

# **Report to the Legislature: Education Evaluation Grant Program**

---

Pursuant to Chapter 46 of the Acts of 2015  
April 2016

**Massachusetts Department of Elementary and Secondary Education**  
75 Pleasant Street, Malden, MA 02148-4906  
Phone 781-338-3000 TTY: N.E.T. Relay 800-439-2370  
[www.doe.mass.edu](http://www.doe.mass.edu)



This document was prepared by the  
Massachusetts Department of Elementary and Secondary Education  
Mitchell D. Chester, Ed.D.  
Commissioner

### **Board of Elementary and Secondary Education Members**

Mr. Paul Sagan, Chair, Cambridge  
Mr. James Morton, Vice Chair, Boston  
Ms. Katherine Craven, Brookline  
Dr. Edward Doherty, Hyde Park  
Dr. Roland Fryer, Concord  
Ms. Margaret McKenna, Boston  
Mr. Michael Moriarty, Holyoke  
Dr. Pendred Noyce, Boston  
Mr. James Peyser, Secretary of Education, Milton  
Ms. Mary Ann Stewart, Lexington  
Mr. Donald Willyard, Chair, Student Advisory Council, Revere

Mitchell D. Chester, Ed.D., Commissioner and Secretary to the Board

The Massachusetts Department of Elementary and Secondary Education, an affirmative action employer, is committed to ensuring that all of its programs and facilities are accessible to all members of the public.

We do not discriminate on the basis of age, color, disability, national origin, race, religion, sex, gender identity, or sexual orientation. Inquiries regarding the Department's compliance with Title IX and other civil rights laws may be directed to the Human Resources Director, 75 Pleasant St., Malden, MA 02148-4906. Phone: 781-338-6105.

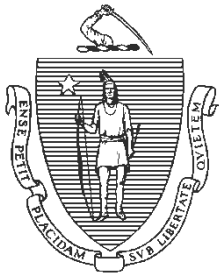
© 2016 Massachusetts Department of Elementary and Secondary Education

*Permission is hereby granted to copy any or all parts of this document for non-commercial educational purposes. Please credit the "Massachusetts Department of Elementary and Secondary Education."*

*This document printed on recycled paper*

Massachusetts Department of Elementary and Secondary Education  
75 Pleasant Street, Malden, MA 02148-4906  
Phone 781-338-3000 TTY: N.E.T. Relay 800-439-2370  
[www.doe.mass.edu](http://www.doe.mass.edu)





# **Massachusetts Department of Elementary & Secondary Education**

---

75 Pleasant Street, Malden, Massachusetts 02148-4906

Telephone: (781) 338-3000  
TTY: N.E.T. Relay 1-800-439-2370

Mitchell D. Chester, Ed.D.  
Commissioner

April 1, 2016

Members of the General Court:

Attached is the Department of Elementary and Secondary Education's first report on the new grant program established in fiscal year 2016 for evaluation of state-funded education programming.

The first two programs to be evaluated through this grant program are the state's academic support program, which provides remediation for students at risk of not meeting the state's high school graduation requirement, and the educator preparation and licensure programs, which work to improve the quality of the state's educator workforce. We have provided research plans from the two research teams selected to conduct these studies in this report and look forward to providing preliminary findings in the near future.

Gathering additional information on these topics will help us improve the quality of the services we deliver to districts as well as to ensure that we are spending the state's fiscal resources wisely. We appreciate your support in helping us to evaluate the implementation, outcomes, and costs and benefits of our state programs.

Sincerely yours,

A handwritten signature in black ink, appearing to read "Mitch D. Chester".

Mitchell D. Chester, Ed.D.  
Commissioner of Elementary and Secondary Education

## Legislative report

The Department of Elementary and Secondary Education respectfully submits this Report to the Legislature: Department of Elementary and Secondary Education Strategic Plan pursuant to Chapter 182 of the Acts of 2008, line item 7010-0050:

*“...provided further, that organizations receiving funds through this item shall report biannually to the house and senate committees on ways and means, the joint committee on education and the joint committee on higher education on: (1) the status and preliminary results of evaluations funded through this item; and (2) any obstacles encountered in access to data or other information that is negatively affecting the completion of the study.”*

The purpose of this funding is to support evaluation of state-funded education programming. The legislature provided \$300,000 for this purpose in fiscal year 2016 and stipulated specific requirements for the analyses to be conducted with the funding.

