What Works Clearinghouse™
ies INSTITUTE OF EDUCATION SCIENCES

Teacher Training, Evaluation, and Compensation February 2018

National Board for Professional Teaching Standards Certification

Intervention Description[1]

The *National Board for Professional Teaching Standards* (*NBPTS*) establishes standards for accomplished teachers and awards professional certification to teachers who can demonstrate that their teaching practices meet those standards. Educators and experts in child development and related fields established the organization, and these experts work to develop and refine the standards for accomplished teaching based on the knowledge and skills that effective teachers demonstrate. The standards reflect five core propositions: (1) effective teachers are committed to students and their learning, (2) effective teachers know the subjects they teach and how to teach those subjects to students, (3) effective teachers manage and monitor student learning, (4) effective teachers think systematically about their practice and learn from experience, and (5) effective teachers are members of learning communities. Those seeking certification from the *NBPTS* must complete a computer-based assessment and three portfolio entries. The certification process can take 1 to 5 years.[2]

Research[3]

The What Works Clearinghouse (WWC) identified five studies of *NBPTS* certification that both fall within the scope of the Teacher Training, Evaluation, and Compensation topic area and meet WWC group design standards. No studies meet WWC group design standards without reservations, and five studies meet WWC group design standards with reservations. Together, these studies included more than 1,316,146 elementary and middle school students in grades 3 to 8 in four states.[4]

According to the WWC review, the extent of evidence for teachers who obtained *NBPTS* certification on the academic achievement of elementary and middle school students was medium to large for two student outcome domains—English language arts achievement and mathematics achievement. No studies meet WWC group design standards in the four other student outcome domains or the 11 teacher outcome domains, so this intervention report does not report on the effectiveness of *NBPTS*-certified teachers for those domains.[5] (See the Effectiveness Summary on p. 6 for more details of effectiveness by domain.)

Effectiveness

NBPTS-certified teachers had mixed effects on mathematics achievement and no discernible effects on English language arts achievement for students in grades 3 through 8.

Report Contents

This intervention report presents findings from a systematic review of *National Board for Professional Teaching Standards Certification* 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[6]

Outcome domain	Rating of effectiveness	Improvement index *(percentile points)*		Number of studies	Number of students	Extent of evidence
		Average	Range			
Mathematics achievement	Mixed effects	+1	0 to +2	3	1,316,146	Medium to large
English language arts achievement	No discernible effects	+2	0 to +4	4	1,242,454	Medium to large

Intervention Information

Background

The *NBPTS* was founded in 1987. The organization continues to update the standards and award certifications. Address: 1525 Wilson Blvd., Ste. 700, Arlington, VA 22209. Web: http://www.nbpts.org/. Telephone: (703) 465-2700.

Intervention details

The *NBPTS* offers certificates in 16 content areas for teachers working in pre-K through grade 12. For many of the content areas, certificates are available for students in different age groups. In general, to be eligible for certification, a teacher must hold a bachelor's degree and a valid state teaching license, and must have completed 3 years of teaching. Requirements vary for teachers pursuing the Career and Technical Education, School Counseling, and World Language certifications.

The certification process includes tasks associated with each of four components: (1) content knowledge, (2) differentiation in instruction, (3) teaching practice and learning environment, and (4) effective and reflective practitioner. Candidates receive an assessment score for each component. To achieve certification, candidates must achieve or exceed the minimum individual scores for each component and a minimum combined score across the four components. Candidates select the components they choose to attempt in a given year, must complete a first attempt at all components within 3 years, and have up to 5 years to achieve the required minimum scores for all components. Those who do not attain the minimum score(s) can retake components up to two times within that time frame.

The first component, content knowledge, is assessed through a computer-administered test consisting of three constructed response exercises and 45 multiple-choice items, specific to each certification area. The content knowledge assessment takes a minimum of 2.5 hours to complete. The differentiation in instruction component is assessed via a written reflection on students' work and includes a collection of students' work and a commentary connecting the teacher's instructional choices to students' growth. The teaching practice and learning environment component is assessed via a written self-reflective analysis of teaching practice. Scores for this component are based on video recordings of teachers' interactions with their students and the teachers' written analyses of those interactions. To demonstrate the effective and reflective practitioner component, candidate teachers must document their knowledge and use of assessment and their collaboration with families and colleagues, and they must comment on how those activities affected students' learning.

Teachers who obtained *NBPTS* certification before 2017 must fulfill certain requirements to renew their certification every 10 years. This process requires demonstrating professional growth through recordings of teaching and students' work, as well as a written analysis of teaching practices and plans for continued professional growth. Those certified in 2017 and after will be required to maintain their certification every 5 years.

Cost

As of April 2017, *NBPTS* certification candidates pay a $75 registration fee and $475 for each of the four components of certification; thus, the total minimum cost for certification is $1,975. Additional fees apply for candidates who have to repeat requirements to complete a component or change a certification area during the application process. For teachers certified before 2017, the fee for certification renewal is $1,250. Some states and localities provide subsidies to cover part of the cost of certification. Many states and school districts offer salary increases or bonuses for teachers who become certified through the *NBPTS*.

Research Summary

The WWC identified 39 eligible studies that investigated the effects of *NBPTS*-certified teachers on academic achievement for elementary and middle school students. An additional 109 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 148 studies are in the References section, which begins on p. 9.

Table 2. Scope of reviewed research

Grades	3–8
Delivery method	Whole class
Intervention type	Teacher level

The WWC reviewed 38 eligible studies against group design standards. No studies are randomized controlled trials that meet WWC group design standards without reservations, and five studies use quasi-experimental designs that meet WWC group design standards with reservations. This report summarizes those five studies. The remaining 33 studies do not meet WWC group design standards.

The WWC reviewed one eligible study against pilot regression discontinuity design standards. This study does not meet WWC pilot regression discontinuity design standards.

Summary of studies meeting WWC group design standards without reservations

No studies of the effects of *NBPTS*-certified teachers meet WWC group design standards without reservations.

Summary of studies meeting WWC group design standards with reservations

Cowan and Goldhaber (2016) examined the effectiveness of *NBPTS*-certified teachers compared with other teachers in their schools using a quasi-experimental design in elementary and middle schools in Washington state. The authors compared the academic achievement of students receiving instruction from an *NBPTS*-certified teacher with those receiving instruction from a non–*NBPTS*-certified teacher. The authors measured mathematics and English language arts achievement using state-required end-of-year standardized tests. The analytic sample (that is, the sample used for study analysis) included 1,312,657 students (110,634 taught by *NBPTS*-certified teachers and 1,202,023 taught by comparison group teachers) for the mathematics achievement domain and 1,234,924 students (113,129 taught by *NBPTS*-certified teachers and 1,121,795 taught by comparison group teachers) for the English language arts achievement domain in grades 4–8, from the 2005–06 to 2012–13 school years. Because the authors examined achievement across multiple school years, the reported sample sizes may count some individual students more than once. Cowan and Goldhaber (2016) also reported subgroup findings for school level, certification subject area, English learners, students receiving special education, students eligible for free or reduced-price lunch, and schools with low prior achievement. In addition, they reported subgroup findings for what they described as "apparently random samples" of these same groups of students, in which there was no evidence of students being sorted into particular classrooms based on demographic characteristics. Appendix D reports these supplemental findings, which do not factor into the intervention's rating of effectiveness.

Fisher and Dickenson (2005) examined the effectiveness of *NBPTS*-certified teachers compared with other teachers using a quasi-experimental design in elementary and middle schools across South Carolina. The authors compared the academic achievement of students receiving instruction from an *NBPTS*-certified teacher with those receiving instruction from a non–*NBPTS*-certified teacher. The authors measured mathematics and English language arts achievement using state-required end-of-year standardized tests. Depending on the grade taught, *NBPTS*-certified teachers had an average of between 13.7 to 17.8 years of experience, whereas comparison group teachers had an average of between 10.4 to 14.1 years of experience. The analytic sample included 3,336 students (1,668 taught by *NBPTS*-certified teachers and 1,668 taught by comparison group teachers) for the mathematics achievement domain and 3,938 students (1,969 taught by *NBPTS*-certified teachers and 1,969 taught by comparison group

teachers) for the English language arts achievement domain in grades 4–8, during the 2003–04 school year. Fisher and Dickenson (2005) also reported subgroup findings for individual grades and by free or reduced-price lunch eligibility status. Appendix D reports these supplemental findings, which do not factor into the intervention's rating of effectiveness.

Gardner (2010) examined the effectiveness of *NBPTS*-certified teachers compared with other teachers using a quasi-experimental design in nine elementary schools in Brevard County and Seminole County Public School districts in Florida. The author compared the academic achievement of students receiving instruction from an *NBPTS*-certified teacher with those receiving instruction from a non–*NBPTS*-certified teacher. The author measured English language arts achievement using the Scholastic Reading Inventory standardized test. The analytic sample included 3,592 students (535 taught by *NBPTS*-certified teachers with a graduate degree and 3,057 taught by comparison group teachers with a graduate degree) in grade 5, during the 2008–09 school year.

Silver (2007) examined the effectiveness of *NBPTS*-certified teachers compared with other teachers using a quasi-experimental design in elementary schools in North Carolina. The author compared the academic achievement of students receiving instruction from an *NBPTS*-certified teacher with those receiving instruction from a non–*NBPTS*-certified teacher. The author measured English language arts achievement using state-required end-of-grade assessments. The analytic sample included 62 teachers (31 *NBPTS*-certified teachers and 31 comparison group teachers) in grades 3, 4, and 5 during the 2002–03 through 2004–05 school years.[7]

Stephens (2003) examined the effectiveness of *NBPTS*-certified teachers compared with other teachers using a quasi-experimental design in elementary schools in two large school districts in South Carolina. The author compared the academic achievement of students receiving instruction from an *NBPTS*-certified teacher with those receiving instruction from a non–*NBPTS*-certified teacher. The author measured mathematics achievement using state-required end-of-year standardized tests. The analytic sample included 153 students (72 taught by *NBPTS*-certified teachers and 81 taught by comparison group teachers) in grade 4, during the 2001–02 school year.

Effectiveness Summary

The WWC review of studies of teachers obtaining *NBPTS* certification 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. The five studies of *NBPTS*-certified teachers that met WWC group design standards reported findings in two of the six domains for student outcomes: (1) mathematics achievement and (2) English language arts achievement. The studies 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 *NBPTS*-certified teachers on students in grades 3–8. 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. 42.

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
Mixed effects *Evidence of inconsistent effects.*	In the three studies that reported findings, the estimated impact of the intervention on outcomes in the *mathematics achievement* domain was positive and statistically significant in one study, and neither statistically significant nor large enough to be substantively important in the other two studies.
Extent of evidence	**Criteria met**
Medium to large	Three studies that included 1,316,146[a] students reported evidence of effectiveness in the *mathematics achievement* domain.[b]

[a] The reported sample sizes may count some individual students more than once because some studies examined data from multiple school years.

[b] Stephens (2003) included 12 schools. Cowan and Goldhaber (2016) and Fisher and Dickenson (2005) did not report the number of schools included in their studies.

Three studies that meet WWC group design standards with reservations reported findings in the mathematics achievement domain.

Cowan and Goldhaber (2016) examined one outcome in the mathematics achievement domain: the authors created a standardized achievement measure (called a *z*-score) based on two state standardized assessments measured in different school years (before 2010, the Washington Assessment of Student Learning; thereafter, the Measures of Student Progress). The authors found, and the WWC confirmed, a positive and statistically significant effect of *NBPTS*-certified teachers on mathematics achievement. The WWC characterizes this study finding as a statistically significant positive effect. Supplemental findings presented in Appendix D do not factor into the intervention's rating of effectiveness.

