Making an Impact

Impacts of the Retired Mentors for New Teachers program

Dale DeCesare
Abby McClelland
Augenblick, Palaich and Associates

Bruce Randel
Century Analytics

Key findings

A randomized controlled trial of a two-year intervention using retired mentors to support teachers who were in their first three years, conducted in collaboration with Aurora Public Schools in Colorado and using 2013/14 and 2014/15 data, yielded several causal findings:

- At the end of the first year math achievement was significantly higher among students taught by teachers in the program group than among students taught by teachers in the business-as-usual group.
- Although the differences were not statistically significant, reading achievement was higher among students taught by teachers in the program group than among students taught by teachers in the business-as-usual group.
- The program's effect on teacher evaluation ratings and teacher retention was not significant after two years.

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences
U.S. Department of Education

REL CENTRAL
Regional Educational Laboratory
At Marzano Research

Summary

This study evaluates the impact of the Retired Mentors for New Teachers program, a two-year mentoring program at the elementary school level developed by Aurora Public Schools in Colorado. Many of the district's schools serve a large percentage of economically disadvantaged children, experience high teacher turnover, and hire newer, less experienced teachers. The program addresses these challenges using master educators who recently retired from the district to provide tailored one-on-one mentoring to new teachers. The program requires mentees to meet weekly one-on-one with their mentor and monthly in school-level groups over the course of two years.

This study was undertaken by Regional Educational Laboratory (REL) Central in collaboration with Aurora Public Schools. It used a randomized controlled trial to assess the impacts of the Retired Mentors for New Teachers program on student achievement, teacher retention, and teacher evaluation ratings during the 2013/14 and 2014/15 school years. As part of the study, the district's elementary school teachers were randomly assigned to either a group that received only the district's typical mentoring support (the business-as-usual group) or a group that received both typical mentoring support and added support from a retired mentor under the Retired Mentors for New Teachers program (the program group). The business-as-usual support involved first-year teachers being assigned to work with a more experienced "buddy" teacher for about 15 hours over one school year. The study team collected data on resources and costs associated with running the program and generated a return on investment estimate. The study team also used teacher and mentor surveys and focus groups to investigate whether the program was implemented with fidelity to its model.

The following key findings emerged from the study. The first four are causal findings; the last four are exploratory findings from exploratory analyses:

- *At the end of the first year math achievement was significantly higher among students taught by teachers in the program group than among students taught by teachers in the business-as-usual group.* The difference represents nearly 12 percent of the academic growth in math that students are expected to make over an entire year (Northwest Evaluation Association, 2011). This is equivalent to one month of additional instruction time in a typical nine-month school year. A similar positive effect was observed at the end of the second year, with an effect size of 0.06, although it was not statistically significant.
- *While the differences were not statistically significant, reading achievement was also higher among students taught by teachers in the program group than among students taught by teachers in the business-as-usual group.* The difference in reading achievement after the second year was positive, with an effect size of 0.07, but not statistically significant.
- *The program had no effect on teacher evaluation outcomes.* The likelihood of a teacher receiving a higher overall evaluation rating (for example, proficient instead of partially proficient) was lower for teachers in the program group than for teachers in the business-as-usual group, but the difference was not statistically significant.
- *Although more teachers in the program group than teachers in the business-as-usual group left the district after two years, the effect of the program on teacher retention was not significant.* Three teachers left the district after the first year: one from

the program group and two from the business-as-usual group. More teachers left after the second year: 11 from the program group and 5 from the business-as-usual group. The differences between the two groups were not statistically significant.

- *Increased hours of mentoring were associated with higher teacher retention in the second year among teachers who participated in the program.* Mentees who received more hours of mentoring were more likely to stay in the district. The odds of a mentee staying in the district doubled with each additional 10 hours of mentoring. The sharpest increase in retention occurred with each additional hour received after 25 hours. Teachers who stayed in the district could have received more mentoring hours simply because they were there to receive them. The analysis of the effect of mentoring hours was conducted only on program group teachers and therefore was not part of the randomized trial impact study. Nevertheless, the potential for a causal connection between increased mentoring and increased retention may be worth further study.

- *The program had an average annual local cost of approximately $171 per student.* That cost was based on local resource prices. The representative national cost was about $237 per student, based on a coversion of the local cost usng a national database.

- *Exploratory analysis suggested that the program could yield a return on investment that may pay back the annual cost of the program more than 15 times over through increased student earnings over time.* The study team conducted an in-depth analysis of the program's costs, including staff time and resources. The study team also compared the program's statistically significant impact findings with regard to student math achievement to those of a rigorous national study on the relationship between increased student test scores due to teachers' influence and student outcomes in adulthood. The combined cost and impact analysis found that student lifetime earnings increases could pay back the annual cost of the program more than 15 times over.

- *Overall, the program was implemented with fidelity to its intended model.* The fidelity of program implementation was assessed on four components: provision of highly qualified mentors, mantees' access to their mentor in the summer prior to the first year, mentor team meetings, and coaching and direct support for mentees. Overall, the program met implementation thresholds for all components except mantees' access to their mentor prior to the first year.

Contents

Why this study?

Teacher quality is the most important school-related factor affecting student academic achievement (Organisation for Economic Co-operation and Development, 2005). Yet research indicates that beginning teachers, especially teachers in their first two years, are less effective than are teachers with more experience (Clotfelter, Ladd, & Vigdor, 2006; Rice, 2010). Research also indicates that 40–50 percent of U.S. public school teachers leave teaching within five years of entering the profession (Ingersoll & Strong, 2011). Such turnover represents a considerable cost to districts and schools, with some estimates finding that overall teacher attrition (including new and veteran teachers) costs the United States more than $2 billion a year (Alliance for Excellent Education, 2014; Barnes, Crowe, & Schaefer, 2007). Because of the potential costs, in terms of both teacher effectiveness and turnover, it is important for education leaders and policymakers to have access to rigorous research on innovative approaches that can improve new teachers' instruction practice and retention in the profession.

Providing new teachers with an experienced, effective mentor is one such possible innovative approach. In fact, several studies show the promise of mentoring as an effective intervention for addressing teacher retention and student achievement (Glazerman et al., 2010; Gray, Taie, & O'Rear, 2015; Ingersoll & Strong, 2011; Smith & Ingersoll, 2004; Strong, 2006).

The current study contributes to the existing research on mentoring in several important ways. First, it uses a randomized controlled trial. Most research studies examining mentoring use nonexperimental designs. Few studies provide experimental evidence assessing mentoring program effects.

Second, the study collects data on program costs and return on investment. Such data can help district leaders understand both the resources that may be needed to implement a similar program in their own jurisdiction and the potential returns over time. This information is particularly important for districts operating in constrained fiscal environments.

Third, the current study focuses on the use of retired educators as mentors. This is critical because:

- The nation's teaching population is older than it has ever been, and at least 1.8 million K–12 teachers and leaders are reaching retirement age between 2010 and 2020 (Foster, 2010).
- There is evidence that more than half of baby boomer teacher retirees are interested in further employment, and about 70 percent of those interested in such employment indicate interest in working in new education roles (Foster, 2010).
- Retired educators can have scheduling flexibility to observe or meet with a mentee teacher before, during, or after the school day and are a potentially high-quality, lower cost resource (compared, for instance, to the cost of releasing teachers from their classroom duties to serve as mentors).

The convergence of the ongoing need to support new teachers, the promise of mentoring as an intervention, and the aging of the U.S. teacher population mean there is a need to rigorously examine the costs and potential impacts of using retired educators to mentor new teachers. This study addresses this need by looking at the impacts and costs

The convergence of the ongoing need to support new teachers, the promise of mentoring as an intervention, and the aging of the U.S. teacher population mean there is a need to rigorously examine the costs and potential impacts of using retired educators to mentor new teachers

of a district program that uses recently retired master educators to mentor new teachers in high-need elementary schools.

Mentoring approaches in Aurora Public Schools

Aurora Public Schools is the fifth-largest school district in Colorado, with 64 schools and more than 40,000 students in grades preK–12. This student population is diverse, and many of the district's elementary schools are high-poverty schools.

The district's business-as-usual mentoring approach at all grade levels is a "buddy" model, similar to that used in many other districts. First-year teachers are assigned to work with a more experienced "buddy" teacher, usually from the same school, for one year (table 1). A mentor's experience, availability, and willingness to work with new teachers on top of all other normal classroom duties can vary substantially, and there is no formal process for selecting or assigning qualified buddy mentors. There is no reduction in the mentor's course load or other responsibilities in order to make time for mentoring. The mentor is paid a $500 stipend for the school year. The district expects a mentor to spend approximately 15 contact hours with a mentee over the year, but district monitoring of the mentoring process is limited. The only reporting requirement is paperwork at the end of the school year stating that mentoring was provided. School and district leaders generally do not set a high priority on providing common planning periods for mentors and mentees, and there are no expectations for the amount of time mentors are expected to spend observing classrooms or offering feedback to mentees on instruction.

The business-as-usual mentoring approach for first-year teachers in Aurora Public Schools is a "buddy" model; the district's Retired Mentors for New Teachers program uses recently retired master educators to provide one-on-one mentoring to probationary teachers

The mentoring provided under the buddy mentoring model is typically driven by the willingness and availability of the mentor to spend time with the mentee. This availability can vary considerably, and the lack of a common schedule between mentor and mentee means that mentoring time must be planned by each individual mentor–mentee pair.

Some mentees in the buddy mentoring program indicate an unwillingness to share their challenges with more senior teachers in the building for fear that such information could reach their principal and be used against them in performance evaluations. The lack of assurance that mentoring is separate from the evaluation process can thus impede the formation of a relationship of trust between mentor and mentee.

The district's Retired Mentors for New Teachers program—which operates only at the elementary school level—uses master educators (teachers and principals) who recently retired from the district to provide one-on-one mentoring to probationary teachers (teachers who are in their first three years with the district, regardless of prior teaching experience). According to the program's theory of change and logic model, the performance of probationary teachers, especially those serving in schools with a high percentage of at-risk students, can be improved by receiving two years of consistent mentoring (including classroom observation and feedback) from a retired educator who is well versed in the district's expectations and who has no input into the mentee's evaluation (figure 1). In particular, mentoring is not designed to be one size fits all but is instead tailored to each mentee.

To participate in the program, retired educators must have at least five years of teaching experience in the district (most had 10 or more years in the district and 20 or more years in education) and a record of excellence as reflected by student achievement, performance

Table 1. Comparison of mentoring approaches in Aurora Public Schools, 2013/14 and 2014/15

Component	Business-as-usual approach ("buddy" model)		Retired Mentors for New Teachers approach	
	First year	Second year	First year	Second year
Meetings between mentors	None	None	At least quarterly meetings of mentors, typically led by the district head of professional development, for sharing and reviewing mentoring strategies and district priorities	At least quarterly meetings of mentors for sharing and reviewing mentoring practice strategies
Summer professional development for mentees[a]	None	None	Two half-days at the end of July, prior to start of the school year	None
Coaching and in-classroom support	Teachers in their first year are assigned a "buddy" mentor in the building. Mentors are expected to log at least 15 hours of contact time, but there are no expectations for either frequency of visits or that visits include classroom observations	None	An average of one hour of individualized, tailored support per week, per mentee. Mentees in their first year also receive business-as-usual support and are assigned a "buddy" mentor.	An average of one hour of individualized, tailored support per week, per mentee
Cohort support	None	None	One- to two-hour meeting per month of mentee cohorts before, during, or after school	One- to two-hour meeting per month of mentee cohorts before, during, or after school
Principal meetings with mentors	None. Mentors fill out end-of-year paperwork to confirm that mentoring took place	None	Quarterly meeting at each school site	Quarterly meeting at each school site

a. Summer professional development, though considered a typical program component, did not take place during the current study because of the need to conduct random assignment of teachers at the start of the first school year

Source: Authors' compilation based on data from Aurora Public Schools.

reviews, leadership, and reputation as an effective educator. The district head of professional development interviews each mentor to confirm compatibility.