In the grant competition, the agency prioritized proposals related to the state academic support program for students at risk of not meeting the high school graduation requirement and the state’s educator preparation and licensure programs. ESE has substantial data available to evaluate these programs, they are of high priority to the agency, and the agency has unanswered research questions that an evaluation could shed light on. ESE also allowed proposals to evaluate other state-funded education programs so long as the proposals studied a topic of significant interest to the agency and sufficient data were available to meet the legislative requirements for this grant program.

Per the statutory requirement, proposals specified how the researcher would analyze the following six areas:

1. The quantifiable effect of the program on the population enrolled in the program;
2. Fidelity of program implementation;
3. An estimate of the cost to the Commonwealth of the education problem being addressed through the program;
4. A comparison of the cost of the program and the estimated short-term and long-term benefits received by program recipients through the program;
5. Data limitations in estimating the effect of the program; and
6. Recommendations for further study.

ESE received five applications: three for the educator preparation and licensure priority and two for the academic support priority. No applicants proposed projects outside the two prioritized programs.

After review, the UMass Donahue Institute was awarded a \$150,000 grant to study the state’s academic support program and the American Institutes of Research was awarded a \$150,000 grant to study educator preparation and licensure. Each began their work by preparing a detailed research plan specifying the research questions and timeline for the project. These plans are

attached below. Per statutory requirement, future reports will provide additional information about the status and preliminary findings from the projects, as well as any difficulties the researchers have encountered in securely accessing appropriate data.



## Evaluation of Academic Support Programs

### Evaluation Plan

---

Prepared for the Massachusetts Department of Elementary and Secondary Education by the UMass Donahue Institute

Submitted: January 26, 2016



## Overview

The Massachusetts state legislature funds the Academic Support grants annually to enhance academic support services so that all eligible students will meet the state's Competency Determination—based on performance on the state's high school MCAS assessments in English language arts, mathematics, and science and technology/engineering—that is required for high school graduation.

The Academic Support grants fund four types of school year and summer grant programs: 1) district and approved private special education schools and collaboratives; 2) one-stop career centers and higher education institutions; 3) work and learning programs; and 4) collaborative partnerships for student success. The first of these grant programs is based on a formula that includes the number of high school students who scored at the “Failing” level on their most recently available high school MCAS exam in ELA, mathematics, and STE. The other three grants are awarded through a competitive application process.

Overall, school year 2013–14 participants in programs funded by the Academic Support grants were 1.9 times as likely to have met the state's MCAS testing requirements by November 2014 as eligible youth who did not participate (Massachusetts Department of Elementary and Secondary Education, 2014). During FY14, nearly \$8.9 million in Academic Support grants were awarded. The grants served 10.1 percent of the 153,100 eligible students from the classes of 2003–2018, at an average cost of \$575 per student. Funding cuts resulted in a reduction to \$4.1 million in FY15, and a similar amount is anticipated for FY16. In the past, approximately 450 grants have been awarded per year across the Commonwealth, based on local needs and program scope, with award amounts in fiscal year 2016 projected to range from approximately \$5,000 to \$350,000.

The Applied Research and Program Evaluation (ARPE) group of the UMass Donahue Institute (UMDI) is conducting an evaluation of the programs funded by the Academic Support grants, with the objective of helping the state legislature and the Massachusetts Department of Elementary and Secondary Education (ESE) understand program outcomes, implementation, costs, and benefits. UMDI has extensive experience conducting program evaluations of educational initiatives both within and outside of Massachusetts, and has completed numerous studies of similar scope for ESE that utilize the state databases required for the proposed study.

This evaluation plan describes activities through which the program evaluation will assess the effectiveness of the Academic Support programs, as well as provide timely feedback to support ESE's management of the programs and future decisions with regard to replication and sustainability.

Implementation of the evaluation plan will be a collaborative effort involving ESE and their evaluation contractor, the University of Massachusetts Donahue Institute (UMDI). This collaboration is intended to ensure that:

- Evaluation activities are perceived by participants as a critical aspect of their involvement in the program;
- Evaluation activities are designed to minimize burden on field activities and personnel; and
- Evaluation plans and findings are reviewed in a timely manner to inform decision-making regarding evaluation design and/or the programs themselves, as appropriate.

A separate work plan provides additional information about instruments to be developed and data collection activities to be conducted, and the timing of these activities. Combined, these documents represent the most accurate current status of the evaluation plan.

