes from publicly available sources: intervention websites (tntpteachingfellows.org, tntp.org, and www.nycteachingfellows.org, downloaded September 2016) and the research literature (Clark et al., 2013). The What Works Clearinghouse (WWC) requests developers review the intervention description sections for accuracy from their perspective. The WWC provided the developer with the intervention description in November 2016, and the WWC incorporated feedback from the developer. Further verification of the accuracy of the descriptive information for this intervention is beyond the scope of this review.

[2] The exact nature of the support *TNTP Teaching Fellows* provides its teachers has varied over time and across local programs.

[3] Reviews of the studies in this report used the standards from the WWC Procedures and Standards Handbook (version 3.0) and the Teacher Training, Evaluation, and Compensation review protocol (version 3.2). The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

[4] The literature search reflects documents publicly available as of July 2016. The WWC released a single study review of Clark et al. (2013) in May 2014 and modified it in September 2015. Some of the effect sizes and two improvement index values reported in the single study review differ from the effect sizes reported in this intervention report because the WWC calculated the effect sizes in this intervention report using unadjusted standard deviations provided by the study authors at the WWC's request. The single study review and intervention report effect sizes differ by no more than 0.03 standard deviations. The improvement index for the subgroup analysis comparing *TNTP Teaching Fellows* teachers in their first 3 years of teaching to non-*TNTP Teaching Fellows* teachers in their first 3 years of teaching changed from +5 to +6. The improvement index for the subgroup analysis comparing *TNTP Teaching Fellows* teachers in their first 3 years of teaching to non-*TNTP Teaching Fellows* teachers with more than 3 years of experience changed from –4 to –3. Both the single study review and this intervention report characterize the study as having an indeterminate effect in the mathematics achievement domain. In addition, the single study review incorrectly reported the intervention group mean as –0.47 for the high school subgroup; this intervention report presents the correct value of –0.41 in Appendix D.1.

[5] Absence of conflict of interest: This intervention report includes a study conducted by staff from Mathematica Policy Research. Because Mathematica Policy Research is one of the contractors that administers the WWC, staff members from a different organization reviewed the study. The lead methodologist, a WWC quality assurance reviewer, and an external peer reviewer reviewed this report.

[6] Clark et al. (2013) did not name the school districts or states included in the study.

[7] Please see the Teacher Training, Evaluation, and Compensation review protocol (version 3.2) for a list of all outcome domains.

[8] For criteria used to determine the rating of effectiveness and extent of evidence, see the WWC Rating Criteria on p. 21. These improvement index numbers show the average and range of individual-level improvement indices for all findings across the studies.

[9] *TNTP Teaching Fellows* defines *admitted* applicants as those applicants who began pre-service training.

[10] Differences between intervention and comparison group teachers in background characteristics (for example, demographics and educational background) might reflect the type of teacher that *TNTP Teaching Fellows* attracts and selects. In other words, teachers' background characteristics could be considered part of the intervention.

[11] In a sensitivity analysis, Clark et al. (2013) also presented complier average causal effect estimates of the effectiveness of *TNTP Teaching Fellows* teachers. The authors reported that the findings from this sensitivity analysis were consistent with the findings from their main analysis; in both analyses, the authors did not find a statistically significant effect of *TNTP Teaching Fellows* teachers on mathematics achievement, and they estimated an effect size of zero.

[12] Please see the Teacher Training, Evaluation, and Compensation review protocol (version 3.2) for a list of all outcome domains.

[13] Clark et al. (2013) contained two studies examining the effectiveness of teachers from two different interventions, *Teach For America* and *TNTP Teaching Fellows*. This report reviews findings for only the *TNTP Teaching Fellows* study.

[14] Eligible math courses included any middle school math course and the following high school courses: general math (for example, pre-algebra or remedial math), Algebra I, Algebra II, and Geometry.