Fisher and Dickenson (2005) examined one outcome in this domain: the Palmetto Achievement Challenge Test. The authors did not find a statistically significant effect of teachers with *NBPTS* certification 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.

Stephens (2003) examined one outcome in mathematics achievement: the Palmetto Achievement Challenge Test. The author did not find a statistically significant effect of teachers with *NBPTS* certification 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.

Thus, for the mathematics achievement domain, one study showed a statistically significant positive effect and two studies showed indeterminate effects. This results in a rating of mixed effects, with a medium to large extent of evidence.

Summary of effectiveness for the English language arts achievement domain

Table 4. Rating of effectiveness and extent of evidence for the English language arts achievement domain

Rating of effectiveness	Criteria met
No discernible effects *No affirmative evidence of effects.*	In the four studies that reported findings, the estimated impact of the intervention on outcomes in the *English language arts achievement* domain was neither statistically significant nor large enough to be substantively important.
Extent of evidence	**Criteria met**
Medium to large	Four studies that included 1,242,516[a] students reported evidence of effectiveness in the *English language arts achievement* domain.[b]

[a] The reported sample sizes may count some individual students more than once because some studies examined data from multiple school years.

[b] Gardner (2010) included all elementary schools in Brevard County and nine elementary schools in Seminole County. Cowan and Goldhaber (2016), Fisher and Dickenson (2005), and Silver (2007) did not report the number of schools included in their studies.

Four studies that met WWC group design standards with reservations reported findings in the English language arts achievement domain.

Cowan and Goldhaber (2016) examined one outcome in the English language arts achievement domain: the authors combined two state-standardized assessments measured in different school years (before 2010, the Washington Assessment of Student Learning; thereafter, the Measures of Student Progress). The authors did not find a statistically significant effect of *NBPTS*-certified teachers on English language arts 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, Cowan and Goldhaber (2016) found, and the WWC confirmed, seven statistically significant positive effects of *NBPTS*-certified teachers on English language arts achievement for the following student subgroups: (1) students in elementary school classrooms; (2) students eligible for free or reduced-price lunch in elementary school classrooms; (3) students receiving special education in elementary school classrooms; (4) students in middle school classrooms; (5) students in middle school classrooms (analyzed with cohort-by-track fixed effects); (6) students of teachers with Early Adolescence: English Language Arts (EA/ELA) certifications in middle school classrooms; and (7) students of teachers with EA/ELA certifications in middle school classrooms (analyzed with cohort-by-track fixed effects).

Fisher and Dickenson (2005) examined one outcome in this domain: the Palmetto Achievement Challenge Test. The authors did not find a statistically significant effect of *NBPTS*-certified teachers on English language arts achievement. The WWC-calculated average effect size was not large enough to be considered substantively important. The WWC characterizes these study findings 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, Fisher and Dickenson (2005) found, and the WWC confirmed, four statistically significant positive effects for the following student subgroups: (1) grade 4 students, (2) grade 8 students eligible for free or reduced-price lunch, (3) grade 4 students not eligible for free or reduced-price lunch, and (4) grade 7 students not eligible for free or reduced-price lunch.

Gardner (2010) examined one outcome in the English language arts domain: the Scholastic Reading Inventory. The author did not find a statistically significant effect of *NBPTS*-certified teachers on English language arts achieve-

ment. The WWC-calculated average effect size was not large enough to be considered substantively important. The WWC characterizes these study findings as an indeterminate effect.

Silver (2007) examined one outcome: the North Carolina End-of-Grade Reading assessment. The author used both the scale scores and the percentage of students meeting proficiency requirements for this measure. The author did not find a statistically significant effect of *NBPTS*-certified teachers on English language arts achievement. The WWC-calculated average effect size was not large enough to be considered substantively important. The WWC characterizes these study findings as an indeterminate effect.

Thus, for the English language arts achievement domain, four studies showed indeterminate effects. This results in a rating of no discernible effects, with a medium to large extent of evidence.

References

Studies that meet WWC group design standards without reservations

None.

Studies that meet WWC group design standards with reservations

Cowan, J., & Goldhaber, D. (2016). National Board certification and teacher effectiveness: Evidence from Washington state. *Journal of Research on Educational Effectiveness, 9*(3), 233–258. Retrieved from https://eric.ed.gov/?id=EJ1106512

Additional source:

Cowan, J., & Goldhaber, D. (2015). *National Board certification and teacher effectiveness: Evidence from Washington*. Technical Report 2015-1, Center for Education Data and Research, Seattle, WA. Retrieved from https://eric.ed.gov/?id=ED558082

Fisher, S., & Dickenson, T. (2005). *A study of the relationship between the National Board Certification status of teachers and students' achievement: Technical report.* Columbia: South Carolina Dept. of Education.

Gardner, D. J. (2010). *The effectiveness of state certified, graduate degreed, and National Board certified teachers as determined by student growth in reading* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED522796

Silver, K. T. (2007). *The National Board effect: Does the certification process influence student achievement?* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3280759)

Stephens, A. D. (2003). *The relationship between National Board certification for teachers and student achievement* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3084814)

Studies that do not meet WWC group design standards

Abernathy, D. F. (2009). *Affluence and influence: A study of inequities in the age of excellence* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3355826) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Ajimatanrareje, F. (2014). *An examination of teacher's certification or non-certification on students achievement* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3578849) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Antunez, F. (2015). *The effectiveness of the National Board Certification as it relates to the Advanced Placement Calculus AB exam* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 10154930) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Brown, A. L. (2012). *The cost effectiveness of a bonus pay plan for National Board Certified teachers in high poverty elementary schools in an urban school district in Florida* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3569611) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Buecker, H. L. (2010). *Quality teaching in addressing student achievement: A comparative study between National Board certified teachers and other teachers on the Kentucky Core Content Test results* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED527825 The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Cantrell, S., Fullerton, J., Kane, T. J., & Staiger, D. O. (2008). *National Board Certification and teacher effectiveness: Evidence from a random assignment experiment* (NBER Working Paper No. 14608). Cambridge, MA: National Bureau of Economic Research. Retrieved from https://eric.ed.gov/?id=ED503841 The study does not meet

WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Cavalluzzo, L. C. (2004). *Is National Board Certification an effective signal of teacher quality?* Alexandria, VA: CNA Corporation. Retrieved from https://eric.ed.gov/?id=ED485515 The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Childs, D. E., Jr. (2006). *Elementary school National Board certified teachers and student achievement* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3224419) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Chingos, M. M., & Peterson, P. E. (2011). It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review, 30*(3), 449–465. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Clark, S. B. (2012). *The effects of National Board Certification on student achievement* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED545934 The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Clotfelter, C. T., Ladd, H., & Vigdor, J. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review, 26*(6), 673–682. Retrieved from https://eric.ed.gov/?id=EJ781075 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:

Clotfelter, C. T., Ladd, H., & Vigdor, J. (2007). *How and why do teacher credentials matter for student achievements?* (CALDER Working Paper 2). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research. Retrieved from https://eric.ed.gov/?id=ED509655

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources, 41*(4), 778–820. Retrieved from https://eric.ed.gov/?id=EJ750956

Ladd, H., Clotfelter, C., & Vigdor, J. (2007). *How and why do teacher credentials matter for student achievements?* (NBER Working Paper 12828). Cambridge, MA: National Bureau of Economic Research. Retrieved from https://eric.ed.gov/?id=ED501923

Ladd, H. F., Sass, T. R., & Harris, D. N. (2007). *The impact of National Board certified teachers on student achievement in Florida and North Carolina: A summary of the evidence prepared for the National Academies Committee on the Evaluation of the Impact of Teacher Certification by NBPTS.* Washington, DC: The National Academies.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. *Journal of Human Resources, 45*(3), 655–681. Retrieved from https://eric.ed.gov/?id=EJ889247 The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Diaz, K. A. (2013). *Employing National Board certification practices with all teachers: The potential of cognitive coaching and mentoring* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED552760 The study does not meet WWC group design standards because the measures of effectiveness cannot be attributed solely to the intervention.

Falaney, P. E. (2007). *National Board for Professional Teaching Standards certification: Does it impact student learning?* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3257510) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National Board Certification as a signal of effective teaching. *Review of Economics and Statistics, 89*(1), 134–150. Retrieved from https://eric.ed.gov/?id=ED490921 The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Harris, D. N., & Sass, T. R. (2009). The effects of NBPTS-certified teachers on student achievement. *Journal of Policy Analysis and Management, 28*(1), 55–80. Retrieved from https://eric.ed.gov/?id=EJ822730 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:

Harris, D. N., & Sass, T. R. (2007). *The effects of NBPTS-certified teachers on student achievement* (Working Paper 4). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research (CALDER). Retrieved from https://eric.ed.gov/?id=ED509659

Ladd, H. F., Sass, T. R, & Harris, D. N. (2007). *The impact of National Board certified teachers on student achievement in Florida and North Carolina: A summary of the evidence prepared for the National Academies Committee on the Evaluation of the Impact of Teacher Certification by NBPTS.* Washington, DC: The National Academies.

Helding, K., & Fraser, B. (2013). Effectiveness of National Board Certified (NBC) teachers in terms of classroom environment, attitudes and achievement among secondary science students. *Learning Environments Research, 16*(1), 1–21. Retrieved from https://eric.ed.gov/?id=EJ996744 The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Kitts, A. S. (2011). *The relationship of student achievement and level of teacher certification: A quantitative study* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3459673) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Locklear, R. D. (2013). *A comparative study of National Board certified teachers and non-National Board certified teachers on student achievement in selected rural elementary schools in North Carolina* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3581531) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

McColskey, W., Stronge, J. H., Ward, T. J., Tucker, P. D., Howard, B., Lewis, K., & Hindman, J. L. (2005). *Teacher effectiveness, student achievement, and National Board Certified teachers.* Arlington, VA: National Board for Professional Teaching Standards. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

McCullough, M. T. (2011). *Impact of National Board certification, advanced degree, and socio-economic status on the literacy achievement rate of 11th grade students in Arkansas* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED535894 The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

McRae, J. S. (2014). *Advancing the science of hiring teachers: An analysis of the effects of teacher characteristics on student achievement* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3682166) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Morgigno, R. C. (2012). *The effects of National Board certified teachers on student achievement in Mississippi high schools* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED547197 The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Rouse, W. A. (2004). *An examination of student test results: National Board-Certified teachers and non-National Board-Certified teachers* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3120274) 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:

Rouse, W., & Hollomon, H. L. (2005). A comparison of student test results: Business and marketing education National Board Certified teachers and non-National Board teachers. *The Delta Pi Epsilon Journal, 47*(3), 128–142. Retrieved from https://eric.ed.gov/?id=EJ748223

Rouse, W. A., Jr. (2008). National Board Certified teachers are making a difference in student achievement: Myth or fact? *Leadership and Policy in Schools, 7*(1), 64–86. Retrieved from https://eric.ed.gov/?id=EJ811558