## **Project Goals and Objectives**

The goal of the evaluation is to support ESE and the state legislature in their ongoing efforts to increase the high school graduation rate, thereby improving the quality of life of individuals as well as the work force, economy, and civic engagement in the Commonwealth. The evaluation's objectives are to advance this goal by addressing the following six areas:

1. The quantifiable effects of the Academic Support programs on student attainment of a Competency Determination as well as impacts on related outcomes (i.e., graduation, dropout, and attendance).
2. Fidelity of implementation of the Academic Support programs, as well as perspectives of program personnel regarding program implementation and its relationship to program effectiveness.
3. The economic cost to the Commonwealth of students who do not achieve their Competency Determination.
4. The cost to the Commonwealth of the Academic Support grants, and the benefits received by program participants.
5. Limitations of the evaluation's estimates of Academic Support programs' effects.
6. Recommendations for future research relevant to the quantifiable impacts, fidelity of implementation, costs, and benefits of the Academic Support programs.

In service of these objectives, the evaluation has several primary components, to be finalized collaboratively with ESE. First, a rigorous quasi-experimental design, with closely matched comparison groups selected using propensity score matching strategies, will be used to estimate program impacts on attainment of a Competency Determination, graduation and dropout status, and attendance rates. Second, fidelity of program implementation will be assessed by a survey of



administrators from each Academic Support program. Third, phone interviews of selected program personnel will be conducted to obtain their perspectives regarding aspects of program implementation and its relationship to program effectiveness. Fourth, existing models of the economic costs of high school dropouts will be applied to dropout rates in Massachusetts, to estimate a total cost to the Commonwealth. Fifth, existing economic models will be utilized to calculate program benefits for students, and those benefits will be compared to annual program expenditures. Last, UMDI will describe limitations of the evaluation's quantitative estimates of program impacts, as well as providing recommendations for future research relevant to the Academic Support programs.

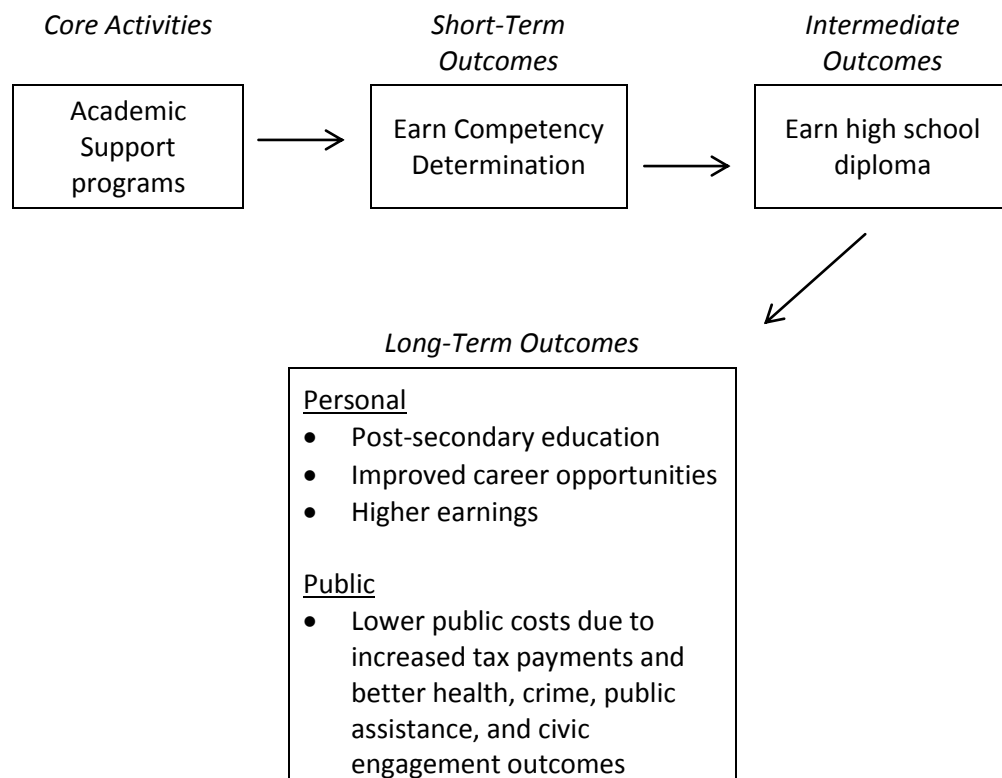
### **Strategies for Managing Risks and Constraints**

The study design is intended to be responsive to identified constraints as reflected in the bid solicitation, the most significant of those constraints being limited resources and time. The study relies on ESE to provide timely access to specific datasets, which will enable UMDI to meet the study's objectives in an efficient and cost-effective manner. A portion of the work will be dependent on self-reported survey and interview data, and UMDI will work with ESE to mitigate related validity risks (e.g., low response rate, unrepresentative sample). In addition, although a rigorous quantitative design is proposed, threats to statistical conclusion validity can not be eliminated entirely. These are discussed in detail in the "Data Limitations in Estimating Program Effects" section below.