The program also calls for individual mentee–mentor meetings and group meetings with other mentees (led by a retired mentor) in the same school. The mix of one-on-one and group meetings allows mentors to tailor support to each mentee while building a model of collaboration across mentees that can be sustained after mentoring ends (see table 1). Mentees in their first year also receive the district's business-as-usual support and are assigned a "buddy" mentor in addition to their retired mentor.

Retired educators can use a mix of supports, including modeling the classroom setup, co-teaching, classroom observation and feedback, modeling lesson design and delivery, analysis and reflection on data, literature studies, and organizing mentee visits to observe model classrooms. The program relies on each mentor's professional judgment and expertise to determine an appropriate mix of approaches for each mentee.

Figure 1. Theory of change and logic model for the Retired Mentors of New Teachers program in Aurora Public Schools

Theory of change: By providing probationary teachers (teachers who are in their first three years with the district, regardless of prior teaching experience) in high-need schools with regular coaching and support from a recently retired master educator, the Retired Mentors for New Teachers program enhances probationary teachers' instruction practices, thus increasing student achievement on district assessments and improving teacher evaluations and retention.

Current situation and need diagnosis	Major intervention components	Improved outcomes
Students in high-need elementary schools continue to underperform in reading and math Probationary teachers require added support to deliver effective reading and math instruction Retaining teachers in high-need schools remains an ongoing challenge High teacher turnover forces higher percentages of probationary teachers into high-need schools	Pair probationary teachers in high-need elementary schools with recently retired master educators who have: · History of success in the district · Years of experience with district expectations and student challenges · Flexibility to meet with probationary teachers before, during, or after school · No input into teacher evaluations Mentoring provided over two school years Two half-days of summer professional development for probationary teachers Individualized mentoring and classroom support Cohort support through meetings with other probationary teachers Mentor meetings with principals to ensure common understanding of school priorities Quarterly mentor meetings to discuss and continually improve practice	Improved instruction delivered by probationary teachers increases their students' achievement as measured by math and reading assessments Improvement in principal evaluations of probationary teachers Reduced probationary teacher turnover

Source: Authors' compilation based on data from Aurora Public Schools.

What the study examined

This study focused on eight research questions. The first four are causal, and the remaining four are exploratory:

- What is the impact of the Retired Mentors for New Teachers program on elementary school students' scores on math benchmark assessments?
- What is the impact of the Retired Mentors for New Teachers program on elementary school students' scores on reading benchmark assessments?
- What is the impact of the Retired Mentors for New Teachers program on teacher evaluation ratings?
- What is the impact of the Retired Mentors for New Teachers program on teacher retention?
- What is the association between hours of mentoring and teacher retention?
- What is the cost of the Retired Mentors for New Teachers program?
- What is the potential return on investment of the Retired Mentors for New Teachers program?
- Was the Retired Mentors for New Teachers program implemented with fidelity to its model?

The study used a randomized controlled trial to address the four causal impact questions and exploratory analyses to address the other four questions. Box 1 summarizes the data and outcome measures used in the study, and appendixes A–D provide additional details on data and methodology.

Box 1. Data and outcome measures

Data

This study was conducted using data on 77 teachers at 11 Title I elementary schools in Aurora Public Schools for the 2013/14 and 2014/15 school years (see appendix A for more details). Participating schools had at least 80 percent of students eligible for the federal school lunch program, a proxy for low-income status. Teachers in the study were all probationary teachers, defined as teachers in their first three years with the district, regardless of prior teaching experience (which ranged from 0 to 15 years). At the start of the study all teachers in the study taught core classroom subjects (math or reading) in grades 1–5 and had an average of 1.4 years of experience in the district. Teachers in the business-as-usual group had an average of 3 years of experience in education, and teachers in the program group had an average of 3.7 years of experience in education. None of the differences in teacher characteristics between the groups was significant.

Within each participating school, half the probationary teachers were randomly assigned to receive the district's business-as-usual mentoring support, and the other half received the Retired Mentors for New Teachers program plus business-as-usual support. Each retired mentor was assigned a caseload of three to seven mentees (typically three to four per school) in one to two elementary schools.

Outcome measures

Student math and reading achievement was measured using scores on the Measures of Academic Progress assessment, administered to students in the early fall and late spring of each school year. The assessment is an online adaptive test with separate sections for math and reading. Students' fall scores were the baseline, and spring scores were the outcome measure. Student achievement impacts were estimated comparing average assessment scores for students taught by teachers in the program group with assessment scores for students taught by teachers in the business-as-usual group. The analysis controlled for students' fall scores to improve the precision of the estimates and to adjust for any differences in student achievement that existed before exposure to the program. Sensitivity analyses of student outcome impact estimates were also conducted (see appendix B).

Teacher retention was measured using data from the district's department of human resources. Teachers were considered to have stayed if they were still teaching in the district at the start of the following school year, even if they changed grade levels or schools. Teachers were considered to have left if they were not teaching in the district at the start of the following school year. Teacher retention was measured twice: at the beginning of the second year (2014/15) and at the beginning of what would have been the third year (2015/16).

Teacher evaluation outcomes were measured using district evaluation data. Teachers received scores for five components: demonstrating mastery of and pedagogical expertise in the content taught; establishing a safe, inclusive, and respectful learning environment for a diverse population of students; planning and delivering effective instruction and creating an environment that facilitates learning for students; reflecting on practice; and demonstrating

(continued)

Box 1. Data and outcome measures *(continued)*

leadership. Teachers also received an overall composite evaluation score based on a weighted average of their component scores. Both component scores and the overall score were converted to a rating of basic, partially proficient, proficient, accomplished, or exemplary.

The study team used an elements approach to collect program cost information as recommended by Levin & Belfield (2013). The elements included personnel costs, facilities costs, equipment and materials costs, and other program costs (see appendix C for more details).

To calculate the program's return on investment, the study team used an approach similar to that used by Chetty, Friedman, and Rockoff (2011), which focused on student impacts in a single year as a result solely of teacher effectiveness. Under this approach, the costs associated with the key components of the program represent the district's investment in the program, and outcomes in adulthood represent the return on the district's investment.

An implementation analysis was conducted to assess whether the program operated with fidelity to its design, to facilitate interpretation of impact findings, and to inform replication of the program in other districts. Surveys of teachers in the program and teachers in business-as-usual groups were conducted at the end of each year, as were focus groups with mentors and teachers in the program group. Monthly records of support were completed online by mentors to track mentee dosage and key areas of mentor support (see appendix D for more details).

What the study found

This section presents findings on the Retired Mentors for New Teachers program, including the impacts on student math achievement, student reading achievement, teacher retention, and teacher evaluation ratings; the association between hours of mentoring and teacher retention; the cost of operating the program; the potential return on investment; and the fidelity of implementing the program.

At the end of the first year math achievement was significantly higher among students taught by teachers in the program group than among students taught by teachers in the business-as-usual group

After the first year of the program, students taught by teachers in the program group scored 1.4 points higher on the spring Measures of Academic Progress math assessment than did students taught by teachers in the business-as-usual group (table 2). This difference was statistically significant. Elementary school students are expected to gain an average of 12.1 points over the course of a school year (Northwest Evaluation Association, 2011). The difference in math achievement observed for students in the study represents nearly 12 percent of the academic growth a student is expected to make over an entire year and is roughly equivalent to one month of additional instruction time in a typical nine-month school year. See appendix E for complete regression results for math achievement at the end of the first year.

After the first year of the program, students taught by teachers in the program group scored 1.4 points higher on the spring Measures of Academic Progress math assessment than did students taught by teachers in the business-as-usual group

At the end of the second year, math achievement was also higher among students taught by teachers in the program group than among students taught by teachers in the business-as-usual group, but the difference was not statistically significant (see table 2). The results are somewhat sensitive to model specification and adjustment for multiple comparisons, meaning that including or excluding specific variables in the model could change the findings (see appendix B).

Table 2. Scores on the Measures of Academic Progress math assessment among elementary school students in Aurora Public Schools taught by teachers in the program group compared with teachers in the business-as-usual group, by year

| Year | Adjusted mean score (standard deviation) | | Difference (standard error) | *p*-value | Effect size[c] |
	Students taught by teachers in the business-as-usual group[a]	Students taught by teachers in the program group[b]			
First year (2013/14)	178.9 (21.28)	180.3 (22.32)	1.4* (0.70)	.045	0.064
Second year (2014/15)	178.3 (21.66)	179.7 (24.22)	1.4 (0.86)	.113	0.061

* Significantly different from zero at the .05 level, two-tailed test. The Benjamini–Hochberg adjustment for multiple comparisons lowers the *p*-value threshold to .025, and the finding is not significant with the adjustment applied.

Note: A regression model with a dichotomous indicator for treatment status was used to estimate average differences between the two groups. The model included prior achievement, as measured by fall Measures of Academic Progress math scores, student demographics, teacher education and experience, class size, grade level, and dummy variables representing school sites. The sample included 1,189 students and 70 teachers in the first year and 1,105 students and 62 teachers in the second year. Standard errors were estimated with adjustment for clustering of students within teachers.

a. Refers to teachers who received mentoring through the district's typical "buddy" mentoring approach.

b. Refers to teachers who received mentoring through the Retired Mentors for New Teachers program in addition to the district's typical "buddy" mentoring approach.

c. Calculated by dividing the difference in adjusted mean score between students taught by teachers in the business-as-usual group and students taught by teachers in the program group by the pooled standard deviation of the outcome variable.

Source: Authors' analysis of data from Aurora Public Schools for 2013/14 and 2014/15; see appendix A.

Reading achievement was higher among students taught by teachers in the program group than among students taught by teachers in the business-as-usual group, but the differences were not statistically significant

At the end of the first year the difference in Measures of Academic Progress reading assessment scores between students taught by teachers in the program group and students taught by teachers in the business-as-usual group was 0.3 (or 0.014 standard deviation) and not statistically significant. At the end of the second year the difference between reading scores was 1.42 points, the equivalent of an effect size of 0.066, and approached but did not reach statistical significance at the 0.05 level (table 3). These positive effects on reading achievement in the second year are essentially the same in magnitude as the effects on student achievement in math in both years. The results are somewhat sensitive to model specification, meaning that including or excluding specific variables in the model sometimes changes the findings, as detailed in appendix B.

The Retired Mentors for New Teachers program had no effect on teacher evaluation outcomes

In both the first and second years of the study there was no significant difference in the likelihood of teachers in either group receiving a higher overall evaluation rating (for example, proficient instead of partially proficient; table 4). There were also no clear differences in the five evaluation component ratings between teachers in the two groups.

The difference in reading assessment scores between students taught by teachers in the program group and students taught by teachers in the business-as-usual group was 0.3 points at the end of the first year and 1.42 points at the end of the second year; neither difference was statistically significant

Table 3. Scores on the Measures of Academic Progress reading assessment among elementary school students in Aurora Public Schools taught by teachers in the program group compared with teachers in the business-as-usual group, by year

| Year | Adjusted mean score (standard deviation) | | Difference and standard error | p-value | Effect size[c] |
	Students taught by teachers in the business-as-usual group[a]	Students taught by teachers in the program group[b]			
First year (2013/14)	185.3 (20.86)	185.6 (22.29)	0.3 (0.70)	.637	0.014
Second year (2014/15)	172.6 (20.73)	174.0 (21.44)	1.4 (0.81)	.078	0.066

Note: A regression model with a dichotomous indicator for treatment status was used to estimate average differences between the two groups. The model included prior achievement, as measured by fall Measures of Academic Progress reading scores, student demographics, teacher education and experience, class size, grade level, and dummy variables representing school sites. The sample included 1,155 students and 66 teachers in the first year and 964 students and 57 teachers in the second year. Standard errors were estimated with adjustment for clustering of students within teachers.

a. Refers to teachers who received mentoring through the district's typical "buddy" mentoring approach.

b. Refers to teachers who received mentoring through the Retired Mentors for New Teachers program in addition to the district's typical "buddy" mentoring approach.

c. Calculated by dividing the difference in adjusted mean score between students taught by teachers in the business-as-usual group and students taught by teachers in the program group by the pooled standard deviation of the outcome variable.