[15] These sample characteristics are the simple average of the characteristics that the authors reported separately for the intervention and comparison groups. The differences were statistically significant for the following student characteristics: age (14.31 years for the *TNTP Teaching Fellows* group vs. 14.27 years for the comparison group, $p = .005$), percentage eligible for free or reduced-price lunch (73.7% for the *TNTP Teaching Fellows* group vs. 75.9% for the comparison group, $p = .017$), percentage Black (50.4% for the *TNTP Teaching Fellows* group vs. 48.8% for the comparison group, $p = .047$), and percentage Hispanic (34.9% for the *TNTP Teaching Fellows* group vs. 36.8% for the comparison group, $p = .038$). The *TNTP Teaching Fellows* and comparison groups differed by less than two percentage points for each of the remaining demographic characteristics; none of the differences were statistically significant.

[16] Clark et al. (2013) analyzed a fifth teacher experience subgroup: *TNTP Teaching Fellows* teachers with more than 3 years of experience vs. non-*TNTP Teaching Fellows* teachers in their first 3 years of teaching. However, the authors did not report findings for this analysis due to small sample sizes.

[17] Other math-related subjects included statistics, engineering, computer science, finance, economics, physics, and astrophysics. College competitiveness was defined based on *Barron's Profiles of American Colleges 2003*.

Recommended Citation

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2017, June). *Teacher Training, Evaluation, and Compensation intervention report: TNTP Teaching Fellows.* Retrieved from https://whatworks.ed.gov

WWC Rating Criteria

Criteria used to determine the rating of a study

Study rating	Criteria
Meets WWC group design standards without reservations	A study that provides strong evidence for an intervention's effectiveness, such as a well-implemented RCT.
Meets WWC group design standards with reservations	A study that provides weaker evidence for an intervention's effectiveness, such as a QED or an RCT with high attrition that has established equivalence of the analytic samples.

Criteria used to determine the rating of effectiveness for an intervention

Rating of effectiveness	Criteria
Positive effects	Two or more studies show statistically significant positive effects, at least one of which met WWC group design standards for a strong design, AND No studies show statistically significant or substantively important negative effects.
Potentially positive effects	At least one study shows a statistically significant or substantively important positive effect, AND No studies show a statistically significant or substantively important negative effect AND fewer or the same number of studies show indeterminate effects than show statistically significant or substantively important positive effects.
Mixed effects	At least one study shows a statistically significant or substantively important positive effect AND at least one study shows a statistically significant or substantively important negative effect, but no more such studies than the number showing a statistically significant or substantively important positive effect, OR At least one study shows a statistically significant or substantively important effect AND more studies show an indeterminate effect than show a statistically significant or substantively important effect.
Potentially negative effects	One study shows a statistically significant or substantively important negative effect and no studies show a statistically significant or substantively important positive effect, OR Two or more studies show statistically significant or substantively important negative effects, at least one study shows a statistically significant or substantively important positive effect, and more studies show statistically significant or substantively important negative effects than show statistically significant or substantively important positive effects.
Negative effects	Two or more studies show statistically significant negative effects, at least one of which met WWC group design standards for a strong design, AND No studies show statistically significant or substantively important positive effects.
No discernible effects	None of the studies shows a statistically significant or substantively important effect, either positive or negative.

Criteria used to determine the extent of evidence for an intervention

Extent of evidence	Criteria
Medium to large	The domain includes more than one study, AND The domain includes more than one school, AND The domain findings are based on a total sample size of at least 350 students, OR, assuming 25 students in a class, a total of at least 14 classrooms across studies.
Small	The domain includes only one study, OR The domain includes only one school, OR The domain findings are based on a total sample size of fewer than 350 students, AND, assuming 25 students in a class, a total of fewer than 14 classrooms across studies.