As with any program, there is also a risk that implementation will not progress as planned, or that new information will lead to needed or desired adjustments. UMDI has a long history of working collaboratively with clients to understand and address developing needs throughout an engagement. It would be our intent to have regular contact as needed with ESE personnel so that issues affecting the research plans could be identified and adjustments made accordingly.

## Conceptual Framework and Evaluation Questions

UMDI aims to evaluate the implementation, impact, costs, and benefits of the Academic Support programs. The evaluation questions are based on the logic model below and guidance that was provided by ESE. The program's theory of change is that the Academic Support programs lead to earning a Competency Determination, which leads to earning a high school diploma. The diploma then leads to a series of positive outcomes, both personal (e.g., post-secondary education, higher earnings) and for the public (e.g., increased tax payments, improved health outcomes).



The study will focus on the following evaluation questions:

1. What are the impacts of the programs supported through the state's Academic Support grants on student achievement and engagement (i.e., MCAS, graduation, dropout, and attendance)? Do the impacts of each program vary by instructional focus (i.e., ELA, mathematics, and STE)? Which programs are most effective, and for which student subgroups?

2. To what extent are grantees implementing the Academic Support programs with fidelity? What are the perspectives of program personnel with regard to aspects of implementation and its relationship to program effectiveness?
3. What is the estimated economic cost to the Commonwealth of students who do not achieve their competency determination?
4. What is the cost to the Commonwealth of the Academic Support grants that are the focus of this proposal? How does that cost compare to the estimated short-term and long-term benefits received by program participants through the Academic Support programs?
5. In what ways do available data limit estimates of program effects?
6. What further research could provide additional insights relevant to the quantifiable impacts, fidelity of implementation, costs, and benefits of the Academic Support programs?

## Research Design and Analysis

**1. Program impacts.** A quasi-experimental, matched comparison group design will be used to evaluate the effects of participation in the Academic Support programs on student outcomes (i.e., MCAS performance, graduation and dropout status, and attendance rates), compared to closely matched students who do not participate in these interventions. This method will provide strong statistical controls and baseline equivalence between intervention and comparison students, which will substantially minimize threats to statistical conclusion validity.

ESE has reported that students participating in Academic Support interventions are 1.9 times as likely as eligible non-participants to earn a Competency Determination. While this strongly suggests that Academic Support interventions are having a positive impact, program impacts may be systematically underestimated or overestimated, because participating and non-participating students may have differed in important ways prior to intervention. For example, students who participated may have had better attendance than non-participants, which could suggest higher motivation to graduate. If so, the analysis just cited may have overestimated program impacts. Conversely, participants may have struggled more academically than eligible non-participants, in which case past analyses may have underestimated program impacts for the entire eligible population.

To minimize these validity threats, this study will use propensity score matching methods to obtain an unbiased estimate of program impacts by balancing observable characteristics of students so that the characteristics of the students in the intervention and comparison groups are nearly the same. Propensity score matching reduces the set of student characteristics into a single composite score, which provides an important diagnostic tool for assessing the comparability of the two groups (Rubin, 1997). The treatment and comparison groups will be matched on gender, race/ethnicity, free and reduced price lunch status, English language learner (ELL) status, special education (SPED) status, attendance, and prior MCAS scores.

We anticipate that the very large sample of students available through ESE's statewide datasets will make it possible for propensity score matching to reduce baseline differences between intervention and comparison students to within the parameters specified in the standards document of the U.S. Department of Education's What Works Clearinghouse (2014). In addition, statistical adjustments will be conducted for key measurable variables not included in the matching protocol to minimize additional sources of potential selection bias. This approach is common for quasi-experimental studies that are designed to assess the impact of a particular program, such as those being considered in this study (Guo and Fraser, 2010).

UMDI will assess program impacts for participants from school years 2012–13 and 2013–14, the two most recent years for which relevant outcome data are available. Utilizing two years of data will yield a larger sample than would be available from considering only 2013–14, while still producing results which reflect the recent status of the programs. UMDI will discuss with ESE the appropriateness of pooling data across years, which reflects an assumption that the effects of participation in the intervention will be similar across years of the study.