Source: Authors' analysis of 2013/14 and 2014/15 data from Aurora Public Schools; see appendix A.

Table 4. Odds ratios for receiving a higher evaluation rating between teachers in the program group and teachers in the business-as-usual group, by evaluation component and year

| Evaluation component | First-year evaluations (2013/14) | | | Second-year evaluations (2014/15) | | |
	Odds ratio	Standard error	p-value	Odds ratio	Standard error	p-value
Overall	0.45	0.21	.090	0.77	0.38	.598
Component 1: Demonstrating mastery of and pedagogical expertise in the content taught	0.62	0.27	.280	0.56	0.28	.248
Component 2: Establishing a safe, inclusive, and respectful learning environment for a diverse population of students	0.94	0.42	.887	0.80	0.36	.625
Component 3: Planning and delivering effective instruction and creating an environment that facilitates learning for students	0.76	0.34	.536	0.58	0.31	.301
Component 4: Reflecting on practice	0.59	0.26	.234	0.50	0.23	.126
Component 5: Demonstrating leadership	0.99	0.46	.983	1.49	0.72	.415

Note: An ordinal logistic regression model with a dichotomous indicator for treatment status was used to estimate average differences in the odds of achieving each higher step on the evaluation rating scale between teachers in the program group and teachers in the business-as-usual group. The model included teacher demographics. An odds ratio greater than 1 indicates that being in the program group was associated with higher odds of receiving a higher evaluation rating, an odds ratio less than 1 indicates that being in the program group was associated with lower odds of receiving a higher evaluation rating, and a value of 1 indicates that the odds of receiving a higher evaluation rating were the same for both groups.

Source: Authors' analysis of 2013/14 and 2014/15 data from Aurora Public Schools; see appendix A.

The effect of the Retired Mentors for New Teachers program on teacher retention was not significant after two years

Three teachers left Aurora Public Schools after the first year: one from the program group and two from the business-as-usual group. More teachers left after the second year: 11 from the program group and 5 from the business-as-usual group. Thus the retention rate was lower among teachers in the program group than among teachers in the business-as-usual group. Participating in the mentor program did not, on its own, have a significant effect on an individual teacher's odds of remaining in the district after two years (table 5).

Increased hours of mentoring were associated with higher teacher retention in the second year among teachers in the program group

Among teachers in the program group who remained teaching in the district at the beginning of year two, those who received more hours of mentoring from their retired mentor were significantly more likely to stay in the district after year two than were teachers who received fewer hours (figure 2). The odds of a mentee staying in the district doubled with each additional 10 hours of mentoring. Even one additional hour of mentoring increased the odds of staying by 12 percent, a statistically significant result ($p = 0.004$). At 40 hours of mentoring and above, teachers approach 100 percent likelihood of staying in the district (see figure 2). Seventy-four percent of teachers in the program group received 40 or more hours of mentoring from a retired mentor. All teachers in the program group who received more than 60 hours of mentoring over the two-year period stayed in the district. The amount of mentoring that teachers in the program group received over the two years of the study ranged from 7 hours to 167 hours, with an average of 65 hours. The analysis excludes the single teacher in the program group who left the district after the first year.

Since this analysis was correlational in nature, conclusions cannot be drawn about the direction of the effect. One possibility is that hours of mentoring had a positive impact on retention. However, there are at least two alternative explanations. First, teachers who remained in the district after year two could have received more hours of mentoring simply because they were there to receive them. Second, teachers motivated to stay in the district could have sought additional hours of mentoring.

The retention rate was lower among teachers in the program group than among teachers in the business-as-usual group, but participating in the mentor program did not, on its own, have a significant effect on an individual teacher's odds of remaining in the district after two years

Table 5. Odds ratios for staying in Aurora Public Schools among teachers in the program group and teachers in the business-as-usual group, by year

Year	Odds ratio	Standard error	*p*-value
First year (2013/14)	1.00	1.58	.528
Second year (2014/15)	0.48	0.29	.220

Note: The program group refers to teachers who received mentoring through the Retired Mentors for New Teachers program in addition to the district's typical "buddy" mentoring approach. The business-as-usual group refers to teachers who received mentoring through the district's typical "buddy" mentoring approach. A logistic regression model with a dichotomous indicator for treatment status was used to estimate average differences between the odds of retention for the two groups. The model included teacher demographics. An odds ratio greater than 1 indicates that being in the program group was associated with higher odds of retention, an odds ratio less than 1 indicates that being in the program group was associated with lower odds of retention, and a value of 1 indicates that the odds of retention were the same for both groups.

Source: Authors' analysis of 2013/14 and 2014/15 data from Aurora Public Schools; see appendix A.

Figure 2. Among teachers in the program group who stayed in the district after the first year (2013/14), those who received more hours of mentoring over two years were significantly more likely to return to Aurora Public Schools for a third year than were teachers who received fewer hours
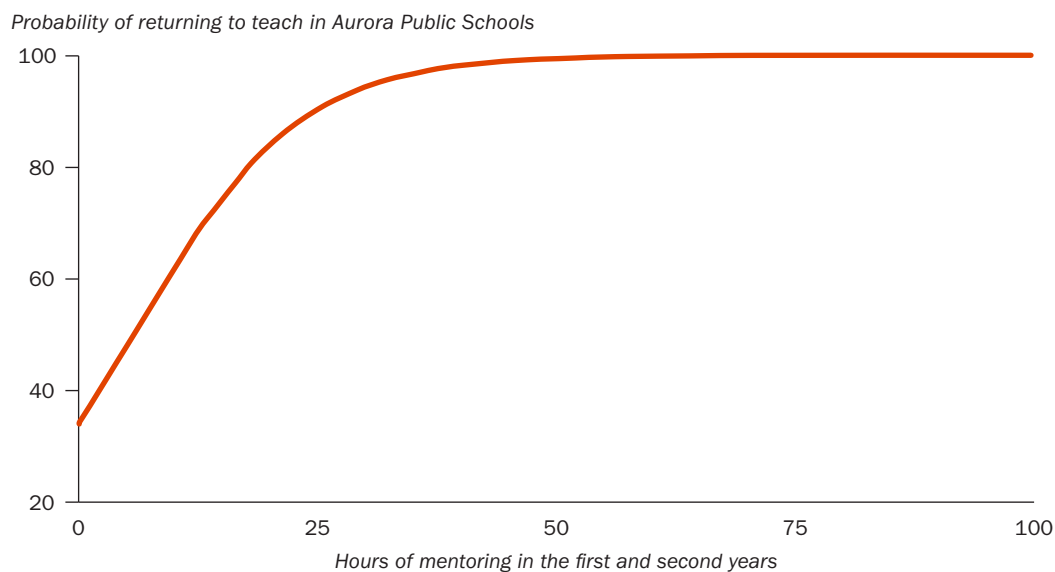
Probability of returning to teach in Aurora Public Schools

Hours of mentoring in the first and second years

The odds of a mentee staying in the district doubled with each additional 10 hours of mentoring. Even one additional hour of mentoring increased the odds of staying by 12 percent. At 40 hours of mentoring and above, teachers approach 100 percent likelihood of staying in the district

Note: The program group refers to teachers who received mentoring through the Retired Mentors for New Teachers program in addition to the district's typical "buddy" mentoring approach.

Source: Authors' analysis of 2013/14 and 2014/15 data from Aurora Public Schools; see appendix A.

Table 6. Comparison of average annual local and national costs of the Retired Mentors for New Teachers program in Aurora Public Schools, 2013/14 and 2014/15 ($)

Cost	Average annual program cost	Average annual cost per student	Average annual cost per mentee
Local Aurora Public Schools cost	126,015	171	3,765
National cost	174,355	237	5,210

Source: Authors' calculations based on 2013/14 and 2014/15 data from Aurora Public Schools and Center for Benefit-Cost Studies of Education (2015).

The Retired Mentors for New Teachers program had an annual local cost of about $171 per student

The average annual cost to Aurora Public Schools per student served by teachers in the program, based on local resource prices, was about $171 (table 6; see appendix C for the elements considered in the analysis of program cost). The study team also used the Center for Benefit-Cost Studies of Education (2015) national database to convert the local cost to a representative national cost of $237, based on national average salary and benefit rates for similar personnel.

The Retired Mentors for New Teachers program demonstrates a return on investment that may pay back the annual cost of the program more than 15 times over through increased student earnings over time

After the first year the effect size of the difference in math test scores between students taught by teachers in the program group and students taught by teachers in the

Table 7. Potential long-term returns from a 0.06 effect size increase in math assessment scores, by student outcome measure, 2013/14 and 2014/15

Student outcome measures	Impact of 0.06 effect size increase in math assessment scores
College attendance	The probability of attending college at age 20 increases 0.3 percentage points (relative to a mean of 37.8 percent)
Income	Student lifetime earnings increase $2,760
Teenage birth	The probability of having a teenage birth is reduced 2.3 percentage points (relative to a mean of 8 percent)

Source: Authors' calculations based on 2013/14 and 2014/15 data from Aurora Public Schools and Chetty et al. (2011; see appendix C).

business-as-usual group was 0.06. To translate this effect into practical terms, the study team reviewed the literature on the relationship between increased student test scores due to teachers' influence and student outcomes in adulthood. One study explored the long-term effects of having a teacher who increases student test scores in a single school year when the student is in elementary school solely as a result of teacher effectiveness, not as a result of specific curriculum or other nonteacher factors (Chetty et al., 2011). The findings of that study are relevant to the test score gains produced by teachers in the program group in the current study. The findings include increases in the probability of college attendance, increases in annual earnings, and reductions in the probability of teenage births (table 7).

The estimated $2,760 increase in lifetime income for the 704 students taught math in the first year by teachers in the program group amounts to a potential combined increase of nearly $2 million over the earnings of students taught by teachers in the business-as-usual group. That figure takes into account both a discount rate of 5 percent (a dollar earned in the future is worth less than a dollar earned today because of the opportunity for earnings from investing today's dollar) and fadeout of the program's effect over time.

These long-term potential outcomes—associated with a single year of increased student assessment scores in elementary school—are found even when student characteristics, including parents' income and education, are controlled for. The estimated lifetime earnings increase could therefore potentially pay back the annual cost of the program more than 15 times over.

Overall, the Retired Mentors for New Teachers program was implemented with fidelity to its intended model

The fidelity of program implementation was assessed on four components: provision of highly qualified mentors, mentees' access to their mentor prior to the first year, mentor team meetings, and coaching and direct support for mentees. The study team worked with district leaders to establish thresholds for each component to provide benchmarks for what the district considers adequate implementation of the program. Overall, the program met implementation thresholds with fidelity for three of the four components. See appendix D for the component thresholds and the full results for this analysis.

The estimated $2,760 increase in lifetime income for the 704 students taught math in the first year by teachers in the program group amounts to a potential combined increase of nearly $2 million over the earnings of students taught by teachers in the business-as-usual group

The only component that was not implemented with fidelity was mantees' access to summer professional development with their mentor for two half-days prior to the first year. Ordinarily, the program is implemented from late July through early May. But during the first year of the study, support was shortened by almost two months. Mentoring did not fully start until October because of the added time needed to enroll schools in the study, conduct random assignment, and conduct student assessments. So support was curtailed for mentees, largely because of considerations involving this study. In addition, the program typically focuses on first-year teachers and those new to the district. But to increase the statistical power of the study, second- and third-year probationary teachers were included.