Saderholm, J. (2007). *Science inquiry learning environments created by National Board certified teachers* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3286743) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Sanders, W. L., Ashton, J. J., & Wright, S. P. (2005). *Comparison of the effects of NBPTS certified teachers with other teachers on the rate of student academic progress. Final report.* Arlington, VA: National Board for Professional Teaching Standards. Retrieved from https://eric.ed.gov/?id=ED491846 The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Sato, M., Ruth, C. W., & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of National Board Certification. *American Educational Research Journal, 45*(3), 669–700. Retrieved from https://eric.ed.gov/?id=EJ807296 The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Smith, T. W., Appalachian State University Office for Research on Teaching. (2005). *An examination of the relationship between depth of student learning and National Board Certification status.* Boone, NC: Office for Research on Teaching, Appalachian State University. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Strobel, T. L. (2011). *The effect of National Board Certification on student achievement in career and technology education* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3454027) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Stronge, J. H., Ward, T. J., Tucker, P. D., Hindman, J. L., McColsky, W., & Howard, B. (2007). National Board certified teachers and non-National Board certified teachers: Is there a difference in teacher effectiveness and student achievement? *Journal of Personnel Evaluation in Education, 20*(3-4), 185–210. Retrieved from https://eric.ed.gov/?id=EJ789880 The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Vandevoort, L. G., Amrein-Beardsley, A., & Berliner, D. C. (2004). National Board certified teachers and their students' achievement. *Education Policy Analysis Archives, 12*(46). Retrieved from https://eric.ed.gov/?id=EJ853513 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:

Vandevoort, L. G. (2004). *National Board certified teachers and student achievement* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3123636)

Vitale, T. M. (2008). *What is the relationship between National Board Certification and the achievement results of third grade students in a local central Florida school district?* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3319281) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Welborn, T. M. (2016). *Do students that have a National Board certified teacher have higher scores on standardized achievement tests in Mississippi?* (Doctoral dissertation, Mississippi College). The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

Study that does not meet WWC pilot regression-discontinuity design standards

Goldhaber, D., & Hansen, M. (2009). National Board certification and teachers' career paths: Does NBPTS certification influence how long teachers remain in the profession and where they teach? *Education Finance and Policy*, 4(3), 229–262. Retrieved from https://eric.ed.gov/?id=EJ849857 The study does not meet WWC pilot regression discontinuity design standards because it has high or unknown levels of attrition and does not demonstrate continuity of the outcome-forcing variable relationship.

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

Adams, A. (2016). *Teacher leadership: A little less conversation, A little more action research* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 10107569) This study is ineligible for review because it is out of scope of the protocol.

Allen, P. R. (2012). *Understanding the relationship between students' reading achievement and teachers' self-regulation patterns in grades K-3* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3578849) This study is ineligible for review because it is out of scope of the protocol.

Amos, J. L. (2013). *Supporting teachers: The role of reflection in professional learning* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED552435 This study is ineligible for review because it is out of scope of the protocol.

Angle, J. M. (2006). *Science teacher efficacy, National Board certification, and other teacher variables as predictors of Oklahoma students' end-of-instruction (EOI) Biology I test scores* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3211667) This study is ineligible for review because it does not use an eligible design.

Angulo, S. R. (2010). *Highly qualified: The perceptions of student learning and pedagogy related to mathematics of National Board certified teachers of urban Latino students* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED519381 This study is ineligible for review because it does not use an eligible design.

Bailey, A. T. (2010). *Leadership skills of North Carolina principals with certification from the National Board of Professional Teaching Standards* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3415796) This study is ineligible for review because it does not use an eligible design.

Balbach, A. B. M. (2012). *A survey of Pennsylvania school principals' perceptions of the National Board for Professional Teaching Standards certification process and the leadership roles of National Board certified teachers* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED546678 This study is ineligible for review because it does not use an eligible design.

Baratz-Snowden, J. (1993). Assessment of teachers: A view from the National Board for Professional Teaching Standards. *Theory into Practice, 32*(2), 82–85. Retrieved from https://eric.ed.gov/?id=EJ467924 This study is ineligible for review because it does not use an eligible design.

Beck, L. D. (2009). *The current state of professional development in Appalachia* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3380502) This study is ineligible for review because it is out of scope of the protocol.

Belson, S. I., & Husted, T. A. (2015). Impact of National Board for the Professional Teaching Standards certification on student achievement. *Education Policy Analysis Archives, 23*(91). Retrieved from https://eric.ed.gov/?id=EJ1084031 This study is ineligible for review because it does not use an eligible design.

Benigno, S. C., Jr. (2005). *A comparison of student scores on the Mississippi curriculum test of students taught by National Board certified teachers and non-National Board certified teachers* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3209666) This study is ineligible for review because it does not use an eligible design.

Benz, J. (1997). Measuring up: A personal journey through National Board certification in art. *Art Education, 50*(5), 20–24, 49–50. Retrieved from https://eric.ed.gov/?id=EJ566853 This study is ineligible for review because it does not use an eligible design.

Bivins, E. B. (2001). *A journey toward teaching mastery: Influences of National Board Certification on personal and professional development* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3007767) This study is ineligible for review because it does not use an eligible design.

Bohen, D. B. (2001). Strengthening teaching through national certification. *Educational Leadership, 58*(8), 50–53. Retrieved from https://eric.ed.gov/?id=EJ637143 This study is ineligible for review because it does not use an eligible design.

Boulden, S. M. (2011). *A mixed methods examination of the impact of National Board certified teachers in central Kentucky* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3475034) This study is ineligible for review because it does not use an eligible design.

Bowen, K. C. (2010). The relation of teachers' reflective judgment and conceptions of teaching and learning. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 70*(11-A), 4164. This study is ineligible for review because it does not use an eligible design.

Boyd, W. L., & Reese, J. P. (2006). Great expectations: The impact of the National Board for Professional Teaching Standards. *Education Next, 6*(2), 50–57. Retrieved from https://eric.ed.gov/?id=EJ763324 This study is ineligible for review because it does not use an eligible design.

Bozeka, J. L. (2015). *The professional development experiences of four Nationally Board certified teachers of reading-English language arts* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3730111) This study is ineligible for review because it does not use an eligible design.

Brenneman, L. (2010). *Wyoming teacher perceptions of teacher quality: Effects of National Board Certification and teacher education level* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED529578 This study is ineligible for review because it does not use an eligible design.

Bryant, A. J. (2010). *Perception of high-stakes testing by National Board certified teachers* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3407615) This study is ineligible for review because it is out of scope of the protocol.

Bumgarner, H. J. (2015). *The National Board Certification process as professional development: Perceptions about the impact that characteristics of the process had on professional growth* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3727984) This study is ineligible for review because it does not use an eligible design.

Cabezas, C. C. (2006). *The influence of highly qualified teacher designation, and other teacher variables, on student achievement* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3226820) This study is ineligible for review because it does not use an eligible design.

Cain, C. E. (2002). *Principal perceptions of National Board certified teachers* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3069437) This study is ineligible for review because it is out of scope of the protocol.

Cannata, M., McCrory, R., Sykes, G., Anagnostopoulos, D., & Frank, K. A. (2010). Exploring the influence of National Board certified teachers in their schools and beyond. *Educational Administration Quarterly, 46*(4),

463–490. Retrieved from https://eric.ed.gov/?id=EJ898243 This study is ineligible for review because it does not use an eligible design.

Cast, D. (2014). *The perceived impact of the National Board Certification process on Arkansas teachers* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3617652) This study is ineligible for review because it does not use an eligible design.

Chandler, K. D. (2005). *Paradigms, pedagogy and practice: Perspectives of National Board certified teachers in regard to reading* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3187853) This study is ineligible for review because it does not use an eligible design.

Collins, E. L. (2012). *A comparative study between National Board certified teachers' versus non-National Board certified teachers' perceived responsibility for student achievement* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3490050) This study is ineligible for review because it is out of scope of the protocol.

Corcoran, S. P., & Evans, W. N. (2008). The role of inequality in teacher quality. In K. Magnuson & J. Waldfogel (Eds.), *Steady gains and stalled progress: Inequality and the Black-White test score gap* (pp. 212–249). New York, NY: Russell Sage Foundation. This study is ineligible for review because it is out of scope of the protocol.

Craig, C. J. (2003). *Missouri school administrators' perceptions of the effectiveness of National Board certified teachers* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3102883) This study is ineligible for review because it does not use an eligible design.

Dagenhart, D. B., O'Connor, K. A., Petty, T. M., & Day, B. D. (2005). Giving teachers a voice. *Kappa Delta Pi Record, 41*(3), 108–111. Retrieved from https://eric.ed.gov/?d=EJ773876 This study is ineligible for review because it is out of scope of the protocol.

Davis, A., Wolf, K., & Borko, H. (1999). Examinees' perceptions of feedback in applied performance testing: The case of the National Board for Professional Teaching Standards. *Educational Assessment, 6*(2), 97–128. Retrieved from https://eric.ed.gov/?id=EJ604331 This study is ineligible for review because it does not use an eligible design.

Dickinson, G. K. (2006). *Achieving National Board certification for school library media specialists: A study guide.* Chicago, IL: American Library Association. This study is ineligible for review because it does not use an eligible design.

Diezi, C. (2004). *The effect of National Board-certified teachers on curriculum, instructional practices, and assessment decisions* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3147071) This study is ineligible for review because it does not use an eligible design.

Digby, A. D., & Avani, N. (2003). Moving toward a research agenda: Key questions for teacher educators on the role and impact of the National Board for Professional Teaching Standards. *Issues in Teacher Education, 12*(1), 9–17. Retrieved from https://eric.ed.gov/?id=EJ676788 This study is ineligible for review because it does not use an eligible design.

Fox, R. K., White, C. S., & Kidd, J. K. (2011). Program portfolios: Documenting teachers' growth in reflection-based inquiry. *Teachers and Teaching, 17*(1), 149–167. Retrieved from https://eric.ed.gov/?id=EJ911526 This study is ineligible for review because it is out of scope of the protocol.

Frank, K. A., Sykes, G., Anagnostopoulos, D., Cannata, M., Chard, L., Krause, A., & McCrory, R. (2008). Does NBPTS certification affect the number of colleagues a teacher helps with instructional matters? *Educational Evaluation and Policy Analysis, 30*(1), 3–30. Retrieved from https://eric.ed.gov/?id=EJ786470 This study is ineligible for review because it is out of scope of the protocol.

Galluzzo, G. R. (2005). Performance assessment and renewing teacher education: The possibilities of the NBPTS standards. *Clearing House, 78*(4), 142. Retrieved from https://eric.ed.gov/?id=EJ713922 This study is ineligible for review because it does not use an eligible design.

Gee, R. L. (2016). *A National Board certified teacher in the principalship: A qualitative analysis of leadership behaviors* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3735133) This study is ineligible for review because it does not use an eligible design.

Gitomer, D. (2007). *The impact of the National Board for Professional Teaching Standards: A review of research.* Princeton, NJ: Educational Testing Service. Retrieved from https://eric.ed.gov/?id=EJ1111636 This study is ineligible for review because it does not use an eligible design.

Goldhaber, D. (2006). National Board teachers are more effective, but are they in the classrooms where they're needed the most? *Education Finance and Policy, 1*(3), 372–382. Retrieved from https://eric.ed.gov/?id=EJ902830 This study is ineligible for review because it does not use an eligible design.

Hacke, W. (2010). *Meta-analysis comparing student outcomes for National Board certified teachers and non-National Board certified teachers* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED520141 This study is ineligible for review because it does not use an eligible design.

Hall, A. W. (2012). *National Board Certification: The impact on teaching practices of three elementary teachers* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED546387 This study is ineligible for review because it does not use an eligible design.

Harris, W. L. (2013). *The effect of National Board certified teachers on mathematics achievement for students in a Title I school* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED563808 This study is ineligible for review because it does not use an eligible design.