Statistical tests will then be performed to determine if differences in targeted outcomes are significant for groups and subgroups of interest. Multi-level regression and logistic regression will be used to determine if there are differences between groups on key performance measures. Hierarchical (i.e., multi-level) modeling will account for nesting effects associated with the specific school that students attend, thereby improving estimates of the impacts of the Academic Support programs. Results will be presented separately for each of the four funding streams. A set of planned contrasts among treatment conditions and subgroups will be examined to determine if the intervention is more or less effective for some. These subgroups are instructional focus (i.e., ELA, mathematics, and STE), grade level, gender, race/ethnicity, low-income status, and ELL status.

The impact of the intervention for students from different strata of schools/districts will also be examined. That is, there may be reason to believe that the intervention was more or less effective for students from different types of schools/districts. UMDI will collaborate with ESE to identify criteria of greatest interest for defining school/district strata for the purposes of this study (e.g., school/district accountability level, total enrollment, average Academic Support grant amount per participating student).

Carefully selected covariates will be utilized in each of these analyses to minimize potential for bias. These analyses will be used to determine if some subgroups are more or less likely than others to achieve targeted outcomes. The primary unit of analysis will be the student, although some results will be presented at the district level or for each of the four Academic Support grant programs separately.

We anticipate that power to detect significant differences in the comparisons just described will be substantial. Based on data from ESE's past reports to the state legislature, we anticipate a total sample size of approximately 33,600 students (16,500 for FY13, and 17,100 for FY14) across the four Academic Support grant programs. As previously noted, ESE has reported that students participating in Academic Support interventions are 1.9 times as likely as eligible non-participants to earn a Competency Determination. While the proposed analysis design may reduce the magnitude of impacts found for program participants, the strong findings from past ESE analyses suggest that the Academic Support interventions have had substantial impacts. UMDI could therefore reasonably expect to observe at least moderate effects. This combination

of large sample size and moderate effect size suggests that the proposed analyses should have sufficient power to detect any significant differences between treatment and control groups.

Both descriptive and inferential statistics will be computed for the student-level participation data. These analyses will provide important information about the ways in which participating students differ from eligible but non-participating students, and how eligible students differ from ineligible students. Summary statistics (e.g., range, median, mean, and standard deviation) will be presented, as relevant, for number of hours served by school type, subject, term received services, program completion status (i.e., completed, enrolled, withdrew, participated until receipt of 220+ score), schedule type (during school, extended school, weekend/vacation, summer vacation), gender, race/ethnicity, SPED status, ELL status, FRL status, MCAS achievement, attendance, and EWIS risk level.<sup>1</sup>

To help interpret the quantitative findings, UMDI will review existing ESE data sources relevant to the Academic Support programs (e.g., reports to the legislature,) as well as the survey and interview data collected from sites by UMDI for this study.

If subgroups of interest have high levels of missing data, statistical techniques (e.g., sample weighting) will be used to account for these differences. We do not anticipate that this will be a problem, however, for two reasons. First, our review of ESE's reports to the legislature regarding the Academic Support grants, combined with our experience with datasets relevant to the proposed study, suggest that there will be minimal missing data. This is because districts have already provided ESE with the unique, state-assigned student identifiers (SASIDs) of participating students, and the SASIDs can be used to obtain student demographic characteristics and performance outcome indicators from ESE's statewide, student-level databases. Second, the potential for bias caused by missing data will be mitigated by the large sample size.

Study Sample. The study sample is well defined. ESE has provided clear protocols for identifying eligible and participating students, and the research team can readily identify the schools and districts from which participants were selected. This is important because it will allow the research team to clearly identify the treatment group, and to define a sample pool from which a rigorously matched comparison group of eligible non-participants can be constructed.

ESE's FY14 Report to the Legislature states that 153,108 students in the classes of 2003–2018 were eligible for services in FY14, and that 11 percent (approximately 17,100 students) were expected to be served. Similar numbers of students were eligible and served in prior years (e.g., 16,500 for FY13). With nearly 90 percent of eligible students not being served, there are an ample number of students from which comparison groups could be identified.

---

<sup>1</sup> UMDI will use EWIS data to calculate descriptive statistics only, as EWIS risk level data are not intended to predict student outcomes.

The high percentage of eligible non-participants also permits comparison group students to be drawn from districts that participated in the intervention. This is advantageous in terms of reducing possible sources of error, due to unmeasured differences between intervention and comparison groups that are attributable to the schools and districts in which students are nested. It also strengthens the evaluation design because all of the districts in the state with the highest number of eligible students participated in the intervention, leaving no appropriate non-participating comparison districts.