In addition to studying the fidelity with which the program was implemented, the study team also gathered data on educator perspectives. These perspectives were gathered through analysis of mentee survey data and qualitative analysis of themes emerging from mentee and mentor focus groups:

- Both mentees and mentors in focus groups valued the two-year length of the program. The second year was viewed as important in allowing mentors and mentees to fully leverage the relationship that was built during the first year.
- Both mentors and mentees in focus groups indicated that the first year of mentoring often focused on classroom management. This was viewed as critical to boost mentee confidence to tackle improvement in instruction and content in the second year.
- Mentors and mentees in focus groups noted that mentors, as retirees, had the scheduling flexibility to plan meetings at times most convenient for mentees (before, during, or after school).
- Mentees and mentors in focus groups also reported that mentors' explicit none-valuative role created a relationship of trust with mentees that distinguished the support from other coaching.
- All mentees participating in end-of-study focus groups indicated that they valued participating in the program.
- Most mentees reported in surveys that they did not want to decrease the number of mentoring visits they received. In fact, the percentage of mentees who wanted the frequency of retired mentor visits to either increase or stay the same was 81 percent after the first year (69 percent wanted the frequency of visits to stay the same, and 12 percent wanted it to increase) and 89 percent after the second year (71 percent wanted visit frequency to stay the same, and 18 percent wanted it to increase).
- Mentees in focus groups highly valued having a consistent weekly mentor presence in their classroom.
- Mentees in focus groups highly valued immediate (within 24 hours) feedback on instruction. This immediate feedback differentiated the program from other coaching.
- The most frequent way retired mentors engaged with mentees was to observe instruction and provide feedback. Mentees highly valued this support, and support grew from the first year to the second year (figure 3).

The only component of the Retired Mentors for New Teachers program that was not implemented with fidelity was mantees' access to summer professional development with their mentor for two half-days prior to the first year

Figure 3. The value mentees placed on feedback based on observation by mentors was high and grew from the first year to the second of the Retired Mentors for New Teachers program in Aurora Public Schools

Percent of mentees

First year (2013/14)

Not valuable 3%
Somewhat valuable 9%
Valuable 41%
Very valuable 47%

Second year (2014/15)

Not valuable 0%
Somewhat valuable 11%
Valuable 21%
Very valuable 68%
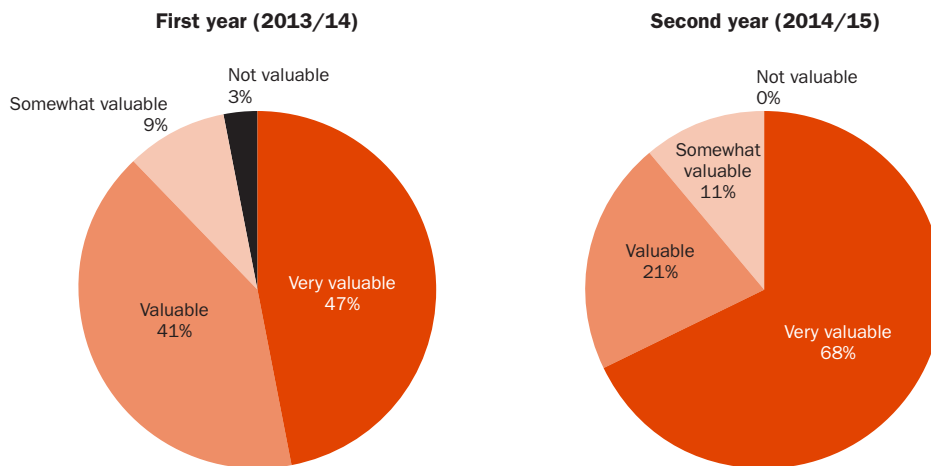
Note: *n* = 32 for 2013/14, and *n* = 28 for 2014/15.

Source: Authors' analysis of data from mentee surveys in 2013/14 and 2014/15.

Implications of the study findings

This section discusses implications of this study's findings on student academic achievement, teacher retention, teacher evaluation, program cost, and return on investment.

Student academic achievement

The study found that the Retired Mentors for New Teachers program had a significant positive effect on math achievement among elementary school students after one year and a nonsignificant positive effect on math and reading achievement after the second year. The differences in math and reading achievement after the first year (1.4 points) and the difference in math achievement after the second year (1.4 points) between students taught by teachers in the program group and students taught by teachers in the business-as-usual group were the same, suggesting that the Retired Mentors for New Teachers program improved student math achievement in the first year and continued to do so in the second year. The lack of statistical significance in the second year may be due to teacher and student attrition (see figures A1 and A2 in appendix A) rather than a lack of effect on achievement; however, the lack of statistical significance indicates that the findings are suggestive only.

Even with these higher achievement outcomes, students taught by teachers in the program group lag behind national grade-level norms (Northwest Evaluation Association, 2015) by one full standard deviation. This suggests that, despite the significant positive effect on student math achievement after one year, the Retired Mentors for New Teachers program alone may not be sufficient to close the achievement gap for students in Title I Aurora Public Schools. The program may help students move toward meeting national achievement norms and could be combined with other research-proven approaches to further improve student achievement.

Despite the significant positive effect on student math achievement after one year, the Retired Mentors for New Teachers program alone may not be sufficient to close the achievement gap for students in Title I Aurora Public Schools

The student achievement effects of the Retired Mentors for New Teachers program are larger than the two-year impact estimates from a recent randomized controlled trial of a comprehensive two-year program for new teachers that provided them with a trained full-time mentor (Glazerman et al., 2010). That study found no statistically significant effect on student achievement during the first two years, but it did find a statistically significant effect of the two-year comprehensive induction and mentoring support when a subset of student achievement data—including students for whom current and prior year data were available—was analyzed after a third year. The effect size after three years in that study was larger (0.20 in math and 0.11 in reading) than the effect size after two years in the current study.

It was beyond the scope of the current two-year study to analyze student impacts associated with the Retired Mentors for New Teachers program for a third year (a year after the end of the two years of added mentoring support). Studying the student achievement impacts of the program after three or more years could be particularly informative since the program had a statistically significant effect on student achievement (in math) after one year. And such analysis may show whether the impacts on teacher instruction and student achievement change after the intervention has been fully delivered.

Studying the student achievement impacts of the program after three or more years may show whether the impacts on teacher instruction and student achievement change after the intervention has been fully delivered

Teacher evaluation

The current study found no observable effect on teacher evaluation after two years of comprehensive orientation and mentoring and other support for new teachers. In the Aurora Public Schools' evaluation system, as in many other districts', school leaders such as the principal or assistant principal are heavily involved in observing and rating teachers. The current study was carried out across 11 elementary schools over two school years, so at least 11 evaluators rated the study's teachers. Future research could more deeply explore how evaluator characteristics affect the consistency of assigning evaluation ratings and how this might affect the ability of researchers to detect observable effects of mentoring interventions.

It was beyond the scope of the current study to gather data on the characteristics of each evaluator in each school. Such data might include years of experience evaluating teachers, specific training received in how to observe teacher performance for evaluation purposes, number of classroom evaluation visits per teacher, and duration of classroom evaluation visits. Those data could help researchers understand whether teacher evaluation ratings are related to evaluator characteristics and whether variation in evaluator experience reduces the consistency in evaluation ratings across school sites.

It was also outside the scope of the current study to collect information on the training each evaluator received in using the Aurora Public Schools evaluation rubric, which was fully implemented across all schools for the first time during the 2014/15 school year. Such information could shed light on how evaluators interpreted the rubric's requirements and how varied interpretations affect the consistency of evaluation across sites.

Teacher retention

The current study corroborates the findings of some prior research that found no statistically significant effect on teacher retention after two years of comprehensive orientation,

support, and mentoring for new teachers (Glazerman et al., 2010). However, the current study defined retention as remaining within the same school district as opposed to remaining in the teaching profession even if working in a different district. This distinction is important because recent longitudinal research tracking thousands of teachers over five years found that the percentage of beginning teachers who remained in the profession was larger among those who were assigned a first-year mentor than among those who were not (Gray et al., 2015).

The current study suggests that increased teacher retention after the second year was associated with program teachers receiving more hours of mentoring (particularly for those who received 40 or more hours of mentoring over the course of the two years). While not causal, the relationship between hours of mentoring and increased retention suggests that more investigation is needed on mentoring dosage (hours). For instance, teachers who are more motivated to stay in the district could have sought additional hours of mentoring on their own. Such investigation could further explore how increasing the amount of mentoring received or ensuring that all teachers receive a minimum number of mentoring hours might increase retention.

While not causal, the relationship between hours of mentoring and increased retention suggests that more investigation is needed on mentoring dosage

Further research could explore the relationship between mentee buy-in to the program and the amount of mentoring each mentee received. First, retired mentors consistently reported in focus groups for the study that while most mentees bought into the program, there was variation within and across schools and that mentees with stronger buy-in were more likely to seek out their mentor and to use more hours of support in the process. Conversely, mentees with less buy-in sought to minimize time with their mentor. Future research could explore development of a methodology for assessing differing levels of teacher buy-in to receiving mentoring support and the reasons for differences in levels of buy-in. Coupling data from such an assessment of buy-in with the number of hours of mentoring received could improve understanding of the overall impacts of mentee buy-in on mentoring dosage.

Second, principals' support for the program could affect mentee buy-in (and thus could affect how likely mentees were to seek out mentor support) according to focus groups. In particular, principals in some schools were more explicit in their directions to mentee teachers about expectations for meeting weekly with their mentor. Principals in some schools also included mentors in staff discussions and leadership team meetings, communicating clearly that the mentor was a valued staff member. Additional research could more deeply explore the relationship between principal support and mentee buy-in. Third, focus groups indicated the possibility that mentees who had prior teaching experience outside the district could be less receptive to new teacher mentorship and therefore less likely to seek out retired educator support. Additional research could examine whether overall buy-in, mentoring dosage, and retention differ among teachers with and without prior teaching experience.

More research is needed on why teachers leave the profession or the district where they work. In general, districts do not collect detailed or consistent data from teachers that leave their employment in order to understand whether the motives were professional, personal, or related to performance. Such data, if collected in a systematic, detailed, and reliable way, could inform future policy and practice about teacher retention and help in tailoring mentor support to meet the specific needs of mentees.

Program cost

The average annual cost per student for the Retired Mentors for New Teachers program to Aurora Public Schools was estimated to be $171 in local resource costs. This estimate includes costs that do not require new expenditures from the district. For instance, the cost analysis includes costs of district staff who were already salaried employees and who would therefore be paid regardless of whether the program was operating (such as the district's professional development director, district trainers, and grant office personnel). These staff did not present a new cost to the district to operate the program, but their costs were still included to present the opportunity cost to the district of their time spent on the program.

Cost per student could be lower in the typical operation of the program without the presence of a randomized controlled trial. Retired educators indicated in focus group discussions that they could serve additional mentees in each school with a smaller marginal added cost. They were unable to do so during the study, however, because half of the available probationary teachers in each school were assigned to the study's business-as-usual group.

While the mentor team in the current study was relatively small, the number could grow if the program expands into middle or high schools or if other districts implement a similar program. The impacts of such expansion on overall cost are unclear

The national average annual cost per student is estimated at $237. The largest difference between the national cost and the local cost was in the national rate for mentor pay, which was derived from average national assistant principal salary data and found to be about $10 per hour higher than the $41 per hour that Aurora Public Schools mentors were paid. Assistant principal salary data were used because that job role was the closest position in the database reflective of the skill set of the retired mentor team, even though only two of the eight mentors had career experience as school administrators.

One area of future study could be the impact of a retired mentor program on teacher decisions to retire. While the mentor team in the current study was relatively small (eight mentors), the number could grow if the program expands into middle or high schools or if other districts implement a similar program. The impacts of such expansion on overall cost are unclear. For instance, would veteran teacher decisions to retire in order to join the mentoring program allow districts to hire newer, less expensive teachers to take their place, and how would this affect district costs both for teachers and mentors?