Holland, J. W. (2006). *Are Mississippi students achieving at a higher rate as a result of National Board certified teachers?* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3238928) This study is ineligible for review because it does not use an eligible design.

Holland, T. D. (2011). *How do teacher qualifications impact student achievement in relation to the achievement model established by the Mississippi State Department of Education?* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3455441) This study is ineligible for review because it does not use an eligible design.

Hollandsworth, S. E. (2006). *Best practices of National Board certified teachers and non-Board certified teachers in grades one and two* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3216930) This study is ineligible for review because it does not use an eligible design.

Houston, J. (2014). *Measures of effective teaching: National Board Certification and physical education teachers* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3642770) This study is ineligible for review because it is out of scope of the protocol.

Hunzicker, J. (2011). Teacher learning through National Board candidacy: A conceptual model. *Teacher Education Quarterly, 38(*3), 191–209. Retrieved from https://eric.ed.gov/?id=EJ940649 This study is ineligible for review because it does not use an eligible design.

Hunzicker, J. L. (2006). *The influence of the National Board Certification experience on teacher and student learning.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3244012) This study is ineligible for review because it does not use an eligible design.

Ingvarson, L., & Australian Council for Education Research. (2002). Strengthening the profession? A comparison of recent reforms in the UK and the USA. *ACER Policy Briefs,* Issue 2. Retrieved from https://eric.ed.gov/?id=ED499153 This study is ineligible for review because it is out of scope of the protocol.

Irwin-Beck, D. (2002). *National Board Certification: A descriptive study on its impact as a professional development activity* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3043400) This study is ineligible for review because it does not use an eligible design.

Jackson, L. (2009). *Effect of National Board Certification on retention of teachers in the classroom* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED535235 This study is ineligible for review because it does not use an eligible design.

Jay, J. K. (2003). *Quality teaching: Reflection as the heart of practice.* Lanham, MD: Scarecrow Press. This study is ineligible for review because it is out of scope of the protocol.

Johnson, T. S. (2009). Performing "teacher": A case study of a National Board certified teacher. *English Education, 41*(2), 158–176. Retrieved from https://eric.ed.gov/?id=EJ825830 This study is ineligible for review because it does not use an eligible design.

Kantner, L. A., Bergee, M. J., & Unrath, K. A. (2000). National Board Certification in art and its potential impact on graduate programming in art education. *Arts and Learning Research, 16*(1), 226–239. Retrieved from https://eric.ed.gov/?id=EJ638210 This study is ineligible for review because it does not use an eligible design.

Karaman, A. (2008). *Exploring the meaning of practicing classroom inquiry from the perspectives of National Board certified science teachers* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3301564) This study is ineligible for review because it does not use an eligible design.

Kelley, C., & Kimball, S. M. (2001). Financial incentives for National Board Certification. *Educational Policy, 15*(4), 547–574. This study is ineligible for review because it does not use an eligible design.

Knoeppel, R. C. (2008, November). *Increasing capacity to improve instruction: Are National Board certified teachers the answer?* Paper presented at the annual meeting of the University Council for Educational Administration, Orlando, FL. Retrieved from https://eric.ed.gov/?id=ED525683 This study is ineligible for review because it does not use an eligible design.

Lai, E. R., Auchter, J. E., & Wolfe, E. W. (2012). Confirmatory factor analysis of certification assessment scores from the National Board for Professional Teaching Standards. *The International Journal of Educational and Psychological Assessment, 9*(2), 61–81. This study is ineligible for review because it is out of scope of the protocol.

Laverick, D. M. (2005). *A qualitative study of teachers certified by the National Board for Professional Teaching Standards and their expertise in promoting early literacy* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3165958) This study is ineligible for review because it does not use an eligible design.

Le, H. T. (2015). *The relationship between preexisting teacher quality factors and high school student achievement* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3727121) This study is ineligible for review because it does not use an eligible design.

Lieberman, J. M., & Wilkins, E. A. (2006). The professional development pathways model: From policy to practice. *Kappa Delta Pi Record, 42*(3), 124–128. Retrieved from https://eric.ed.gov/?id=EJ738070 This study is ineligible for review because it does not use an eligible design.

Lucarelli, D. M. (2014). *Does the presence of National Board certified teachers make a difference? Examining standardized test scores and the perceptions of principals in Maryland elementary schools* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED569464 This study is ineligible for review because it does not use an eligible design.

Marshall, B. A. (2011). *Fostering positive classroom environments: The relationship between teacher qualifications, facility management, and perceptions of leadership on student outcomes* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED529053 This study is ineligible for review because it does not use an eligible design.

McDaniel, K. S. (2010). *National Board Certification and student achievement in Title I schools* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED516892 This study is ineligible for review because it does not use an eligible design.

McKenzie Lowery, E. N. (2010). *The relationship between National Board Certification and teachers' perceived use of developmentally appropriate practices* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3414792) This study is ineligible for review because it does not use an eligible design.

McKenzie, E. N. (2013). National Board Certification and developmentally appropriate practices: Perceptions of impact. *Journal of Research in Childhood Education, 27*(2), 153–165. Retrieved from https://eric.ed.gov/?id=EJ1011563 This study is ineligible for review because it does not use an eligible design.

Moore, P. B. (2000). *The effects of K-12 teacher professionalization on attitudes promoting equal education opportunity* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9993438) This study is ineligible for review because it is out of scope of the protocol.

National Board for Professional Teaching Standards. (2003). *National Board for Professional Teaching Standards*. Arlington, VA: Author. This study is ineligible for review because it does not use an eligible design.

Nesmith, B. S. (2011). *An investigation of National Board certified teachers' perceptions of teacher leadership dimensions on school support for teacher leadership involvement in high- and low-performing elementary schools in South Carolina* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED535929 This study is ineligible for review because it does not use an eligible design.

Neustel, S. B. (2001). *A psychometric investigation of NBPTS assessments: A comparative analysis of information functions* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3009701) This study is ineligible for review because it is out of scope of the protocol.

Nichols, L. C. (2016). *National Board certified teachers and methods they use to teach vocabulary* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 10243945) This study is ineligible for review because it does not use an eligible design.

Okpala, C. O., James, I., & Hopson, L. (2009). The effectiveness of National Board certified teachers: Policy implications. *Journal of Instructional Psychology, 36*(1), 29–34. Retrieved from https://eric.ed.gov/?id=EJ840814 This study is ineligible for review because it is out of scope of the protocol.

Palmer, J. L. (2013). *The National Board Certification portfolio process and its influence on teacher reflection* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED548044 This study is ineligible for review because it is out of scope of the protocol.

Park, S., & Oliver, J. S. (2008). National Board Certification (NBC) as a catalyst for teachers' learning about teaching: The effects of the NBC process on candidate teachers' PCK development. *Journal of Research in Science Teaching, 45*(7), 812–834. Retrieved from https://eric.ed.gov/?id=EJ809066 This study is ineligible for review because it does not use an eligible design.

Park, S., Oliver, J. S., Johnson, T. S., Graham, P., & Oppong, N. K. (2007). Colleagues' roles in the professional development of teachers: Results from a research study of National Board Certification. *Teaching and Teacher Education, 23*(4), 368–389. Retrieved from https://eric.ed.gov/?id=EJ756902 This study is ineligible for review because it does not use an eligible design.

Pastore, D. A. (2016). *National Board certification: An analysis of multiple variables on pass rates* (Doctoral dissertation, Washington State University). This study is ineligible for review because it is out of scope of the protocol.

Petty, T. M., Good, A. J., & Handler, L. K. (2016). Impact on student learning: National Board certified teachers' perceptions. *Education Policy Analysis Archives, 24*(49), 1–22. Retrieved from https://eric.ed.gov/?id=EJ1100180 This study is ineligible for review because it does not use an eligible design.

Petty, T. M., O'Connor, K. A., & Dagenhart, D. B. (2010). Was it worth it? Some National Board certified teachers say no! *Educational Forum, 74*(1), 19–24. Retrieved from https://eric.ed.gov/?id=EJ881461 This study is ineligible for review because it does not use an eligible design.

Place, N. A., & Coskie, T. L. (2006). Learning from the National Board portfolio process: What teachers discovered about literacy teaching and learning. *New Educator, 2*(3), 227–246. Retrieved from https://eric.ed.gov/?id=EJ819708 This study is ineligible for review because it does not use an eligible design.

Pool, J., Ellett, C., Schiavone, S., & Carey-Lewis, C. (2001). How valid are the National Board of Professional Teaching Standards assessments for predicting the quality of actual classroom teaching and learning? Results of six mini case studies. *Journal of Personnel Evaluation in Education, 15*(1), 31–48. Retrieved from https://eric.ed.gov/?id=EJ633958 This study is ineligible for review because it does not use an eligible design.

Preach, D. (2013). *Supporting and fostering the development of alternatively certified teachers: Creating a collaborative community* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED552958 This study is ineligible for review because it does not use an eligible design.

Qualls, K. M. (2015). *Teacher in the mirror: Reflective practices of National Board certified teachers* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3688881) This study is ineligible for review because it does not use an eligible design.

Rhoades, J. L. (2010). *National Board certified physical education teachers: A descriptive analysis* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3452256). This study is ineligible for review because it does not use an eligible design.

Rhoades, J. L., & Woods, A. M. (2012). National Board Certified Physical Education Teachers task presentations and learning environments. *Journal of Teaching in Physical Education, 31*(1), 4–20. Retrieved from https://eric.ed.gov/?id=EJ978079 This study is ineligible for review because it does not use an eligible design.

Rorie, L. G. (2014). *Correlation between National Board certified teachers and reading achievement in elementary schools* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED556908 This study is ineligible for review because it does not use an eligible design.

Sato, M., Hyler, M. E., & Monte-Sano, C. (2014). Learning to lead with purpose: National Board certification and teacher leadership development. *International Journal of Teacher Leadership, 5*(1), 1–23. Retrieved from https://eric.ed.gov/?id=EJ1137495 This study is ineligible for review because it does not use an eligible design

Serafini, F. (2005). Taking on the National Board for Professional Teaching Standards: Alignment, recognition and representation. *Current Issues in Education, 8*(21). Retrieved from https://eric.ed.gov/?id=EJ875563 This study is ineligible for review because it does not use an eligible design.

Singleton, R. L. (2010). *The National Board Certification process: A comparison of the perceptions of National Board certified teachers and National Board candidates in West Virginia* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED521770 This study is ineligible for review because it does not use an eligible design.

Sottile, K. M. (2014). *Exploring the relationship between accomplished teaching through National Board Certification for teachers and teacher leadership in New York State* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED568412 This study is ineligible for review because it is out of scope of the protocol.

Standerfer, S. L. (2003). *Perceptions and influences of the National Board for Professional Teacher Certification on secondary choral music teachers: Three case studies* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3083085) This study is ineligible for review because it does not use an eligible design.

Standerfer, S. L. (2008). Learning from the National Board for Professional Teacher Certification (NBPTS) in music. *Bulletin for the Council of Research in Music Education,* (176), 77–88. This study is ineligible for review because it does not use an eligible design.

Starnes, R. J. (2013). *National Board certified teachers in Pennsylvania: A study of motivation and persistence* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED553192 This study is ineligible for review because it does not use an eligible design.

Stone, J. E. (2002). *The value-added achievement gains of NBPTS-certified teachers in Tennessee: A brief report.* Retrieved from https://eric.ed.gov/?id=ED472132 This study is ineligible for review because it does not use an eligible design.

Sullivan, D. (2010). *An examination of National Board certified teachers' views of the professional impact of National Board Certification* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED526460 This study is ineligible for review because it does not use an eligible design.