Separate comparison samples will be identified for each intervention, because different populations of students are served by each intervention, and determining the efficacy of each intervention can be best achieved by selecting comparison samples that are as similar to each of the different treatment samples as possible.

Data Sources. The ESE student-level databases described below will be used from all years needed for the proposed analyses.

- **SIMS.** SIMS data provide outcomes (e.g., attendance) as well as student demographics and ELL, FRL, and SPED status. October and June SIMS files would be requested for this study.
- **MCAS.** MCAS data provide information on students' performance on the MCAS exams.
- **EWIS.** EWIS data are used to describe students' level of risk for not reaching important academic milestones (e.g., risk of not passing all courses as a freshmen)
- **4- and 5-year graduation files.** ESE prepares 4- and 5-year graduation files each year, which are used (in part) to calculate district graduation rates. UMDI could derive the information provided in these files (as well as the annual dropout files described next) from multiple years of October and June SIMS data if needed, but obtaining them from ESE would be more efficient.
- **Annual dropout files.** ESE prepares an annual dropout file that is used (in part) to calculate district dropout rates.
- **Academic Support student data.** As part of the annual reporting for the Academic Support programs, for each student who received services sites report the SASID, hours of service provided, and participation/completion status. These data will identify all students in the intervention groups and provide additional information to be utilized in the impact analyses.

- **Historical CD file.** ESE has maintained historical files that identify students who have attained their Competency Determination at particular points in time. These data are used to identify eligible students, and to report on outcomes relevant to this study.
- **Eligibility file.** ESE has maintained a list of all eligible students in each district for each year of the grant. These files are used to identify eligible students.

**2. Program implementation.** ESE and the legislature requested evaluation of implementation fidelity as well as perspectives of program personnel regarding specific aspects of implementation and its relationship to program effectiveness. UMDI will conduct a survey of all grantees to evaluate implementation fidelity, and phone interviews of selected program sites to obtain perspectives of program personnel. UMDI will draft all data collection instruments in collaboration with ESE personnel. These will be revised through an iterative process of feedback and revision, and final instruments will be agreed upon prior to data collection and analysis.

Survey of implementation fidelity. Implementation fidelity refers to the degree to which an intervention is delivered as intended. Therefore, the first step in measuring implementation fidelity is determining what the intended intervention was. The Academic Support programs do not prescribe a specific program model; instead, they provide required and suggested program dimensions. For example, for the Allocation Grant program, some of the required program dimensions are that students must be selected based on specified MCAS score cutoffs, student-to-teacher ratios must be no more than 10:1, and the district may not charge any students for participating. Some of the suggested program dimensions include the following:

- Clear evidence is provided that the Educational Proficiency Plans (EPP) and other student success plans are used to identify knowledge gaps for individual students and that instruction is targeted appropriately.
- Teachers are involved with helping to plan engaging, hands-on, and relevant curriculum that addresses the aforementioned gaps in students' knowledge and skills.
- The program supports and provides continuity with classroom instruction.
- The program design provides opportunities for students to demonstrate their learning and for teachers to provide feedback to the students and parents/guardians.

For each of the four Academic Support programs, UMDI will develop an online survey that asks program administrators to provide self-reported levels of implementation of required and suggested program dimensions. It will also ask for designation of the program categories listed in the bid solicitation (i.e., straight remediation, online tutorials, project-based learning, work-based learning, service learning, etc.) as well as others that might be identified as relevant during the instrument development process, in collaboration with ESE. Descriptive statistics of survey findings will be reported to indicate fidelity of program implementation for each grant program. Having reviewed the narrative evaluation templates that are available online for the Academic



Support programs, we believe that this approach will provide a more detailed and quantifiable estimate of the fidelity of program implementation than has been available in the past.

The survey will be administered in relation to the school year 2015–2016 and summer 2016 programs. In consultation with ESE, UMDI may propose to incorporate these survey items into the required end-of-program reporting process that is already in place for the school year 2015–16 and summer 2016 grants, as this may reduce the burden on respondents, reduce duplication of the information requested, and substantially increase response rates.