Return on investment

The current study found that the 704 students taught math by teachers in the program group in the first year could earn nearly $2 million more combined than the students taught by teachers in the business-as-usual group. This figure, which is based on the increase in math assessment scores among students taught by teachers in the program group, takes into account a discount rate. It also draws on existing research that takes into account fadeout of the program's effect over time (Chetty et al., 2011). The estimated lifetime earnings increase could therefore pay back the annual cost of the program more than 15 times over.

In addition to increased lifetime income, other long-term benefits to society are related to increases in student academic achievement (Chetty et al., 2011). In fact, an increase in assessment scores equivalent to that experienced by students taught by teachers in the program group in the first year of the current study is associated with increased probability of college attendance and reduced probability of teenage births.

Research has not quantified the value of these additional returns. The return on investment analysis for this study includes only the returns demonstrated by potential increased annual earnings of students taught by teachers in the program group. Additional research quantifying dollar values associated with increased probability of attending college at age 20 and reduced probability of teenage births could change the estimated long-term return on investment estimates for the program.

Limitations of the study

This study has four main limitations. The first three limitations concern the generalizability of the study findings, and the other concerns the study's analytic approach.

First, the study used a sample of convenience. All study participants were from Aurora Public Schools and agreed to participate in the study, which limits the generalizability of the study findings to this voluntary sample. The findings may be generalizable to districts with similar characteristics as Aurora Public Schools, especially in terms of teachers and students.

Second, the study used a sample of teachers with varying degrees of experience in the teaching profession and in the district. Teachers at the start of the study ranged from 0 to 3 years of experience in the district and from 0 to 15 years of prior experience in the teaching profession. The findings are not generalizable to different samples of teachers, such as a sample composed entirely of teachers in their first year of the profession.

Third, the findings are not generalizable to other mentoring programs or interventions or to the practice of mentoring. Nor can they be generalized to implementing the Retired Mentors for New Teachers Program under different conditions, such as if the summer professional development had occurred or if the program had started on time each year. In other words, the findings provide evidence of the impact of the Retired Mentors for New Teachers Program only under the implementation conditions observed in this study.

Fourth, results from the sensitivity analyses suggest that the impact of the Retired Mentors for New Teachers program was influenced to some degree by the specification of the statistical models used to estimate impacts. Specifically, impact estimation was affected by the inclusion of moderator effects of the grade level that teachers were teaching and by the inclusion of covariates that described teachers' years of experience teaching and years of experience in the district.

The findings of the study are not generalizable to different samples of teachers, such as a sample composed entirely of teachers in their first year of the profession, or to other mentoring programs or interventions or to the practice of mentoring

This appendix provides details on the study data, outcome measures, baseline equivalence, and methodology.

Data

The study's randomized controlled trial was conducted at 11 elementary school sites in Aurora Public Schools, which is located in the Denver, Colorado, metropolitan area. The district's enrollment in 2014 was 41,729, making it the fifth-largest district in the state (table A1). At least 80 percent of students in each of the study schools were eligible for the federal school lunch program.

In the first year of the study, all participating teachers taught core classroom subjects (math or reading) in grades 1–5 and were in their first three years of teaching at Aurora Public Schools. The number of teachers from each school who participated in the study ranged from 2 to 13, and teachers had an average of about 1.4 years of experience in the district.

Because some teachers taught only math or only reading, it was possible for students to be assigned a teacher in the program group for one subject and a teacher in the business-as-usual group for the other. The study addressed this issue by analyzing math and reading impacts separately.

Teachers were randomly assigned to the treatment (program group) and control (business-as-usual group) conditions within schools. Within each school, teachers were assigned an initial random number via a random-number generator and then sorted by grade level and random number. The first teacher in the list was then assigned another random number from 0 to 1.0000; if that number was between 0 and 0.5000, the teacher was assigned to the program group, and if that number was between 0.5001 and 1.000, the teacher was assigned to the business-as-usual group. Assignment of the remaining teachers in each school alternated down the list, based on the assignment of the first teacher. For example, if the first teacher was randomly assigned to the business-as-usual group, the second teacher was assigned to the program group, the third teacher to the business-as-usual group, and so on.

Table A1. Aurora Public Schools enrollment characteristics, 2014/15

Student characteristic	Percent of students
Eligible for the federal school lunch program	70
Non-White	82
Hispanic	54
Black (non-Hispanic)	18
Asian	5
Multiracial	4
White (non-Hispanic) students	18

Note: Total district enrollment is 41,729. Components may not sum to total because of rounding.

Source: Authors' analysis of 2014/15 data from the Colorado Department of Education.

Assignment was carried out as soon as a complete participant list (schools and probationary teachers within each school) was available but after students had been assigned to teachers. Approximately equal numbers of teachers in each grade level were assigned to the two groups (table A2).

Of the 77 teachers in the first year, 39 were randomly assigned to the business-as-usual group, and 38 were randomly assigned to the program group (figure A1). There was no crossover between the two groups. The characteristics of the teachers in the two groups were similar, with no significant differences (table A3). Similarly, the demographic characteristics of students assigned to teachers in the two groups were comparable (see table A3).

In the first year of the study (2013/14), 35 teachers assigned to the business-as-usual group and 35 teachers assigned to the program group taught math (figure A1). Teachers in the program group who taught math had 704 students on their rosters, and teachers in the business-as-usual group who taught math had 678 students on their rosters. Thirty-three teachers assigned to each group taught reading. Teachers in the business-as-usual group who taught reading had 692 students on their rosters, and teachers in the program group who taught reading had 663 students on their rosters.

Students on class rosters were administered the Measures of Academic Progress assessment in both the fall and spring of the 2013/14 school year. Aurora Public Schools had used the Measures of Academic Progress assessment for several years prior to the beginning of the study. Although at the time of the study, giving the Measures of Academic Progress assessment was voluntary for some schools and it was not used in the district's accountability system, all schools participating in the study gave the assessment.

Students were included in the first-year analytic sample only if they had data for both the fall and spring administrations of the Measures of Academic Progress assessment. In math these data were obtained for 575 students taught by teachers in the business-as-usual group and 614 students taught by teachers in the program group. In reading these data were obtained for 561 students taught by teachers in the business-as-usual group and 594 students taught by teachers in the program group.

In the second year of the study (2014/15), 32 teachers assigned to the business-as-usual group and 30 teachers assigned to the program group taught math (figure A2). Teachers in the

Table A2. Number of teachers randomly assigned to program and business-as-usual groups for the Retired Mentors for New Teachers program in Aurora Public Schools, 2013/14 and 2014/15

Grade level	Number of teachers in the business-as-usual group	Number of teachers in the program group	Total number of teachers
1	11	11	22
2	8	6	14
3	7	5	12
4	6	9	15
5	7	7	14
Total	39	38	77

Source: Authors' analysis.

Table A3. Characteristics of students and teachers in the business-as-usual and program groups in Aurora Public Schools, 2013/14 and 2014/15 (percent, unless otherwise noted)

Characteristic	Math		Reading	
	Business-as-usual group	Program group	Business-as-usual group	Program group
First year (2013/14)				
Student characteristics				
Male	52.1	51.2	52.4	50.2
Eligible for the federal school lunch program	88.5	89.6	88.8	90.5
Gifted and talented	1.2	1.2	1.0	2.1
Has an Individualized Education Program	10.0	10.0	10.1	9.0
Hispanic	73.0	73.0	72.8	72.1
Asian	4.9	4.9	5.5	5.0
Black	12.3	12.3	11.7	14.7
White	5.0	5.0	5.0	4.4
Other race/ethnicity	3.3	3.3	3.4	2.5
Teacher characteristics				
Male	11.1	26.5	11.8	12.5
Years in education	2.8	3.2	2.8	3.8
Years in Aurora Public Schools	1.6	1.8	1.8	1.7
Has a master's degree or higher	30.6	41.2	29.4	40.6
Students per teacher	23.4	24.4	22.8	23.9
Second year (2014/15)				
Student characteristics				
Male	52.8	54.8	52.4	50.0
Eligible for the federal school lunch program	92.5	93.1	93.0	98.9
Gifted and talented	1.0	2.6	1.2	2.4
Has an Individualized Education Program	8.4	9.6	8.8	9.6
Hispanic	71.9	71.9	74.2	75.4
Asian	6.8	5.1	6.5	4.8
Black	13.3	12.4	12.5	14.9
White	5.1	6.2	3.3	6.4
Other race/ethnicity	2.4	3.9	2.9	3.5
Teacher characteristics				
Male	10.3	23.7	30.8	52.6
Years in education	3.0	3.7	3.0	3.7
Years in Aurora Public Schools	1.7	1.7	1.7	1.7
Has a master's degree or higher	28.2	39.5	28.2	42.1
Student per teacher	17.5	18.6	17.3	17.0

Note: The business-as-usual group refers to teachers who received mentoring through the district's typical "buddy" mentoring approach. The program group refers to teachers who received mentoring through the Retired Mentors for New Teachers program as well as business-as-usual district mentoring.

Source: Authors' analysis of 2013/14 and 2014/15 data from Aurora Public Schools.

Figure A1. Flow chart showing how participating teachers and students in Aurora Public Schools were assigned in the analytic samples for the first year of the study (2013/14)

PARTICIPANTS DIAGRAM
First year

Eligible teachers

Randomization:
n = 77

Allocated to business-as-usual group:
n = 39

Included: *n* = 39
Crossovers: *n* = 0

Allocated to program group:
n = 38

Included: *n* = 38
Crossovers: *n* = 0

Taught reading:
n = 33
Students on the roster of randomly assigned reading teachers:
n = 692

Taught math:
n = 35
Students on the roster of randomly assigned math teachers:
n = 678

Taught math:
n = 35
Students on the roster of randomly assigned math teachers:
n = 704

Taught reading:
n = 33
Students on the roster of randomly assigned reading teachers:
n = 663

Analytic sample

Students with math data from both the fall and spring administrations of the Measures of Academic Progress assessment:
n = 575

Students with reading data from both the fall and spring administrations of the Measures of Academic Progress assessment:
n = 561

Business-as-usual teachers who left APS:
n = 2

Analytic sample

Students with math data from both the fall and spring administrations of the Measures of Academic Progress assessment:
n = 614

Students with reading data from both the fall and spring administrations of the Measures of Academic Progress assessment:
n = 595

Program teachers who left APS:
n = 1

APS is Aurora Public Schools

Note: Thirty-three teachers in the business-as-usual group and 33 teachers in the program group taught both subjects. The business-as-usual group refers to teachers who received mentoring through the district's typical "buddy" mentoring approach. The program group refers to teachers who received mentoring through the Retired Mentors for New Teachers program in addition to the district's typical "buddy" mentoring approach.

Source: Authors' analysis.

Figure A2. Flow chart showing how participating teachers and students in Aurora Public Schools were assigned in the analytic samples for the second year of the study (2014/15)

PARTICIPANTS DIAGRAM
Second year

Eligible teachers

Randomization:
n = 77

Allocated to business-as-usual group:
n = 39

Allocated to program group:
n = 38

Left APS between first and second year:
n = 2

Left APS between first and second year:
n = 1

Remaining in APS in second year:
n = 37

Remaining in APS in second year:
n = 37

Included: *n* = 37
Crossovers: *n* = 0

Included: *n* = 37
Crossovers: *n* = 0

Taught reading:
n = 29
Students on the roster of randomly assigned reading teachers:
n = 667

Taught math:
n = 32
Students on the roster of randomly assigned math teachers:
n = 781

Taught math:
n = 30
Students on the roster of randomly assigned math teachers:
n = 769

Taught reading:
n = 28
Students on the roster of randomly assigned reading teachers:
n = 637

Analytic sample

Analytic sample

Students with math data from both the fall and spring administrations of the Measures of Academic Progress assessment:
n = 572

Students with math data from both the fall and spring administrations of the Measures of Academic Progress assessment:
n = 533

Students with reading data from both the fall and spring administrations of the Measures of Academic Progress assessment:
n = 482

Students with reading data from both the fall and spring administrations of the Measures of Academic Progress assessment:
n = 482

Business-as-usual teachers who left APS:
n = 5

Program teachers who left APS:
n = 11

APS is Aurora Public Schools

Note: Twenty-nine teachers in the business-as-usual group and 28 teachers in the program group taught both subjects. The business-as-usual group refers to teachers who received mentoring through the district's typical "buddy" mentoring approach. The program group refers to teachers who received mentoring through the Retired Mentors for New Teachers program in addition to the district's typical "buddy" mentoring approach.