Glossary of Terms

Attrition Attrition occurs when an outcome variable is not available for all subjects initially assigned to the intervention and comparison groups. If a randomized controlled trial (RCT) or regression discontinuity design (RDD) study has high levels of attrition, the validity of the study results can be called into question. An RCT with high attrition cannot receive the highest rating of *Meets WWC Group Design Standards without Reservations*, but can receive a rating of *Meets WWC Group Design Standards with Reservations* if it establishes baseline equivalence of the analytic sample. Similarly, the highest rating an RDD with high attrition can receive is Meets *WWC RDD Standards with Reservations*.

For single-case design research, attrition occurs when an individual fails to complete all required phases or data points in an experiment, or when the case is a group and individuals leave the group. If a single-case design does not meet minimum requirements for phases and data points within phases, the study cannot receive the highest rating of *Meets WWC Pilot Single-Case Design Standards without Reservations*.

Baseline A point in time before the intervention was implemented in group design research and in regression discontinuity design studies. When a study is required to satisfy the baseline equivalence requirement, it must be done with characteristics of the analytic sample at baseline. In a single-case design experiment, the baseline condition is a period during which participants are not receiving the intervention.

Clustering adjustment An adjustment to the statistical significance of a finding when the units of assignment and analysis differ. When random assignment is carried out at the cluster level, outcomes for individual units within the same clusters may be correlated. When the analysis is conducted at the individual level rather than the cluster level, there is a mismatch between the unit of assignment and the unit of analysis, and this correlation must be accounted for when assessing the statistical significance of an impact estimate. If the correlation is not accounted for in a mismatched analysis, the study may be too likely to report statistically significant findings. To fairly assess an intervention's effects, in cases where study authors have not corrected for the clustering, the WWC applies an adjustment for clustering when reporting statistical significance.

Confounding factor 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.

Design The method by which intervention and comparison groups are assigned (group design and regression discontinuity design) or the method by which an outcome measure is assessed repeatedly within and across different phases that are defined by the presence or absence of an intervention (single-case design). Designs eligible for WWC review are randomized controlled trials, quasi-experimental designs, regression discontinuity designs, and single-case designs.

Effect size 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.

Eligibility A study is eligible for review and inclusion in this report if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.

Equivalence A demonstration that the analysis sample groups are similar on observed characteristics defined in the review area protocol.

Extent of evidence An indication of how much evidence from group design studies supports the findings in an intervention report. The extent of evidence categorization for intervention reports focuses on the number and sizes of studies of the intervention in order to give an indication of how broadly findings may be applied to different settings. There are two extent of evidence categories: small and medium to large.

- **small:** includes only one study, or one school, or findings based on a total sample size of less than 350 students and 14 classrooms (assuming 25 students in a class)
- **medium to large:** includes more than one study, more than one school, and findings based on a total sample of at least 350 students or 14 classrooms

Gain scores The result of subtracting the pretest from the posttest for each individual in the sample. Some studies analyze gain scores instead of the unadjusted outcome measure as a method of accounting for the baseline measure when estimating the effect of an intervention. The WWC reviews and reports findings from analyses of gain scores, but gain scores do not satisfy the WWC's requirement for a statistical adjustment under the baseline equivalence requirement. This means that a study that must satisfy the baseline equivalence requirement and has baseline differences between 0.05 and 0.25 standard deviations *Does Not Meet WWC Group Design Standards* if the study's only adjustment for the baseline measure was in the construction of the gain score.

Group design A study design in which outcomes for a group receiving an intervention are compared to those for a group not receiving the intervention. Comparison group designs eligible for WWC review are randomized controlled trials and quasi-experimental designs.

Improvement index Along a percentile distribution of individuals, the improvement index represents the gain or loss of the average individual due to the intervention. As the average individual starts at the 50th percentile, the measure ranges from –50 to +50.

Intervention An educational program, product, practice, or policy aimed at improving student outcomes.

Intervention report A summary of the findings of the highest-quality research on a given program, product, practice, or policy in education. The WWC searches for all research studies on an intervention, reviews each against design standards, and summarizes the findings of those that meet WWC design standards.