Interviews of program personnel. Phone interviews of 30 program personnel will be conducted to obtain their perspectives regarding aspects of program implementation and its relationship to program effectiveness. The semi-structured interview protocol will include open-ended questions corresponding to the following (slightly revised) questions from the “Additional Guidelines and Background Information” document included with the bid solicitation:

1. What are the impacts of the different types of support (straight remediation, online tutorials, project-based learning, work-based learning, service learning, etc.)? What is it about the program’s strategies that make them effective for the particular kinds of students for which they appear to be successful? What strategies are particularly effective for improving English language arts and literacy? For improving mathematics?
2. For the programs that intervene with at-risk 8th graders and 9th graders, how do the programs help with their transition to and success in high school?

ESE would need to facilitate school compliance with the interviews, which will be scheduled to take 20–30 minutes. In consultation with ESE, UMDI will select the sample of 30 interviewee schools to represent all four funding streams, and relevant subgroups (e.g., programs with different types of support and focused on different academic disciplines) within each funding stream on the questions above. Interviews will be conducted during the spring of 2016.

Qualitative data from interviews and focus groups will be managed and analyzed through a multi-step process. Interviews will be recorded, after explaining procedures to assure confidentiality and securing permission from participants. Field notes will then be developed based on the recordings, using common field note forms across analysts. A set of conceptual codes based on the study’s research questions will guide the coding of the field notes. NVivo qualitative analysis software will be utilized for data management, coding, and retrieval of findings. The data will be analyzed through multiple close readings, using a constant comparative method, leading to the development of rich explanations in relation to the research questions (Patton, 1990). Analysis will also emphasize the identification of actionable items that could support future technical assistance and program improvement.

**3. Economic cost to the Commonwealth of high school dropouts.** The Academic Support programs are intended to support students in earning their Competency Determination, which is required for earning a high school diploma. Therefore, to estimate the cost to the Commonwealth of the education problem being addressed through the Academic Support programs, the evaluation will estimate the public costs incurred in relation to high school dropouts. These costs may include lost tax payments and increased public expenditures in the health, criminal justice, and welfare systems; some savings are also realized in the public education system due to lower utilization of higher education resources by high school dropouts and hence reduced government subsidies (Belfield et al., 2012; Sum et al., 2009).

Assessing each of these costs individually is beyond the scope of this evaluation. Instead, UMDI will identify existing approaches to estimating the public costs of high school dropouts and apply those approaches to Massachusetts. Multiple examples of relevant strategies are in existing literature. For example, the Center for Labor Market Statistics (2009, p. 16) estimated that “the average high school dropout will cost taxpayers over \$292,000 in lower tax revenues, higher cash and in-kind transfer costs, and imposed incarceration costs relative to an average high school graduate.” Conducting a similar assessment at the national level, Belfield et al. (2012) estimate that “opportunity youth”—those aged 16–24 who have dropped out of high school and experience a variety of barriers to employment and/or education—create a total taxpayer burden of \$258,240 over their lifetime, with an immediate fiscal cost of \$13,900 annually between the ages of 16–24.

The evaluation will review existing literature to identify the most appropriate models and strategies for estimating the economic costs to the Commonwealth of high school dropouts. The estimate for each high school dropout will be multiplied by the total number of dropouts to estimate a total cost to the state of each year’s cohort. While the evaluation focuses on economic costs, UMDI recognizes that there are also social costs to the Commonwealth related to high school dropouts. However, these are not assessed in the proposed evaluation, due both to available resources and limited evidence as to their magnitudes.

**4. Comparing program costs and benefits received by program recipients.** Assessing program costs is straightforward: they will be defined as the annual amounts allocated in the state budget to each of the four Academic Support grant programs. Assessing benefits received by program recipients will follow an approach similar to the one described above for estimating economic costs to the Commonwealth. In short, a review of existing literature on the costs to individual students of dropping out of high school will lead to selecting strategies for calculating short-term and long-term economic costs per student. These costs will then be applied as an equivalent amount of economic benefit per student to those students who are identified as having received a high school diploma as a result of the Academic Support programs. (As discussed in the section below regarding data limitations in estimating program effects, determinants of

graduation are multi-faceted, but the rigorous, quasi-experimental design of the evaluation will substantially increase the ability to estimate impacts that are specific to the Academic Support programs.)

Numerous estimates appear in the literature that summarize lifetime earnings based on educational attainment. For example, Carnevale et al. (2011) estimated that median lifetime earnings during a 40-year career for those who did not earn a high school diploma or GED is \$973,000 (in 2009 dollars), compared to \$1,304,000 for those whose highest level of educational attainment is a high school diploma—indicating an annual earnings premium for high school graduates of \$8,275 (34 percent) and a lifetime premium of \$331,000. The Bureau of Labor Statistics (2014) provides a similar estimate of the annual earnings premium, finding that median earnings in 2014 for full-time workers over the age of 25 were \$25,425 for those with less than a high school diploma and \$34,803 for those whose highest level of educational attainment was a high school diploma—indicating a 37% earnings premium for high school graduates.