Source: Authors' analysis.

business-as-usual group who taught math had 781 students on their rosters, and teachers in the program group who taught math had 769 students on their rosters. Twenty-nine teachers in the business-as-usual group and 28 teachers in the program group taught reading. Teachers in the business-as-usual group who taught reading had 667 students on their rosters, and teachers in the program group who taught reading had 637 students on their rosters. No teachers who taught only math in the first year started teaching reading in the second year. The same was true with reading. Students on class rosters were administered the Measures of Academic Progress assessment in both the fall and spring of the 2014/15 school year.

Students were included in the second-year analytic sample only if they had data for both the fall and spring administrations of the Measures of Academic Progress assessment. In math, these data were obtained for 572 students taught by teachers in the business-as-usual group and 533 students taught by teachers in the program group. In reading, these data were obtained for 482 students taught by teachers in the business-as-usual group and 482 students taught by teachers in the program group.

Teacher attrition was measured after the first year and after the second year. At the beginning of the second year, Aurora Public Schools provided the study team with a list of study teachers staying to teach for the second year. If a study teacher who had taught during the first year did not stay to teach in the district for the second year, that teacher was considered to have left the study (this included any teachers who left during the first year). Both

Table A4. Teacher and student attrition in Aurora Public Schools during the study period (2013/14 and 2014/15)

Sample	Number with students in analysis file	Number randomly assigned	Attrition rate (percent)
Teachers			
Reading			
Total	66	66	0.0
Business-as-usual	33	33	0.0
Program	33	33	0.0
Differential	na	na	0.0
Math			
Total	70	71	1.4
Business-as-usual	35	36	2.8
Program	35	35	0.0
Differential	na	na	2.8
Students			
Reading			
Total	1,155	1,355	14.8
Business-as-usual	561	663	15.4
Program	594	692	14.2
Differential	na	na	1.2
Math			
Total	1,189	1,382	14.0
Business-as-usual	575	678	15.2
Program	614	704	12.8
Differential	na	na	2.4

na is not applicable.

Source: Authors' analysis based on 2013/14 and 2014/15 data from Aurora Public Schools.

overall and differential attrition was low for both students and teachers over the course of the study (table A4).

Teacher attrition during the second year was measured when the year was complete and determined by which teachers stayed in the district to teach for what would be the third year. After the second year 5 more teachers in the business-as-usual group and 11 more teachers in the program group left.

Outcome measures

The study examined the effect of the Retired Mentors for New Teachers program on four outcomes: student achievement in math, student achievement in reading, teacher retention, and teacher evaluation scores.

Student achievement was measured using the Measures of Academic Progress assessment, a computer-adaptive achievement test in math and reading. The typical assessment is 42–50 questions long and completed by most students in about an hour. It is designed to be used as an interim assessment given multiple times a year and is based on grade level. The assessment is norm-referenced, with scores reported on a Rausch unit scale, an equal-interval scale. For both math and reading, student's Rausch unit score was used.

Measures of Academic Progress technical documentation indicates that the score is precise enough to use the assessment as a basis for decisions concerning individual students (Northwest Evaluation Association, 2004). The technical documentation also demonstrates concurrent validity with high correlations with a number of other state assessments, including the Arizona Instrument to Measure Standards and the Stanford Achievement Test 9 (Northwest Evaluation Association, 2004). The test-retest reliability of the instrument was above 0.8, despite several months separating administrations.

Teacher retention was measured simply by whether an individual teacher stayed in Aurora Public Schools to teach the following school year. A teacher who remained in the district but changed grade levels or schools was considered to be retained. Teacher retention was measured at the beginning of the following school year for each year of the study.

Teacher evaluation data were gathered by the district over the course of the school year, in accordance with a districtwide rubric. Teachers received an overall rating and ratings for five components:
- Demonstrating mastery of and pedagogical expertise in the content taught.
- Establishing a safe, inclusive, and respectful learning environment for a diverse population of students.
- Planning and delivering effective instruction and creating an environment that facilitates learning for students.
- Reflecting on practice.
- Demonstrating leadership.

Each category of evaluation includes several elements, including an observation guide to assist the evaluator in determining the appropriate rating for the teacher. In determining the appropriate rating, evaluators also review a range of artifacts, including student achievement data, parent feedback, lesson plans, and student work. Data on evaluation

ratings for teachers in the business-as-usual and program groups were provided to the study team by the district's human resource office.

Baseline equivalence

To examine how teacher attrition might have affected the baseline equivalence of students assigned to teachers in the business-as-usual and program groups in the second year of the study, the study team conducted a baseline equivalence test using fall student math and reading Measures of Academic Progress assessment scores. In both subjects the effect size of the difference in mean scores between students taught by teachers in the two groups was between 0.06 and 0.11 standard deviation (table A5). Because of this observed difference in the baseline characteristic of the two groups, the analysis includes a statistical adjustment by including the baseline characteristic as a student-level covariate.

Methodology

To address the research questions on student academic achievement, a multilevel linear model was used to estimate program effects. A baseline measure (fall Measures of Academic Progress scores) was included as a covariate to increase the precision of the estimate. The model used to estimate the difference at baseline took the following form:

Level 1

$$Y_{ij} = \beta_{0j} + e_{ij}$$

Level 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{TREAT})_j + \Sigma\gamma_k(\text{SCHOOL})_{kj} + u_{0j}$$

Table A5. Differences in fall Measures of Academic Progress assessment scores between students taught by teachers in the business-as-usual group and students taught by teachers in the program group in Aurora Public Schools, by year

Year and subject	Business-as-usual group[a]	Program group[b]	Difference between groups	Pooled standard deviation	Effect size[c]
First year (2013/14)					
Math	153.6	156.1	2.5	22.1	0.11
Reading	150.2	152.6	2.4	23.0	0.11
Second year (2014/15)					
Math	146.6	147.8	1.3	22.7	0.06
Reading	149.3	151.3	2.0	21.2	0.09

Note: This model calculated the adjusted mean scores for each group using a hierarchical linear model nesting students within teachers. The model contained school dummy variables at the school level to account for the blocking of teachers within school for random assignment. This model excluded other teacher or student covariates at the student level to estimate the uncontrolled student mean nested within schools.

a. Refers to teachers who received mentoring through the district's typical "buddy" mentoring approach.

b. Refers to teachers who received mentoring through the Retired Mentors for New Teachers program in addition to the district's typical "buddy" mentoring approach.

c. Calculated using the Hedge's g calculation.

Source: Authors' analysis of 2013/14 and 2014/15 data from Aurora Public Schools.

where Y_{ij} is the score at baseline for student I in classroom j and TREAT is an indicator variable coded as 1 for teachers randomly assigned to the program group and 0 for teachers randomly assigned to the business-as-usual group, γ_k is a vector of fixed effects for $K - 1$ school dummy variables (grand mean centered) to account for the blocking of teachers within school for random assignment, u_{0j} is the random error in the achievement outcome associated with teacher j, The value of γ_{00} is the estimated mean of the business-as-usual group, the value of γ_{01} is the estimated difference between the business-as-usual group and the program group, and the statistical significance of γ_{01} is the test of group difference.

Level 1 of the impact model is specified as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(PRE_{ij} - \overline{PRE} ._{j})_{ij} + \beta_{2j}(GENDER)_{ij} + \beta_{3j}(FRL\_pre)_{ij} + \beta_{4j}(BLACK)_{ij} +$$
$$\beta_{5j}(WHITE)_{ij} + \beta_{6j}(OTHER)_{ij} + \beta_{7j}(NEP)_{ij} + \beta_{8j}(LEP)_{ij} + \beta_{9j}(Other\ ELL)_{ij} + \beta_{10j}(SPED)_{ij}$$

where Y_{ij} is the student outcome for student i in classroom j, β_{0j} is the regression-adjusted mean outcome for students in classroom j, and β_{1j} is the regression-adjusted difference in the outcome due to the student's score on the fall 2013 administration of the Measures of Academic Progress assessment, which was classroom mean centered so that each classroom intercept would be at the mean level of baseline achievement within each classroom. β_{2j} is the regression-adjusted differences in the outcome for student gender, and GENDER is coded 0 for male and 1 for female. β_{3j} is the regression-adjusted difference in the outcome for student eligibility for the federal school lunch program, and FRL_pre is coded 1 for eligible and 0 for not eligible. β_{4j} through β_{6j} are the regression-adjusted differences in the outcome due to the student's race/ethnicity, where each variable is coded as 1 for the respective racial/ethnic group and 0 otherwise. Hispanic was the excluded reference group, and the included race/ethnicity variables were Black, White, and other, which included Asian, Pacific Islander, and multiracial. β_{7j} through β_{9j} are dummy variables for the various categories of English learner students. β_{10j} is a dummy variable for participation in special education. e_{ij} is the random error in the achievement outcome associated with student i in classroom j.

Level 2 of the model is specified as follows:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(TREAT)_j + \gamma_{02}(\overline{PRE} ._j - \overline{PRE} ..)_j + \gamma_{03}(GENDER)_j +$$
$$\gamma_{04}(MASTERS\_pre)_j + \gamma_{05}(YRS\_TEACH\_pre)_j + \gamma_{06}(YRS\_DISTRICT\_pre)_j +$$
$$\gamma_{07}(CLASS\_SIZE)_j + \Sigma\gamma_q(GRADE\_pre)_{qj} + \Sigma\gamma_k(SCHOOL)_{kj} + u_{0j}$$

where γ_{00} is the regression adjusted business-as-usual group grand mean on the outcome and γ_{01} is the regression adjusted mean difference in the outcome between teachers assigned to the program group and teachers assigned to the business-as-usual group. γ_{02} is the regression slope of the classroom-level score on the fall 2013 administration of the Measures of Academic Progress assessment (grand mean centered) to explain additional between-teacher variance not explained in level 1 of the model. γ_{03} through γ_{07} are the regression adjusted differences in the outcome due to teacher characteristics, and GENDER is coded 0 for male and 1 for female, MASTERS_pre is coded 1 for obtained a master's degree and 0 otherwise, YRS_TEACH_pre is the number of years of teaching experience at baseline, and YRS_DISTRICT_pre is the number of years teaching in APS at baseline. CLASS_SIZE is the number of students in each teacher's classroom and is grand mean centered. γ_q is a vector of grand mean centered $Q - 1$ dummy variables to

account for early career teacher grade level at baseline that will adjust the model for different numbers of teachers in each grade level in the program and business-as-usual groups and also will place the classroom level intercept at the grade-level regression adjusted mean on the MAP test score vertical scale (Enders & Tofighi, 2007). γ_k is a vector of fixed effects for $K - 1$ school dummy variables (grand mean centered) to account for the blocking of teachers within school for random assignment. u_{0j} is the random error in the achievement outcome associated with teacher j.

The effects of the individual-level covariates, β_{1j} though β_{7j}, are fixed at the school level (level 2).

The primary parameter of interest from the above model is γ_{01}, which can be interpreted as the regression-adjusted classroom mean difference between the business-as-usual and program groups. A two-tailed test ($p < .05$) was used to assess the statistical significance of this impact estimate.