Swoger, P. A. (2002). *An investigation of National Board Certification in Mississippi* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3043179) This study is ineligible for review because it does not use an eligible design.

Tingle, S. M. (2014). *Teacher's perceptions of National Board Certification as professional development and evaluation* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3609682) This study is ineligible for review because it does not use an eligible design.

Wade, T. L. (2001). *National Board Certification and new roles for teachers: Impact on turnover and attrition among secondary mathematics teachers in North Carolina* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3036244) This study is ineligible for review because it does not use an eligible design.

Walker, S. A. A. (2001). *An investigation of the relationship between teacher personality and National Board Certification among south Mississippi teachers* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED460114 This study is ineligible for review because it is out of scope of the protocol.

Warner, K. L. (2002). *The effect of professional development experiences on National Board for Professional Teaching Standards candidates' scores in Florida* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3038205) This study is ineligible for review because it is out of scope of the protocol.

Whaley, J. W. S. (2003). *Powerful professional development: A perpetuation theory and network analysis of teachers' perceptions of the National Board for Professional Teaching Standards certification process* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3127196) This study is ineligible for review because it does not use an eligible design.

Whitaker, S. R. (2008). *One National Board certified teacher's post-certification journey with differentiated reading instruction in middle school language arts* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3353850) This study is ineligible for review because it does not use an eligible design.

Whitman, B. A. (2002). *Professional teachers for quality education: Characteristics of teachers certified by the National Board for Professional Teaching Standards* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3070572) This study is ineligible for review because it is out of scope of the protocol.

Wiebke, K. M. (2010). *National Board Certification: The power of one, the potential of twenty.* (Doctoral dissertation). Retrieved from https://eric.ed.gov/?id=ED516893 This study is ineligible for review because it does not use an eligible design.

Willer, D. B. (2014). *Targeting success. An evaluation of information literacy standards: A mixed method approach utilizing the judgments of National Board certified teachers* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3680315) This study is ineligible for review because it does not use an eligible design.

Woods, A. M., & Rhoades, J. (2013). Teaching efficacy beliefs of National Board certified physical educators. *Teachers and Teaching, 19*(5), 507–526. Retrieved from https://eric.ed.gov/?id=EJ1022188 This study is ineligible for review because it does not use an eligible design.

Yeh, S. S. (2010). The cost-effectiveness of NBPTS teacher certification. *Evaluation Review, 34(*3), 220–241. Retrieved from https://eric.ed.gov/?id=EJ883803 This study is ineligible for review because it does not use an eligible design.

Yeh, S. S. (2011). *The cost-effectiveness of 22 approaches for raising student achievement.* Charlotte, NC: Information Age Publishers. Retrieved from https://eric.ed.gov/?id=ED529522 This study is ineligible for review because it is out of scope of the protocol.

Young, Y. Y. (2013). *National Board Certification and its influence on leadership self-efficacy* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3578843) This study is ineligible for review because it does not use an eligible design.

Appendix A.1: Research details for Cowan and Goldhaber (2016)

Cowan, J., & Goldhaber, D. (2016). National Board certification and teacher effectiveness: Evidence from Washington state. *Journal of Research on Educational Effectiveness,* 9(3), 233–258.[8]

Table A1. Summary of findings

Meets WWC Group Design Standards With Reservations

Outcome domain	Sample size	Study findings	
		Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	1,312,657 students	+2	Yes
English language arts achievement	1,234,924 students	+1	No

Setting

This study was conducted in elementary and middle school grades throughout Washington state.

Study sample

This study examined two groups of students: elementary school classrooms, defined as those in self-contained classes, primarily grades 3–5, but some sixth-grade classes; and middle school classrooms, defined as those in non–self-contained classes, primarily grades 7 and 8, with some sixth-grade classes. The students in elementary school classes were examined between the 2005–06 and 2012–13 school years, while students in middle school classes were examined between the 2009–10 and 2012–13 school years. The analytic sample for the mathematics scores includes 110,634 students taught by *NBPTS*-certified teachers, and 1,202,023 students taught by comparison teachers. The analytic sample for the English language arts scores includes 113,129 students taught by *NBPTS*-certified teachers, and 1,121,795 students taught by comparison teachers. Because the study spans multiple school years, individual students may be included more than once in the sample size counts. Demographics are not provided for the full sample of elementary and middle school students. The WWC-calculated weighted average demographics between the elementary and middle school math samples suggest that in the analytic sample, 49% of students were female; about 63% were White, 17% Hispanic, 9% were Asian, 5% Black, 5% multiracial, and 2% were American Indian.[9] Among the students in the sample, about 5% had limited English proficiency, 6% had a learning disability, and 46% were eligible for free or reduced-price lunches.

In addition, the authors present subgroup findings for school level (elementary school or middle school classrooms), *NBPTS*-certification subject area (Middle Childhood: Generalist [MC/Gen], Early/Middle Childhood: Literacy, Reading, and Language Arts [EMC/LRLA], Early Adolescence: English Language Arts [EA/ELA], and Early Adolescence: Math [EA/Math]), special education status, eligibility for free or reduced-price lunch, and schools with low high-poverty rates (Challenging Schools Bonus vs. non-Challenging Schools Bonus). The subgroup findings are reported in Appendix D.[10] The supplemental findings do not factor into the intervention's rating of effectiveness.

Intervention group

The intervention consisted of regular instruction for 1 year by an *NBPTS*-certified teacher.

Comparison group

The comparison consisted of regular instruction for 1 year by a teacher who was not *NBPTS*-certified.

Outcomes and measurement

This study examined one outcome in the mathematics achievement domain and one outcome in the English language arts achievement domain. Both outcomes were measured using the same instrument in a given year, but there was a change in the instruments used during the study. For outcomes prior to spring 2010, student achievement was measured using the Washington Assessment of Student Learning test. This test was replaced with the Measurements of Student Progress assessment in spring 2010. These outcomes were standardized, and the analysis included cohort fixed effects. For a more detailed description of these outcome measures, see Appendix B.

Support for implementation

Teachers are provided incentives to become *NBPTS*-certified teachers, and they are also offered financial incentives to teach in lower performing schools. Prior to 2008, Washington state provided a $3,500 salary incentive for certified teachers, which increased to $5,000 in 2008. Also starting in 2008, Washington state *NBPTS*-certified teachers were offered a $5,000 incentive to teach in lower performing schools. Individual school districts may offer additional incentives such as financial support, release for certification activities, and mentoring.

Appendix A.2: Research details for Fisher and Dickenson (2005)

Fisher, S., & Dickenson, T. (2005). *A study of the relationship between the National Board certification status of teachers and students' achievement: Technical report.* Columbia: South Carolina Dept. of Education.

Table A2. Summary of findings **Meets WWC Group Design Standards With Reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	288 teachers/3,336 students	+2	No
English language arts achievement	406 teachers/3,938 students	+4	No

Setting

This study was conducted in elementary and middle school grades throughout South Carolina.

Study sample

This study examined students in grades 4–8 using a quasi-experimental matched-comparison design. *NBPTS*-certified teachers who taught math or English language arts in grades 4–8 were matched with non-certified teachers who had similar years of teaching experience and who taught in schools with similar school poverty levels and student/teacher ratios as the *NBPTS*-certified teachers. Non-certified teachers who taught in schools with an *NBPTS*-certified teacher or *NBPTS*-applicant teacher were excluded from the comparison group as they may benefit from working collaboratively with certified teachers or applicants. The analytic sample for the mathematics scores includes 1,668 students taught by 144 *NBPTS*-certified teachers, and 1,668 students taught by 144 comparison teachers. The analytic sample for the English language arts scores includes 1,969 students taught by 187 *NBPTS*-certified teachers, and 1,969 students taught by 187 comparison teachers. Approximately 47% of students received free or reduced-price lunch.

In addition, the authors present subgroup findings by grade (4, 5, 6, 7, or 8) and by whether students were eligible for free or reduced-price lunch (eligible or not eligible). The subgroup findings are reported in Appendix D.[11] The supplemental findings do not factor into the intervention's rating of effectiveness.

Intervention group

The intervention consisted of regular instruction in mathematics or English language arts for 1 year by a teacher with *NBPTS* certification. Depending on the grade taught, *NBPTS*-certified teachers had an average of between 13.7 and 17.8 years of experience.

Comparison group

The comparison consisted of regular instruction in mathematics or English language arts for 1 year by a teacher who was not *NBPTS*-certified. Depending on the grade taught, non-certified teachers had an average of between 10.4 and 14.1 years of experience.

Outcomes and measurement

This study examined two outcomes, mathematics achievement and English language arts achievement. Both outcomes were measured using the Palmetto Achievement Challenge Test. For a more detailed description of this outcome measure, see Appendix B.

Support for implementation

NBPTS-certified teachers automatically received an equivalent of 12 credit hours toward the renewal of their teaching certificates, additional annual pay while maintaining *NBPTS* certification, and forgiveness of any loans used to pay for the application fee.

Appendix A.3: Research details for Gardner (2010)

Gardner, D. J. (2010). *The effectiveness of state certified, graduate degreed, and National Board certified teachers as determined by student growth in reading* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3415029)

Table A3. Summary of findings **Meets WWC Group Design Standards With Reservations**

		Study findings	
Outcome domain	Sample size	Average improvement index *(percentile points)*	Statistically significant
English language arts achievement	3,592 students	0	No

Setting

This study took place in two public school districts in Florida; specifically, all elementary schools in Brevard County Public Schools and nine elementary schools in Seminole County Public Schools participated.

Study sample

The students included in this study were in grades 3–5 during school year 2008–09 in Florida. The analytic sample for the mathematics scores includes 535 students taught by *NBPTS*-certified teachers, and 3,057 students taught by comparison teachers. About 70% were White, 12% were Black, 9% were Hispanic, 6% were of mixed race, and 3% were Asian. About 51% were male, less than 3% were English learners, and about 35% qualified for free or reduced-price lunch.

In addition, the author presents subgroup findings by grade (3, 4, or 5) and by the highest degree obtained by the teacher (bachelor's or graduate). The subgroup findings are reported in Appendix D. The supplemental findings do not factor into the intervention's rating of effectiveness.

Intervention group The intervention condition was receiving 1 year of instruction by a teacher with *NBPTS* certification.

Comparison group The comparison condition was receiving 1 year of instruction from teachers without *NBPTS* certification.

Outcomes and measurement This study measured English language arts achievement using the Scholastic Reading Inventory. This test was administered at the beginning of the school year and again at the end of April. For a more detailed description of this outcome measure, see Appendix B.

Support for implementation The study notes that the state of Florida provides a salary bonus to teachers who achieve *NBPTS* certification. No details are provided on this salary bonus system.

Appendix A.4: Research details for Silver (2007)

Silver, K. T. (2007). *The National Board effect: Does the certification process influence student achievement?* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3280759)

Table A4. Summary of findings **Meets WWC Group Design Standards With Reservations**

Outcome domain	Sample size	Study findings	
		Average improvement index *(percentile points)*	Statistically significant
English language arts achievement	62 teachers	+1	No

Setting This study was conducted in elementary school grades 3–5 throughout North Carolina.

Study sample The study examined the effect of *NBPTS*-certified teachers in the first year after they received certification. The author identified 81 teachers in grades 3–5 who received *NBPTS* certification in the 2003–04 school year and matched these teachers to 81 comparison teachers without *NBPTS* certification based on teaching experience, degree level, grade level taught, and school district. Approximately 90% of the teachers were White, 8% were Black, 1% were Hispanic, and less than 1% were Native American, 95% were female, and 72% held bachelor's degrees. The analytic sample included 31 *NBPTS*-certified teachers and 31 comparison teachers without *NBPTS* certification.