Multiple comparison adjustment An adjustment to the statistical significance of results to account for multiple comparisons in a group design study. The WWC uses the Benjamini-Hochberg (BH) correction to adjust the statistical significance of results within an outcome domain when study authors perform multiple hypothesis tests without adjusting the p-value. The BH correction is used in three types of situations: studies that tested multiple outcome measures in the same outcome domain with a single comparison group; studies that tested a given outcome measure with multiple comparison groups; and studies that tested multiple outcome measures in the same outcome domain with multiple comparison groups. Because repeated tests of highly correlated constructs will lead to a greater likelihood of mistakenly concluding that the impact was different from zero, in all three situations, the WWC uses the BH correction to reduce the possibility of making this error. The WWC makes separate adjustments for primary and secondary findings.

Outcome domain A group of closely-related outcomes. A domain is the organizing construct for a set of related outcomes through which studies claim effectiveness.

Quasi-experimental design (QED) A quasi-experimental design (QED) is a research design in which study participants are assigned to intervention and comparison groups through a process that is not random.

Randomized controlled trial (RCT) A randomized controlled trial (RCT) is an experiment in which eligible study participants are randomly assigned to intervention and comparison groups.

Rating of effectiveness For group design research, the WWC rates the effectiveness of an intervention in each domain based on the quality of the research design and the magnitude, statistical significance, and consistency in findings. For single-case design research, the WWC rates the effectiveness of an intervention in each domain based on the quality of the research design and the consistency of demonstrated effects. The criteria for the ratings of effectiveness are given in the WWC Rating Criteria on p. 21.

Regression discontinuity design (RDD) A design in which groups are created using a continuous scoring rule. For example, students may be assigned to a summer school program if they score below a preset point on a standardized test, or schools may be awarded a grant based on their score on an application. A regression line or curve is estimated for the intervention group and similarly for the comparison group, and an effect occurs if there is a discontinuity in the two regression lines at the cutoff.

Single-case design 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.

Standard deviation 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 tend to be spread out over a large range of values.

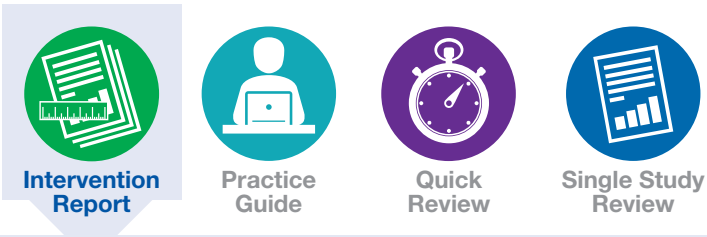
Statistical significance Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p < .05$).

Study rating The result of the WWC assessment of a study. The rating is based on the strength of the evidence of the effectiveness of the educational intervention. Studies are given a rating of *Meets WWC Design Standards without Reservations*, *Meets WWC Design Standards with Reservations*, or *Does Not Meet WWC Design Standards*, based on the assessment of the study against the appropriate design standards. The WWC has design standards for group design, single-case design, and regression discontinuity design studies.

Substantively important A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

Systematic review A review of existing literature on a topic that is identified and reviewed using explicit methods. A WWC systematic review has five steps: 1) developing a review protocol; 2) searching the literature; 3) reviewing studies, including screening studies for eligibility, reviewing the methodological quality of each study, and reporting on high quality studies and their findings; 4) combining findings within and across studies; and, 5) summarizing the review.

Please see the WWC Procedures and Standards Handbook (version 3.0) for additional details.



An **intervention report** summarizes the findings of high-quality research on a given program, practice, or policy in education. The WWC searches for all research studies on an intervention, reviews each against evidence standards, and summarizes the findings of those that meet standards.

This intervention report was prepared for the WWC by Mathematica Policy Research under contract ED-IES-13-C-0010.