To address the research question on teacher evaluation outcomes, a logistic regression model was used to estimate program effects. The outcome variable was binary: 1 if the teacher stayed in Aurora Public Schools to teach the following year and 0 if the teacher did not stay. The odds of staying were predicted based on whether the teacher was in the program group or the business-as-usual group, years in education, years in Aurora Public Schools, gender, and whether they had a master's degree or higher. Whether the teacher was in the program group or the business-as-usual group was the variable of interest in the model. A two-tailed test ($p < 0.05$) was used to assess the statistical significance of this impact estimate.

To address the research question on teacher evaluation outcomes, teacher overall and category ratings (basic, partially proficient, proficient, and accomplished) were converted to numeric ratings, and an ordered logit model used to estimate the impact of the Retired Mentors for New Teachers program. This model was used instead of a linear regression because the categorical ratings are ordinal but do not necessarily have the interval property. The model was estimated separately for the overall evaluation rating and each category rating. The ordered logit model used teacher evaluation rating as the outcome variable and the teacher's treatment status, gender, years in education, years in Aurora Public Schools, and highest level of education as predictor variables. In this model, whether the teacher was in the program group or the business-as-usual group was the variable of interest. A two-tailed test ($p < .05$) was used to assess the statistical significance of this impact estimate.

Appendix B. Sensitivity analysis

In addressing the first and second research questions, the study team conducted several sensitivity analyses to study the robustness of the impact estimate of the Retired Mentors for New Teachers program. Findings of the sensitivity analyses indicate that the impact estimate was somewhat affected by the choice of model (tables B1–B4). Results were also sensitive to application of the Benjamini–Hochberg correction for multiple comparisons. This correction was not applied to the results reported in the body of this report because of the relatively low number of statistical comparisons performed.

The estimated impact on math achievement was sensitive to the exclusion of interaction terms exploring the moderation of treatment effect by teacher grade level (see table B1). When those terms were included, there was no significant difference in math achievement between students in the program and business-as-usual groups.

The estimated impact on reading achievement was sensitive to the inclusion of variables for teacher demographics, including years of experience and years teaching in the district (see table B2). When those variables were excluded, there was a significant difference in reading achievement between students in the program and business-as-usual groups. This is likely caused by the substantial variation in teaching experience among teachers participating in the study.

Table B1. Impact estimate of the Retired Mentors for New Teachers program on students' Measures of Academic Progress math scale score after the first year (2013/14), by model specification

Model	Adjusted mean		Standard deviation		Difference between adjusted means	Standard error	*p*-value
	Business-as-usual group	Program group	Business-as-usual group	Program group			
Benchmark	178.9	180.3	2.67	0.70	1.4	0.70	.045
Sensitivity analysis 1	160.0	164.0	1.35	1.29	4.0	1.37	.002
Sensitivity analysis 2	164.6	168.1	2.31	1.06	3.5	1.06	.001
Sensitivity analysis 4	183.1	181.1	2.75	1.58	-2.0	1.58	.208

Note: Sensitivity analyses were conducted to test the robustness of the benchmark impact estimates to the inclusion and exclusion of covariates. The benchmark model is the model reported in the main text. Sensitivity analysis 1 excludes the variables for individual-level scores on the fall 2013 administration of the Measures of Academic Progress assessment, student gender and race/ethnicity indicators, classroom-level scores on the fall 2013 administration of the Measures of Academic Progress assessment, and teacher characteristics. Sensitivity analysis 2 excludes the variables for teacher transfer (indicating whether teachers changed subject or grade taught between the first and second year of the study) and for first-year teacher characteristics (gender, level of education, years of teaching experience, and years of experience in Aurora Public Schools). Sensitivity analysis 4 tests for possible moderation of the treatment effect by teacher grade level taught by including interaction terms in the model. The table does not include sensitivity analyses 3 and 5 because they were not applicable to this benchmark model.

Source: Authors' analysis of 2013/14 data from Aurora Public Schools; see appendix A.

Table B2. Impact estimate of the Retired Mentors for New Teachers program on students' Measures of Academic Progress reading scale score after the first year (2013/14), by model specification

Model	Adjusted mean		Standard deviation		Difference between adjusted means	Standard error	*p*-value
	Business-as-usual group	Program group	Business-as-usual group	Program group			
Benchmark	185.3	185.6	2.87	0.70	0.3	0.70	.470
Sensitivity analysis 1	158.7	160.9	1.45	1.58	2.2	1.33	.114
Sensitivity analysis 2	196.8	199.4	2.71	1.11	2.6	1.11	.024
Sensitivity analysis 4	173.8	175.4	3.29	1.57	1.6	1.57	.314

Note: Data were regression adjusted to test the robustness of the benchmark impact estimates to the inclusion of covariates. The benchmark model is the model reported in the main text. Sensitivity analysis 1 excludes the variables for individual-level scores on the fall 2013 administration of the Measures of Academic Progress assessment, student gender and race/ethnicity indicators, classroom-level scores on the fall 2013 administration of the Measures of Academic Progress assessment, and teacher characteristics. Sensitivity analysis 2 excludes the variables for teacher transfer (indicating whether teachers changed subject or grade taught between the first and second year of the study). Sensitivity analysis 2 excludes the variables for first-year teacher characteristics (gender, level of education, years of teaching experience, and years of experience in Aurora Public Schools). Sensitivity analysis 4 tests for possible moderation of the treatment effect by teacher grade level by including interaction terms in the model. The table does not include sensitivity analyses 3 and 5 because they were not applicable to this benchmark model.

Source: Authors' analysis of 2013/14 data from Aurora Public Schools; see appendix A.

Table B3. Impact estimate of the Retired Mentors for New Teachers program on students' Measures of Academic Progress math scale score after the second year (2014/15), by model specification

Model	Adjusted mean		Standard deviation		Difference between adjusted means	Standard error	*p*-value
	Business-as-usual group	Program group	Business-as-usual group	Program group			
Benchmark	178.3	179.7	2.97	0.86	1.4	0.86	.113
Sensitivity analysis 1	157.3	160.2	1.67	1.63	2.9	2.40	.075
Sensitivity analysis 3	179.3	180.5	2.84	0.84	1.2	0.84	.169
Sensitivity analysis 4	177.0	179.6	2.38	1.92	2.6	1.92	.177
Sensitivity analysis 5	178.3	179.6	2.97	0.86	1.4	0.86	.117

Note: Data were regression adjusted to test the robustness of the benchmark impact estimates to the inclusion of covariates. The benchmark model is the model reported in the main text. Sensitivity analysis 1 excludes the variables for individual-level scores on the fall 2014 administration of the Measures of Academic Progress assessment, student gender and race/ethnicity indicators, classroom-level scores on the fall 2014 administration of the Measures of Academic Progress assessment, and teacher characteristics. Sensitivity analysis 3 excludes the variables for teacher transfer (indicating whether teachers changed subject or grade taught between the first and second year of the study). Sensitivity analysis 4 tests for possible moderation of the treatment effect by teacher grade level by including interaction terms in the model. Sensitivity analysis 5 tests for potential moderation of the second-year treatment effect by student exposure levels to treatment in the first year. The table does not include sensitivity analysis 2 because it was not applicable to this benchmark model.

Source: Authors' analysis of 2014/15 data from Aurora Public Schools; see appendix A.

Table B4. Impact estimate of the Retired Mentors for New Teachers program on students' Measures of Academic Progress reading scale score after the second year (2014/15), by model specification

Model	Adjusted mean		Standard deviation		Difference between adjusted means	Standard error	*p*-value
	Business-as-usual group	Program group	Business-as-usual group	Program group			
Benchmark	172.6	174.0	2.55	0.81	1.4	0.81	.078
Sensitivity analysis 1	155.5	159.0	1.91	1.85	3.5	1.99	.058
Sensitivity analysis 3	175.2	176.0	2.09	0.78	0.9	0.78	.277
Sensitivity analysis 4	170.6	172.9	2.04	1.69	2.3	1.69	.177
Sensitivity analysis 5	172.8	174.1	2.54	0.80	1.3	0.80	.098

Note: Data were regression adjusted to test the robustness of the benchmark impact estimates to the inclusion of covariates. The benchmark model is the model reported in the main text. Sensitivity analysis 1 excludes the variables for individual-level scores on the fall 2014 administration of the Measures of Academic Progress assessment, student gender and race/ethnicity indicators, classroom-level scores on the fall 2014 administration of the Measures of Academic Progress assessment, and teacher characteristics. Sensitivity analysis 3 excludes the variables for teacher transfer (indicating whether teachers changed subject or grade taught between the first and second year of the study). Sensitivity analysis 4 tests for possible moderation of the treatment effect by teacher grade level by including interaction terms in the model. Sensitivity analysis 5 tests for potential moderation of the second-year treatment effect by student exposure levels to treatment in the first year. The table does not include sensitivity analysis 2 because it was not applicable to this benchmark model.

Source: Authors' analysis of 2014/15 data from Aurora Public Schools; see appendix A.

Appendix C. Elements considered in the analysis of program cost

To research program costs, the study team collected a variety of information on the elements used to operate the Retired Mentors for New Teachers program (table C1). To collect necessary local cost information, the study team participated in all mentor team meetings to directly observe the content, length, and personnel involved in supporting such meetings and to receive updates on program implementation. The study team also conducted interviews several times each program year with district grant management staff and collected data on mentor and mentee time spent on the program. And the study team collected qualitative data through focus groups with mentees and mentors each year to foster its understanding of the resource elements involved in conducting the program.

For each personnel position associated with the program, the study team identified a corresponding personnel position in the national database. The study team then used the Center for Benefit-Cost Studies of Education (2015) national database to translate the local cost to a representative national cost based on national average salary and benefit rates for the corresponding personnel.

Not all elements investigated could be included as actual program costs. For example, technology, food and refreshments, and travel costs were not provided by the district for the program. The majority of program costs were associated with personnel, particularly mentor salaries and benefits.

Table C1. Types of cost elements reviewed for the Retired Mentors for New Teachers program, 2013/14 and 2014/15

Cost element	Description	Average	Unit	Average local cost per unit ($)	Average national cost per unit ($)	Annual average local element cost per student ($)	Annual average national element cost per student ($)
Mentor salaries and benefits	Mentor pay based on hourly rates and hours worked, including benefits	2,864	Hours	41.08	55.34	159.66	215.07
Mentee stipends and benefits	Stipends, including benefits, paid to mentees for time spent outside the regular school day working with mentors	66.7	Hours	29.58	50.37	2.68	4.56
Administrator salaries and benefits	Salaries and benefits based on time spent for various school or district staff required to operate the program, such as the district grant office, professional development office, and research office	77.25[a]	Hours	47.63	53.11	4.99	5.57
Clerical and support staff salaries and benefits	Salaries and benefits associated with any clerical and support staff time used to help operate the program	0	Hours	na	na	0	0

(continued)

Table C1. Types of cost elements reviewed for the Retired Mentors for New Teachers program, 2013/14 and 2014/15 *(continued)*

Cost element	Description	Average	Unit	Average local cost per unit ($)	Average national cost per unit ($)	Annual average local element cost per student ($)	Annual average national element cost per student ($)
Substitute teacher salaries and benefits	Cost associated with providing substitute teachers to cover mentee classrooms in order to allow the mentee and their mentor to visit other model teacher classrooms to inform instruction	23.2	Days	90 salary plus 18.4 percent benefits	283 salary plus 24.8 percent benefits	3.36	11.09
Materials and supplies	Cost associated with providing printed or other materials for mentor team meetings or training sessions provided by the district	80	Pages printed	0.04	0.04	0	0
Technology and equipment	Cost associated with any computers or equipment used to support the mentors in their work as a team or with mentees	0	na	na	na	0	0
Food and refreshments	Cost associated with any food or beverages provided at mentor or mentee meetings	0	na	na	na	0	0
Facilities	Cost associated with providing meeting space in a district facility for mentors to meet as a team and to receive training and support from district leaders	14	Hours	15.19	15.19	0.29	0.29
Travel	Any travel cost paid for by the district to support the program	0	na	na	na	0	0

Note: The program incurred no costs for clerical and support staff, technology and equipment, food and refreshments, or travel. Costs for materials and supplies were incurred by the program but round to $0 when shown on a per student basis. Only a national estimate of per unit facilities costs was available. This national facilities cost estimate was therefore also applied to the local facilities cost. The per student cost calculations are based on 737 students, which is the average number of students served in math over the two years of the program—704 in 2013/14 and 769 in 2014/15.

a. Averaged across four administrator levels.