In addition, the author present subgroup findings by grade (3, 4, or 5). The subgroup findings are reported in Appendix D. The supplemental findings do not factor into the intervention's rating of effectiveness.

Intervention group The intervention condition was receiving 1 year of instruction during the 2004–05 school year by a teacher receiving *NBPTS* certification in the prior school year.

Comparison group

The comparison condition was receiving 1 year of instruction during the 2004–05 school year from teachers without *NBPTS* certification.

Outcomes and measurement

This study measured English language arts achievement using the North Carolina End-of-Grade reading assessment, a state-required test given to all North Carolina public school students in grades 3–8. The author examined the raw score obtained on this assessment, as well as the percent of students scoring above the threshold required to be considered proficient by North Carolina standards.[12] For a more detailed description of this outcome measure, see Appendix B.

Support for implementation

Teachers obtaining *NBPTS* certification are provided with a 12% salary supplement in North Carolina.

Appendix A.5: Research details for Stephens (2003)

Stephens, A. D. (2003). *The relationship between National Board certification for teachers and student achievement* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3084814)

Table A5. Summary of findings **Meets WWC Group Design Standards With Reservations**

Outcome domain	Sample size	Study findings	
		Average improvement index *(percentile points)*	Statistically significant
Mathematics achievement	22 teachers/153 students	0	No

Setting

This study took place in elementary school grades 4 and 5 in two large school districts in South Carolina. One district was described as a suburban district with a total population of 14,759 students across 36 schools. The second district contained urban, suburban, and rural schools with a total of 42,446 students across 85 schools.

Study sample

This study individually matched each of eight teachers with *NBPTS* certification to a teacher without certification. Four of the *NBPTS*-certified teachers taught students in grade 4 and four in grade 5. Individual teachers were matched on the prior year's mathematics achievement of their current students in the instructional year, as well as within a range of the school-level poverty index. Intervention and comparison group teachers were chosen from within each of the participating school districts. The analytic sample includes 72 students taught by the four *NBPTS*-certified teachers, and 81 students taught by the four comparison teachers. The race, gender, and free and reduced-price lunch status of students were not reported. Across all matches, the poverty level ranged from 14.2 to 98.5.

The author presented separate comparisons for each *NBPTS*-certified teacher. Each of these contrasts has a confounding factor since the intervention condition was delivered by a single teacher. An author query was sent to see if aggregate findings were available. The author did not have aggregated findings, so the WWC aggregated the four contrasts for each grade and used these aggregated findings as the contrasts of interest for this review.

Intervention group

The intervention condition was receiving 1 year of instruction in math during the 2001–02 school year by a teacher with *NBPTS* certification. Each teacher had at least 3 years of experience.

Comparison group

The intervention condition was receiving 1 year of instruction in math during the 2001–02 school year by a teacher without *NBPTS* certification. Each teacher had at least 3 years of experience.

Outcomes and measurement

This study measured mathematics achievement using the Palmetto Achievement Challenge Test, a state-required standardized assessment. For a more detailed description of this outcome measure, see Appendix B.

Support for implementation

The state of South Carolina provided a $7,500 bonus for *NBPTS* certification. The two participating school districts provided salary stipends and/or compensation to teachers achieving *NBPTS* certification; no details on these incentives were provided in the study.

Appendix B: Outcome measures for each domain

Mathematics achievement	
Palmetto Achievement Challenge Test	Fisher and Dickenson (2005) used this state assessment to measure achievement for students in grades 4–8. Scaled scores from the 2004 administration were used as the outcome (as cited in Fisher & Dickenson, 2005). Stephens (2003) also used this assessment to measure achievement for students in school years 2000–01 and 2001–02 (as cited in Stephens, 2003). Statewide, students in each grade obtain an average of 100 times their grade level on each assessment, such as 400 for grade 4 and 800 for grade 8 (Fisher & Dickenson, 2005).
Standardized Math Test	Cowan and Goldhaber (2016) created a standardized math score using the Measures of Student Progress and the Washington Assessment of Student Learning for students in grades 3–8. The Washington Assessment of Student Learning was used for school years 2006–07 through fall 2009–10. The Measures of Student Progress was used for the spring of school year 2009–10 and all of school year 2012–13 (as cited in Cowan & Goldhaber, 2016).
English language arts achievement	
North Carolina End-of-Grade Reading Assessment	Silver (2007) used the state-required end-of-grade reading assessment in North Carolina for students in grades 3–5. This is a multiple-choice test aligned to the North Carolina Standard Course of Study and is given to all public school students in North Carolina in grades 3–8. The average test-retest reliability was .86 and the internal consistency ranged from .90 to .94. This outcome was examined in scale score units and in the percent of students meeting proficiency standards for each grade (as cited in Silver, 2007).
Palmetto Achievement Challenge Test	Fisher and Dickenson (2005) used this state assessment to measure achievement for students in grades 4–8. Scaled scores from the 2004 administration were used as the outcome. Statewide, students in each grade obtain an average of 100 times their grade level on each assessment, such as 400 for grade 4 and 800 for grade 8 (Fisher & Dickenson, 2005).
Scholastic Reading Inventory	Gardner (2010) measured English language arts achievement for students in grades 3–5 using the Lexile measure from the Scholastic Reading Inventory (SRI). The Lexile measure is nationally-normed and ranges from 0L to 2000L and provides a metric to assess reading growth over time. The SRI is a reading comprehension assessment where students read brief passages and answer questions about the content. This assessment is taken via computer and has been externally validated for construct and criterion-related validity (as cited in Gardner, 2010).
Standardized English Language Arts Test	Cowan and Goldhaber (2016) created a standardized English language arts score using the Measures of Student Progress and the Washington Assessment of Student Learning for students in grades 3–8. The Washington Assessment of Student Learning was used for school years 2006–07 through fall 2009–10. The Measures of Student Progress was used for the spring of school year 2009–10 and all of school year 2012–13 (as cited in Cowan & Goldhaber, 2016).

Appendix C.1: 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	
Cowan & Goldhaber (2016)[a]								
Standardized Math Test	Elementary and middle school students	15,556 teachers/ 1,312,657 students	0.03 (1.02)	–0.01 (0.99)	0.04	0.04	+2	< .01
Domain average for mathematics achievement (Cowan & Goldhaber, 2016)						**0.04**	**+2**	**Statistically significant**
Fisher & Dickenson (2005)[b]								
Palmetto Achievement Challenge Test	Grades 4–8	288 teachers/ 3,336 students	0.05 (1.00)	0.00 (1.00)	0.05	0.05	+2	.41
Domain average for mathematics achievement (Fisher & Dickenson, 2005)						**0.05**	**+2**	**Not statistically significant**
Stephens (2003)[c]								
Palmetto Achievement Challenge Test	Grade 4	8 teachers/ 153 students	421.66 (13.78)	421.51 (13.16)	0.15	0.01	0	.98
Domain average for mathematics achievement (Stephens, 2003)						**0.01**	**0**	**Not statistically significant**
Domain average for mathematics achievement across all studies						**0.03**	**+1**	**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 WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of each study's domain average was determined by the WWC. Some statistics may not sum as expected due to rounding. na = not applicable.

[a] For Cowan and Goldhaber (2016), a correction for clustering was needed but did not affect whether any of the contrasts were found to be statistically significant. The effect size was calculated using the ordinary least-squares (OLS) coefficient. The single finding presented here is based on an aggregated sample of elementary and middle school students separately reported in the original study. The authors provided unadjusted baseline and post-intervention means and standard deviations for the outcome at the WWC's request. The authors reported *p*-values for some results, but not for the aggregated analysis. The WWC applied a correction for clustering and calculated the *p*-value reported in the table. This study is characterized as having a statistically significant positive effect because the estimated effect for the one measure in this domain is positive and statistically significant. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

[b] For Fisher and Dickenson (2005), a correction for clustering was needed but did not affect whether any of the contrasts were found to be statistically significant. The effect size was calculated using the unadjusted mean and standard deviation calculation. The single finding presented here is based on an aggregated sample of students in grades 4–8 reported separately by grade in the study. Because the outcome measure was not scaled to allow direct comparisons of scores across grades, the WWC standardized the scores and removed between-grade variation in the outcome means prior to aggregating across grades. The authors reported *p*-values for some results, but not for the aggregated analysis. The WWC applied a correction for clustering and calculated the *p*-value reported in the table. 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.

[c] For Stephens (2003), a correction for clustering was needed but did not affect whether any of the contrasts were found to be statistically significant. The single finding presented here is based on an aggregated sample of grade 4 teachers and their students, which were reported separately by teacher in the original study. The effect size was calculated using the unadjusted mean and standard deviation calculation. The author reported *p*-values for some results, but not for the aggregated analysis. The WWC applied a correction for clustering and calculated the *p*-value reported in the table. 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 C.2: Findings included in the rating for the English language arts 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	
Cowan & Goldhaber (2016)[a]								
Standardized English Language Arts Test	Elementary and middle school students	16,081 teachers/ 1,234,924 students	0.03 (0.97)	0.02 (0.99)	0.01	0.02	+1	.24
Domain average for English language arts achievement (Cowan & Goldhaber, 2016)						**0.02**	**+1**	**Not statistically significant**
Fisher & Dickenson (2005)[b]								
Palmetto Achievement Challenge Test	Grades 4–8	374 teachers/ 3,938 students	0.10 (1.00)	0.00 (1.00)	0.10	0.10	+4	.07
Domain average for English language arts achievement (Fisher & Dickenson, 2005)						**0.10**	**+4**	**Not statistically significant**
Gardner (2010)[c]								
Scholastic Reading Inventory	Grade 5 students of teachers with a bachelor's degree	3,592 students	923.93 (218.03)	921.47 (221.12)	2.46	0.01	0	.81
Domain average for English language arts achievement (Gardner, 2010)						**0.01**	**0**	**Not statistically significant**
Silver (2007)[d]								
North Carolina End-of-Grade Reading Assessment	Grade 4 teachers	62 teachers	252.91 (3.74)	252.92 (3.98)	–0.01	–0.00	0	.99
Percent proficient on North Carolina End-of-Grade Reading Assessment	Grade 4 teachers	62 teachers	84.96 (na)	84.10 (na)	0.86	0.07	+3	.77
Domain average for English language arts achievement (Silver, 2007)						**0.04**	**+1**	**Not statistically significant**
Domain average for English language arts achievement across all studies						**0.04**	**+2**	**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 WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of each study's domain average was determined by the WWC. Some statistics may not sum as expected due to rounding. na = not applicable.

[a] For Cowan and Goldhaber (2016), a correction for clustering was needed but did not affect whether any of the contrasts were found to be statistically significant. The effect size was calculated using the ordinary least-squares (OLS) coefficient. The single outcome presented here is based on an aggregated sample of elementary and middle school students separately reported in the original study. The authors provided unadjusted baseline and post-intervention means and standard deviations for the outcome at the WWC's request. The authors reported *p*-values for some results, but not for the aggregated analysis. The WWC applied a correction for clustering and calculated the *p*-value reported in the table. 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.