Source: Author's analysis of 2013/14 and 2014/15 data from Aurora Public Schools; see appendix A.

Appendix D. Implementation fidelity

The study team assessed fidelity of implementation of the Retired Mentors for New Teachers program on four components: provision of highly qualified mentors, mantees' access to the mentor prior to the first year, mentor team meetings, and coaching and direct support for mentees.

Online surveys of teachers in the program and business-as-usual groups were conducted at the end of each year, as were focus groups with teachers in the program group and their mentors. Monthly online records of support were completed by mentors and were used to track mentee dosage and key areas of support.

Sixty-four of the study's 77 teachers (83 percent) responded to the survey after the first year of the program, and 55 teachers (74 percent) responded to the survey after the second year. Thirty-five of 38 teachers in the program group (92 percent) participated in focus groups after the first year of the program, and 28 of 37 program teachers (76 percent) participated in focus groups after the second year. Mentor focus groups had 100 percent attendance after both years of the program. The study team also observed all meetings of the retired mentor team.

The study team worked with district leaders prior to the start of the study to create thresholds in each of the four fidelity components for what the district considers adequate implementation. Overall, it was determined that the program was implemented with fidelity, meeting implementation thresholds for three of the four components. Additional details for each component are discussed below.

Provision of highly qualified mentors

The threshold for meeting this component was 100 percent of mentors having both a minimum of five years working in high-need Aurora Public Schools elementary schools and a record of excellence in the district, as shown through references from district leaders or being appointed to serve as a school or district coach or teacher leader. All mentors in the study met this threshold. The mentor team had an average of 22 years of experience working in the district's high-need schools.

Mantees' access to the irmentor prior to the first year

The threshold for meeting this component was for 100 percent of mentees to meet with their mentor for two half-days during the summer to plan for the coming year, to set up classroom space, and to begin building a professional relationship. However, because the district had not finished enrolling schools in the study until late August 2013, randomization of teachers could not occur until after the start of school, and the first-year summer professional development therefore was not implemented. Summer meetings prior to start of the second program year were not a part of the program's design.

Mentor team meetings

The thresholds for meeting this component were 75 percent of mentors attending at least three quarterly meetings of the mentor team each year to share lessons learned and receive

ongoing training and 60 percent of mentors meeting with their school principals at least four times each year to ensure adequate communication. All mentors attended at least three quarterly mentor team meetings during both years of the study. In practice, seven meetings of the mentor team were held each year, with each meeting lasting two to three hours. The percentage of mentors who met at least four times with their principal was 73 percent in the first year and 60 percent in the second year. Both aspects of this component met the threshold for fidelity in both years.

Coaching and direct support

The primary threshold for meeting this component was providing at least 30 hours of mentor support to 60 percent of mentees each year (table D1). In both years this threshold was met. Overall, the average number of hours received across all mentees was 32 in the first year and 40 in the second year. However, there was considerable variation across mentees (from 6.5 hours to 69.5 hours in the first year and from 12.5 hours to 102 hours in the second year).

In addition to the primary threshold, at least one of two secondary thresholds needed to be met: at least 50 percent of mentees indicating that they received mentoring at least monthly over the course of each school year or at least 75 percent of mentees indicating that they met at least monthly with their retired educator mentor and other mentees in their schools. In both years the program far exceeded the threshold for mentees meeting regularly with their mentors but did not meet the threshold for mentees meeting monthly in groups with their retired mentors. However, because the primary threshold and one of the two secondary thresholds were met, this component was considered implemented with fidelity.

Table D1. Summary of program implementation fidelity components for coaching and direct support

Component	Measure	Result for first year (2013/14)	Result for second year (2014/15)
Primary threshold (required)	Provide at least 30 hours of retired mentor support to at least 60 percent of mentees each year	Threshold met. 61 percent of mentees received 30 or more hours	Threshold met. 66 percent of mentees received 30 or more hours
Secondary threshold	At least 50 percent of mentees indicate retired mentoring was delivered at least monthly	Threshold met. 100 percent of mentees indicated meeting with their retired mentor on at least a monthly basis	Threshold met. 96 percent of mentees indicated meeting with their retired mentor on at least a monthly basis
Secondary threshold	At least 75 percent of mentees indicate meeting at least monthly in groups with other mentees and the retired mentor	Threshold not met. 63 percent of mentees indicated meeting monthly in groups with the retired mentor	Threshold not met. 43 percent of mentees indicated meeting monthly in groups with the retired mentor

Note: At least one secondary threshold had to be met in order for the component to be considered implemented with fidelity.

Source: Authors' analysis of 2013/14 and 2014/15 data from Aurora Public Schools; see appendix A.

Table E1 presents regression results for the analysis of the treatment effect on student math achievement after the first year.

Table E1. Regression results for the analysis of the effect of the Retired Mentors for New Teachers program on student math achievement, 2013/14

Variable	Coefficient	Standard error	p > z
Program teacher	1.39	0.70	045
Constant	178.92	2.67	0.000
Student characteristics			
Male	0.26	0.54	633
2013 fall Measures of Academic Progress score (classroom mean centered)	0.78	0.02	0.000
Eligible for the federal school lunch program	–1.36	1.10	216
Black	0.49	1.07	648
White	0.86	1.46	555
Other race/ethnicity (Asian, Pacific Islander, multiracial)	1.29	0.97	182
Not English proficient	–1.34	0.97	166
Limited English proficient	1.38	0.87	114
Other English learner participant	5.17	1.52	001
Special education	–4.75	1.00	0.000
Grade 2	4.18	1.50	005
Grade 3	4.61	2.57	073
Grade 4	3.83	3.52	276
Grade 5	3.98	4.18	341
Teacher characteristics			
Male	–2.58	1.01	010
Has master's degree or higher	0.57	0.85	501
Years in education	0.03	0.16	826
Years at Aurora Public Schools	0.73	0.48	130
2013 fall Measures of Academic Progress score (grand mean centered)	0.80	0.09	0.000
Students per teacher (grand mean centered)	–0.22	0.11	050
Schools			
School 1	1.38	1.81	446
School 2	1.47	1.77	404
School 3	–0.38	1.33	773
School 4	–2.38	1.62	141
School 5	–4.41	1.65	008
School 6	1.60	1.52	292
School 7	2.26	1.85	222
School 8	6.72	2.80	017
School 9	–4.95	1.82	007
School 10	–0.39	1.33	772

Note: The regression included 1,155 students and 69 teachers.

Source: Authors' analysis of 2013/14 data from Aurora Public Schools; see appendix A.

References

Alliance for Excellent Education. (2014). *On the path to equity: Improving the effectiveness of beginning teachers.* Washington, DC: Author. Retrieved November 17, 2015, from http://all4ed.org/reports-factsheets/path-to-equity/.

Barnes, G., Crowe, E., & Schaefer, B. (2007). *The cost of teacher turnover in five school districts.* Arlington, VA: National Commission on Teaching and America's Future. http://eric.ed.gov/?id=ED497176

Center for Benefit-Cost Studies of Education. (2015). *CBCSE database of educational resource prices.* New York, NY: Columbia University.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2011). *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood* (Working Paper No. 17699). Cambridge, MA: National Bureau of Economic Research. http://eric.ed.gov/?id=ED528374

Clotfelter, C., Ladd, H., & Vigdor, J. (2006). *Teacher-student matching and the assessment of teacher effectiveness.* Cambridge, MA: National Bureau of Economic Research. http://eric.ed.gov/?id=EJ750956

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychology Methods, 12*(2), 121–138. http://eric.ed.gov/?id=EJ766180

Foster, E. (2010). *How Boomers can contribute to student success: Emerging encore career opportunities in K–12 education.* Arlington, VA: National Commission on Teaching and America's Future. http://eric.ed.gov/?id=ED512649

Glazerman, S., Isenberg, E., Dolfin, S., Bleeker, M., Johnson, A., Grider, M., et al. (2010). *Impacts of comprehensive teacher induction: Final results from a randomized controlled study.* Princeton, NJ: Mathematica Policy Research.

Gray, L., Taie, S., & O'Rear, I. (2015). *Public school teacher attrition and mobility in the first five years: Results from the first through fifth waves of the 2007–08 Beginning Teacher Longitudinal Study.* Washington, DC: National Center for Education Statistics, U.S. Department of Education. http://eric.ed.gov/?id=ED544179

Ingersoll, R., & Strong, M. (2011). The impact of induction and mentoring programs for beginning teachers: A critical review of the research. *Review of Educational Research, 81*(2), 201–233. http://eric.ed.gov/?id=EJ923890

Levin, H., & Belfield, C. (2013). *Guiding the development and use of cost-effectiveness analysis in education.* New York, NY: Center for Benefit-Cost Studies of Education, Columbia University. http://eric.ed.gov/?id=EJ1068556

Northwest Evaluation Association. (2004). *Reliability and validity estimates: NWEA Achievement Level Tests and Measures of Academic Progress.* Portland, OR: Author. Retrieved November 23, 2015, from http://images.pcmac.org/Uploads/Jacksonville117/

Jacksonville117/Sites/DocumentsCategories/Documents/Reliability_and_Validity_
Estimates.pdf.

Northwest Evaluation Association. (2011). *NWEA Measures of Academic Progress norma-
tive data.* Portland, OR: Author. Retrieved November 18, 2015, from https://www.nwea.
org/content/uploads/2014/07/MAP-Normative-Data-One-Sheet-Dec11.pdf.

Northwest Evaluation Association. (2015). *NWEA Measures of Academic Progress norma-
tive data.* Portland, OR: Author. Retrieved November 18, 2015, from https://www.nwea.
org/resources/map-normative-data/#sthash.XXmjQs8T.dpuf.

Organisation for Economic Co-operation and Development. (2005). *Teachers matter:
Attracting, developing and retaining effective teachers.* Paris, France: Author. Retrieved
November 18, 2015, from http://www.oecd.org/edu/school/34990905.pdf.

Rice, J. K. (2010). *The impact of teacher experience: Examining the evidence and policy impli-
cations.* Arlington, VA: National Center for Analysis of Longitudinal Data in Educa-
tion Research. http://eric.ed.gov/?id=ED511988

Smith, T., & Ingersoll, R. (2004). What are the effects of induction and mentoring on
beginning teacher turnover? *American Educational Research Journal, 41*(3), 681–714.
http://eric.ed.gov/?id=EJ727495

Strong, M. (2006). *Does new teacher support affect student achievement?* Santa Cruz, CA:
The New Teacher Center. Retrieved November 18, 2015, from http://www.newteacher
center.org/sites/default/files/ntc/main/resources/BRF_DoesNewTeacherSupportAffect
StudentAchievement.pdf.

The Regional Educational Laboratory Program produces 7 types of reports

Making Connections
Studies of correlational relationships
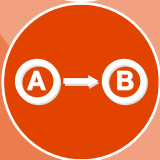
Making an Impact
Studies of cause and effect

What's Happening
Descriptions of policies, programs, implementation status, or data trends

What's Known
Summaries of previous research

Stated Briefly
Summaries of research findings for specific audiences

Applied Research Methods
Research methods for educational settings

Tools
Help for planning, gathering, analyzing, or reporting data or research