[b] For Fisher and Dickenson (2005), a correction for clustering was needed but did not affect whether any of the contrasts were found to be statistically significant. The effect size was calculated using the unadjusted mean and standard deviation calculation. The single finding presented here is based on an aggregated sample of students in grades 4–8 reported separately by grade in the study. Because the outcome measure was not scaled to allow direct comparisons of scores across grades, the WWC standardized the scores and removed between-grade variation in the outcome means prior to aggregating across grades. The authors reported *p*-values for some results, but not for the aggregated analysis. The WWC applied a correction for clustering and calculated the *p*-value reported in the table. 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.

[c] For Gardner (2010), the WWC calculated the intervention group mean using a difference-in-differences approach by adding the impact of the intervention (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the WWC Procedures and Standards Handbook (version 3.0), p. 23 for more information. The WWC did not make corrections for clustering or multiple comparisons. The *p*-value presented here was calculated by the WWC. The WWC was unable to make corrections for clustering because the number of teachers included in the study was unknown. 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.

[d] For Silver (2007), the WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences. The WWC calculated the intervention group mean using a difference-in-differences approach by adding the impact of the intervention (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the WWC Procedures and Standards Handbook (version 3.0), p. 23 for more information. The *p*-values presented here were calculated by the WWC. 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.1a: Supplemental findings for the mathematics achievement domain, elementary grades

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Cowan & Goldhaber (2016)[a,13]								
Standardized Math Test	All students	10,300 teachers/ 742,124 students	0.02 (1.02)	0.00 (1.00)	0.02	0.02	+1	< .01
Standardized Math Test	English learners	10,300 teachers/ 48,631 students	nr	nr	–0.01	nr	nr	> .10
Standardized Math Test	Special education students	10,300 teachers/ 92,937 students	nr	nr	0.03	nr	nr	< .01
Standardized Math Test	FRPL students	10,300 teachers/ 331,924 students	nr	nr	0.01	nr	nr	> .10
Standardized Math Test	Students in high-poverty schools	10,300 teachers/ 331,924 students	nr	nr	0.04	nr	nr	< .05
Standardized Math Test	Teachers have MC/ GEN certifications	11,050 teachers/ 727,768 students	nr	nr	0.02	nr	nr	< .05
Standardized Math Test	Teachers have EMC/LRA certifications	11,050 teachers/ 701,403 students	nr	nr	0.03	nr	nr	< .10
Fisher & Dickenson (2005)[b]								
Palmetto Achievement Challenge Test	Grade 4	98 teachers/ 666 students	414.88 (13.30)	414.16 (13.66)	0.72	0.05	+2	.36
Palmetto Achievement Challenge Test	Grade 5	74 teachers/ 482 students	511.90 (14.16)	511.29 (15.08)	0.61	0.61	+2	.49
Palmetto Achievement Challenge Test	Grade 6	28 teachers/ 546 students	616.58 (15.40)	614.99 (15.05)	1.59	0.10	+4	.03
Palmetto Achievement Challenge Test	Grade 4, FRPL	98 teachers/ 322 students	409.02 (11.42)	409.13 (14.25)	–0.11	–0.01	0	.93
Palmetto Achievement Challenge Test	Grade 5, FRPL	74 teachers/ 250 students	506.01 (11.55)	504.82 (13.52)	1.19	0.09	+4	.34
Palmetto Achievement Challenge Test	Grade 6, FRPL	28 teachers/ 254 students	607.50 (13.83)	607.24 (14.51)	0.26	0.02	+1	.81

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Palmetto Achievement Challenge Test	Grade 4, non-FRPL	98 teachers/ 344 students	420.36 (12.62)	418.86 (11.24)	1.50	0.13	+5	.15
Palmetto Achievement Challenge Test	Grade 5, non-FRPL	74 teachers/ 232 students	518.26 (14.01)	518.27 (13.54)	–0.01	–0.00	0	> .99
Palmetto Achievement Challenge Test	Grade 6, non-FRPL	28 teachers/ 292 students	624.49 (11.98)	621.73 (11.97)	2.76	0.23	+9	< .05

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. nr = not reported. MC/GEN = Middle Childhood: Generalist certificate. EMC/LRLA = Early and Middle Childhood: Literacy, Reading, and Language Arts certificate. FRPL indicates students eligible for free or reduced-price lunch.

[a] For Cowan and Goldhaber (2016), the *p*-values presented here were reported in the original study. A correction for clustering and for multiple comparisons within the elementary school grades was needed and resulted in a WWC-computed critical *p*-value of .005 for special education students, a WWC-computed critical *p*-value of .01 for students in high-poverty schools, a WWC-computed critical *p*-value of .02 for students whose teachers had MC/GEN certifications, a WWC-computed critical *p*-value of .02 for the apparently random sample of students whose teachers had EMC/LRA certifications, and a WWC-computed *p*-value of .03 for the apparently random sample of students; therefore, the WWC does not find these results to be statistically significant. Elementary school classrooms included primarily grades 3–5, with some grade 6 students. Apparently random samples refer to subgroups of schools where the demographic characteristics of the classrooms are similar to the characteristics of the whole school. High-poverty schools are defined as those eligible for the Challenging Schools Bonus, a $5,000 bonus awarded to teachers with *NBPTS*-certification who work in high-poverty schools. Other certifications include all *NBPTS* certification areas except Middle Childhood: Generalist and Early and Middle Childhood: Literacy, Reading, and Language Arts. All analyses included fixed effects for student cohorts. Cohorts were defined by the combination of school, grade, and school year. The number of comparison teachers was estimated by the WWC based on the total number reported by the authors.

[b] For Fisher and Dickenson (2005), the *p*-values presented here were reported in the original study. A correction for clustering and for multiple comparisons within the elementary school grades was needed and resulted in a WWC-computed *p*-value of .09 for grade 6 students not eligible for free/reduced-price lunch; therefore, the WWC does not find the result to be statistically significant. The effect size was calculated using the unadjusted mean and standard deviation calculation.

Appendix D.1b: Description of supplemental findings for the mathematics achievement domain, middle school grades

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Cowan & Goldhaber (2016)[a,14]								
Standardized Math Test	All students	4,535 teachers/ 570,533 students	0.03 (1.02)	–0.02 (0.99)	0.05	0.05	+2	< .01
Standardized Math Test	EL students	4,535 teachers/ 21,912 students	nr	nr	0.06	nr	nr	< .01
Standardized Math Test	FRPL students	4,535 teachers/ 246,335 students	nr	nr	0.06	nr	nr	< .01
Standardized Math Test	Teachers have other certification areas	4,535 teachers/ 514,930 students	nr	nr	0.00	nr	nr	> .05

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Fisher & Dickenson (2005)[b]								
Palmetto Achievement Challenge Test	Grade 7	46 teachers/ 962 student	710.81 (14.64)	710.51 (13.56)	0.30	0.02	+1	.60
Palmetto Achievement Challenge Test	Grade 8	42 teachers/ 680 students	808.26 (12.87)	807.54 (12.84)	0.72	0.06	+2	.17
Palmetto Achievement Challenge Test	Grade 7, FRPL	46 teachers/ 484 students	705.19 (12.85)	705.79 (12.49)	–0.60	–0.05	–2	.50
Palmetto Achievement Challenge Test	Grade 8, FRPL	42 teachers/ 284 students	801.77 (10.44)	801.82 (10.29)	–0.05	–0.01	0	.95
Palmetto Achievement Challenge Test	Grade 7, students non-FRPL	46 teachers/ 478 students	716.51 (14.15)	715.28 (12.97)	1.23	0.09	+4	.11
Palmetto Achievement Challenge Test	Grade 8, non-FRPL	42 teachers/ 396 students	812.91 (12.45)	811.65 (12.94)	1.26	0.01	+4	>.05

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. nr = not reported. FRPL indicates students eligible for free or reduced-price lunch. EL = English learners.

[a] For Cowan and Goldhaber (2016), a correction for multiple comparisons within the middle school grades 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. Middle school classrooms included primarily grades 7–8, with some grade 6 students included. Other certifications include all *NBPTS* certification areas except Early Adolescence: Math. All analyses included fixed effects for student cohorts. Cohorts were defined by the combination of school, grade, and school year. The analyses for students in middle school classrooms and students of teachers with other certification areas in middle school classrooms included student cohort-by-track fixed effects.

[b] For Fisher and Dickenson (2005), a correction for clustering and for multiple comparisons within the table 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 effect size was calculated using the unadjusted mean and standard deviation calculation.

Appendix D.1c: Description of supplemental findings for the mathematics achievement domain, by free/ reduced-price lunch (FRPL) eligibility in grades 4–8

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Fisher & Dickenson (2005)[a]								
Palmetto Achievement Challenge Test	Grades 4–8, FRPL	288 teachers/ 1,594 students	0.00 (1.00)	0.00 (1.00)	0.00	0.00	0	> .99
Palmetto Achievement Challenge Test	Grades 4–8, non-FRPL	4288 teachers/ 1,742 students	0.11 (1.00)	0.00 (1.00)	0.11	0.11	+4	.11

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 Fisher and Dickenson (2005), a correction for clustering was needed but did not affect whether any of the contrasts were found to be statistically significant. The effect size was calculated using the unadjusted mean and standard deviation calculation. The outcomes presented here are based on an aggregated sample of students in grades 4–8 separately reported in the original study. Because the outcome measure was not scaled to allow direct comparisons of scores across grades, the WWC standardized the scores and removed between-grade variation in the outcome means prior to aggregating across grades. The authors reported *p*-values for some results, but not for the aggregated analysis. The WWC applied a correction for clustering and calculated the *p*-value reported in the table.

Appendix D.2a: Description of supplemental findings for the English language arts achievement domain, elementary grades

			Mean (standard deviation)		WWC calculations			
Outcome measure	Study sample	Sample size	Intervention group	Comparison group	Mean difference	Effect size	Improvement index	*p*-value
Cowan & Goldhaber (2016)[a]								
Standardized English Language Arts Test	All students	10,300 teachers/ 742,124 students	0.02 (1.00)	0.00 (1.00)	0.02	0.02	+1	< .01
Standardized English Language Arts Test	EL students	10,300 teachers/ 48,631 students	nr	nr	0.00	nr	nr	> .05
Standardized English Language Arts Test	Special education students	10,300 teachers/ 92,937 students	nr	nr	0.02	nr	nr	< .05
Standardized English Language Arts Test	FRPL students	10,300 teachers/ 331,924 students	nr	nr	0.02	nr	nr	< .01
Standardized English Language Arts Test	Students in high-poverty schools	10,300 teachers/ 105,091 students	nr	nr	0.02	nr	nr	< .10
Standardized English Language Arts Test	Teachers have MC/ GEN certifications	10,300 teachers/ 727,768 students	nr	nr	0.01	nr	nr	> .05
Standardized English Language Arts Test	Teachers have other certifications	10,300 teachers/ 696,335 students	nr	nr	0.03	nr	nr	> .05
Fisher & Dickenson (2005)[b]								
Palmetto Achievement Challenge Test	Grade 4	100 teachers/ 410 students	409.20 (11.24)	407.32 (11.61)	1.88	0.16	+7	.01
Palmetto Achievement Challenge Test	Grade 5	78 teachers/ 374 students	503.83 (11.67)	502.51 (9.76)	1.32	0.12	+5	.08

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Palmetto Achievement Challenge Test	Grade 6	48 teachers/ 848 students	605.78 (14.21)	606.31 (14.16)	–0.53	–0.04	–1	.43
Palmetto Achievement Challenge Test	Grade 4, FRPL	100 teachers/ 188 students	403.31 (10.58)	401.94 (10.96)	1.37	0.13	+5	.22
Palmetto Achievement Challenge Test	Grade 5, FRPL	78 teachers/ 178 students	498.70 (11.18)	497.76 (8.99)	0.94	0.09	+4	.46
Palmetto Achievement Challenge Test	Grade 6, FRPL	48 teachers/ 354 students	599.80 (14.06)	600.19 (12.67)	–0.39	–0.03	–1	.70
Palmetto Achievement Challenge Test	Grade 4, non-FRPL	100 teachers/ 222 students	414.20 (9.21)	411.88 (10.13)	2.32	0.33	+9	.02
Palmetto Achievement Challenge Test	Grade 5, non-FRPL	78 teachers/ 196 students	508.49 (10.08)	506.82 (8.36)	1.67	0.18	+7	.04
Palmetto Achievement Challenge Test	Grade 6, non-FRPL	48 teachers/ 494 students	610.07 (12.71)	610.69 (13.56)	–0.62	–0.05	–2	.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. nr = not reported. MC/GEN = Middle Childhood: Generalist certificate. FRPL indicates students eligible for free or reduced-price lunch. EL = English Learners.

[a] For Cowan and Goldhaber (2016), the *p*-values presented here were reported in the original study. A correction for clustering and for multiple comparisons within the elementary school grades was needed and resulted in a WWC-computed *p*-value of .05 for special education students; therefore, the WWC does not find the result to be statistically significant. Elementary school classrooms included primarily grades 3–5, with some grade 6 students. Apparently random samples refer to subgroups of schools where the demographic characteristics of the classrooms are similar to the characteristics of the whole school. High-poverty schools are defined as those eligible for the Challenging Schools Bonus, a $5,000 bonus awarded to teachers with *NBPTS* certification who work in high-poverty schools. Other certifications include all *NBPTS* certification areas except Middle Childhood: Generalist and Early and Middle Childhood: Literacy, Reading, and Language Arts. All analyses included fixed effects for student cohorts. Cohorts were defined by the combination of school, grade, and school year.

[b] For Fisher and Dickenson (2005), the *p*-values presented here were reported in the original study. A correction for clustering and for multiple comparisons within the elementary school grades was needed and resulted in a WWC-computed critical *p*-value of .006 for grade 4 students and a WWC-computed critical *p*-value of .011 for grade 4 students not eligible for free/reduced-price lunch; therefore, the WWC does not find the results for either outcome to be statistically significant. A correction for clustering was needed and resulted in a WWC-computed *p*-value of .07 for grade 5 students not eligible for free/reduced-price lunch; therefore, the WWC does not find the result to be statistically significant.

Appendix D.2b: Description of supplemental findings for the English language arts achievement domain, middle school grades

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Cowan & Goldhaber (2016)[a]								
Standardized English Language Arts Test	All students	5,811 teachers/ 492,800 students	0.05 (0.95)	0.04 (0.97)	0.01	0.01	+1	< .01
Standardized English Language Arts Test	EL students	5,811 teachers/ 15,212 students	nr	nr	0.03	nr	nr	> .05

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Standardized English Language Arts Test	FRPL students	5,811 teachers/ 210,254 students	nr	nr	0.01	nr	nr	> .05
Standardized English Language Arts Test	Students in high-poverty schools	5,811 teachers/ 107,646 students	nr	nr	0.02	nr	nr	> .05
Standardized English Language Arts Test	Teachers have EA/ELA certifications	5,811 teachers/ 473,693 students	nr	nr	0.01	nr	nr	< .05
Standardized English Language Arts Test	Teachers have other certifications	5,811 teachers/ 442,333 students	nr	nr	0.01	nr	nr	< .05
Fisher & Dickenson (2005)[b]								
Palmetto Achievement Challenge Test	Grade 7	68 teachers/ 898 students	705.71 (11.59)	704.05 (10.80)	1.66	0.15	+6	< .01
Palmetto Achievement Challenge Test	Grade 8	80 teachers/ 1,408 students	806.58 (11.18)	805.27 (11.17)	1.31	0.12	+5	< .01
Palmetto Achievement Challenge Test	Grade 7, FRPL	68 teachers/ 438 students	700.60 (9.81)	700.37 (9.44)	0.23	0.02	+1	.73
Palmetto Achievement Challenge Test	Grade 8, FRPL	80 teachers/ 644 students	802.28 (10.42)	800.20 (9.93)	2.08	0.20	+8	< .01
Palmetto Achievement Challenge Test	Grade 7, non-FRPL	68 teachers/ 460 students	710.57 (11.07)	707.55 (10.86)	3.02	0.28	+11	< .01
Palmetto Achievement Challenge Test	Grade 8, non-FRPL	80 teachers/ 764 students	810.20 (10.50)	809.53 (10.38)	0.67	0.06	+3	.20

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. nr = not reported. EA/ELA = Early Adolescence: English Language Arts certificate.

[a] For Cowan and Goldhaber (2016), a correction for multiple comparisons and for multiple comparisons within the middle school grades 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. Middle school classrooms included primarily grades 7–8, with some grade 6 students included. High-poverty schools are defined as those eligible for the Challenging Schools Bonus, a $5,000 bonus awarded to teachers with *NBPTS* certification who work in high-poverty schools. Other certifications include all *NBPTS* certification areas except Early Adolescence: English Language Arts. All analyses included fixed effects for student cohorts. Cohorts were defined by the combination of school, grade, and school year. The analyses for students in middle school classrooms, students of teachers with EA/ELA certifications in middle school classrooms, and students of teachers with other certification areas in middle school classrooms included cohort-by-track fixed effects.

[b] For Fisher and Dickenson (2005), the *p*-values presented here were reported in the original study. A correction for clustering and for multiple comparisons within the middle school grades was needed and resulted in a WWC-computed *p*-value of .08 for grade 7 students, .09 for grade 8 students; therefore, the WWC does not find the results to be statistically significant. A correction for clustering and multiple comparisons was needed and resulted in a WWC-computed critical *p*-value of .008 for grade 8 students eligible for free/reduced-price lunch and a WWC-computed critical *p*-value of .008 for grade 7 students not eligible for free/reduced-price lunch; therefore, the WWC does not find the result for either outcome to be statistically significant.

Appendix D.2c: Description of supplemental findings for the English language arts achievement domain, by free/reduced-price lunch (FRPL) eligibility in grades 4–8

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Fisher & Dickenson (2005)[a]								
Palmetto Achievement Challenge Test	FRPL students	374 teachers/ 1,802 students	0.10 (1.00)	0.00 (1.00)	0.10	0.10	+4	.11
Palmetto Achievement Challenge Test	Non-FRPL students	374 teachers/ 2,136 students	0.11 (1.01)	0.00 (1.00)	0.11	0.11	+4	.07

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 Fisher and Dickenson (2005), a correction for clustering was needed but did not affect whether any of the contrasts were found to be statistically significant. The effect size was calculated using the unadjusted mean and standard deviation calculation. The outcomes presented here are based on an aggregated sample of students in grades 4–8 that were separately reported in the original study. Because the outcome measure was not scaled to allow direct comparisons of scores across grades, the WWC standardized the scores and removed between-grade variation in the outcome means prior to aggregating across grades. The authors reported *p*-values for some results, but not for the aggregated analysis. The WWC applied a correction for clustering and calculated the *p*-value reported in the table.

Endnotes

[1] The descriptive information for this intervention comes from publicly available sources, specifically intervention websites (http://www.nbpts.org/ and http://www.boardcertifiedteachers.org/, downloaded April 2017). 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 April 2017, 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 maximum amount of time and the requirements to achieve *NBPTS* certification have varied over time.

[3] The literature search reflects documents publicly available by March 2017. 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 (TTEC) 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. The WWC released a single study review of Goldhaber and Anthony (2007) in 2016. This study was previously reviewed in a grant competition in 2016 and was rated as meets standards with reservations. The study was reviewed again under the TTEC protocol for this product and was rated does not meet standards. The difference was based on the grant competition rating a contrast that met standards that is not eligible for the TTEC protocol: comparing newly-certified teachers with teachers who failed certification. In consultation with the TTEC area content experts, we determined this contrast was out of the scope of this review, as the comparison teachers had received some portions of the intervention, and therefore did not represent an untreated condition.

[4] Studies included different locations. Cowan and Goldhaber (2016) included all school districts in Washington state; Fisher and Dickenson (2005) included all school districts in South Carolina; Gardner (2010) included the Brevard County and Seminole County Public School Districts in Florida; Silver (2007) included all school districts in North Carolina; and Stephens (2003) included two counties in South Carolina. Stephens (2003) did not name the included counties.

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

[6] For criteria used to determine the rating of effectiveness and extent of evidence, see the WWC Rating Criteria on p. 42. These improvement index numbers show the average and range of individual-level improvement indices for all findings across the studies.

[7] The study did not report the number of students taught by the teachers, and the author did not respond to an author query.

[8] The WWC identified one additional source related to Cowan and Goldhaber (2016). The study does not contribute unique information to Appendix A.1 and is not listed here.

[9] Weighted averages for each demographic were calculated by weighting the elementary and middle school demographic characteristics by their share of the total student sample examined in the study.

[10] The study also examined the effect of subgroups of teachers on student mathematics and English language arts achievement based on whether the teacher passed *NBPTS* certification on the first or second attempt, and their scores for each attempt; these contrasts are ineligible for review because they do not focus on a subgroup of interest in the Teacher Training, Evaluation, and Compensation review protocol.

[11] Fisher and Dickenson (2005) also examined outcomes using hierarchical linear models and what the authors refer to as a "pilot analysis," which included all teachers and students observed without any matching to balance baseline achievement; these contrasts do not meet WWC group design standards because equivalence of the analytic intervention and comparison groups is necessary and not demonstrated.

[12] The study examined outcomes in both the 2003–04 and 2004–05 school years. However, the WWC review focused only on the outcomes measured in the 2004–05 school year, as all intervention teachers were fully *NBPTS*-certified at the beginning of this school year. These teachers were still in the certification process at the beginning of the 2003–04 school year, and therefore, students in the intervention condition did not receive a full year of instruction from *NBPTS*-certified teachers. In addition, the WWC used the 2002–03 school year as the baseline for assessing equivalence of the intervention and comparison conditions, for the same reason.

[13] Cowan and Goldhaber also present several mathematics and English language arts impact estimates among a subgroup of elementary school students they refer to as an "apparently random sample." This subgroup was identified by limiting to students whose classroom demographic characteristics were similar to the school-level demographics. In other words, there was no evidence of student sorting by classrooms. These findings were generally of the same magnitude as those using the full sample of students, but most were not statistically significant.

[14] Cowan and Goldhaber also present several mathematics and English language arts impact estimates among middle school students using cohort-by-track fixed effects. These findings did not differ from the analyses of the same outcome using only cohort fixed effects.

Recommended Citation

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2018, February). *Teacher Training, Evaluation, and Compensation intervention report: National Board for Professional Teaching Standards Certification.* 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 without reservations, 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 without reservations, 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.

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. 41.

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.

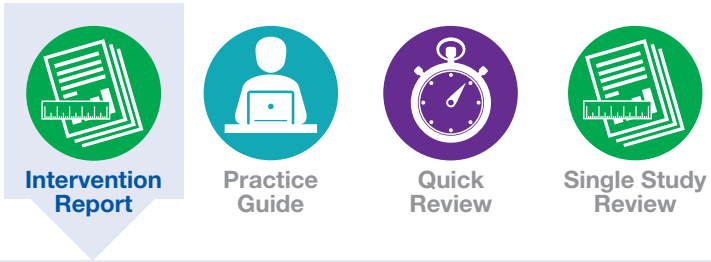
Statistical significance Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p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

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.