Estimating short-term economic costs of dropping out of high school are more problematic than estimating lifetime costs, primarily because a high percentage of the potential comparison group is enrolled in college, with minimal earnings. Looking only at a short-term timeframe reveals an economic *advantage* to dropping out of high school compared to enrolling in college. Baum et al. (2013, p. 13) demonstrate that this short-term advantage is in fact realized for high school graduates who do not enroll in college, and it would presumably also be realized by high school dropouts who enter the workforce. However, as Baum et al. (2013) show, the average college graduate regains the lost ground within about a decade compared to high school graduates, and presumably more quickly in relation to high school dropouts.

Belfield et al. (2012) partially address these limitations, investigating short-term impacts on earnings for opportunity youth (as defined above) compared to employed youth ages 16–24. They find an annual earnings advantage of \$9,760 for the employed youth. However, this comparison does address the issue of short-term costs for all high school dropouts, because dropouts are presumably represented in both groups. UMDI will continue to seek appropriate models for addressing this question, but our conclusions from an initial literature review suggest that developing high quality estimates of short-term economic costs may be beyond the scope of this evaluation. Nonetheless, Belfield et al. argue that “The [short-term] burden understates the true costs” (p. 17). We agree, and would suggest that the evaluation resources devoted to economic benefits of the program be focused on long-term benefits (i.e., lifetime and annualized lifetime economic benefits). In addition, UMDI will report on important short-term benefits that are not strictly economic, such as relative rates of employment, criminal justice system involvement, negative health outcomes, and receipt of public assistance.

**5. Data Limitations in Estimating Program Effects.** The evaluation is designed to provide a robust estimate of program effects. Nonetheless, the available data have some limitations, which are discussed here along with potential strategies to address them. First, many of the sites participating in the Academic Support programs are also participating in other interventions that could influence the outcomes of interest (i.e., MCAS performance, graduation and dropout status, and attendance rates). Therefore, the effects of the Academic Support interventions are partially confounded with the effects of other present and past school and district initiatives. Nonetheless, this evaluation is well-designed to detect changes due to current ESE programs. One way that we address this challenge is through our quasi-experimental design, in which students are closely matched on baseline characteristics, thereby increasing the chance that observed effects can be attributed to the intervention.

Second, it may be difficult to identify a comparison group for the group that ESE refers to as “post 12<sup>th</sup> grade” students, defined as students from the classes of 2003–2015 who have not yet met their Competency Determination. Each year ESE identifies students who are eligible for Academic Support programs and collects information relevant to participating students, but it is not clear if ESE gathers comparable information for eligible non-participants from the post 12<sup>th</sup> grade group. Lack of an adequate sample from which to draw a rigorous comparison group may prevent UMDI from applying the quasi-experimental methods proposed in this study to examine impacts of the Academic Support interventions on post 12<sup>th</sup> grade students. However, important conclusions about program impacts can still be drawn. For example, UMDI will compare MCAS outcomes for post 12<sup>th</sup> grade program participants to younger students for whom appropriate comparison data are available.

Third, starting in the 2013–14 school year, all students in the Boston Public Schools will be designated as receiving free and reduced price lunch (FRL) in the state’s student information management system, which will impede our ability to match students by income status. UMDI is aware that ESE is working with researchers to manage the transition from the past FRL variable in SIMS to the new “economically disadvantaged” (ECODIS) variable. We look forward to contributing to that conversation through this research project, and we will apply any guidance provided by ESE to our analysis and reporting efforts.

**6. Future research.** As the study progresses, UMDI will identify additional questions that could be answered by future research on the implementation, impacts, costs, and benefits of the Academic Support programs. The following are some initial ideas and suggestions:

- A more extensive study of the range of program types within each grant program and the details of their implementation, including promising practices, successes, challenges, technical assistance needs, student perspectives, and lessons learned to share with other schools and districts. This work could be focused particularly on the programs shown to

be most effective in the current study, with implications for replication in sites whose current strategies appear to have smaller impacts.

- A longitudinal study to investigate more closely the actual long-term outcomes of program participants compared to eligible non-participants. Investigating domains such as education, employment, earnings, health, criminal justice, and public assistance could lead to more refined estimates of short-term and long-term program costs and benefits.

## Deliverables and Timeline

The evaluation is designed to provide valuable, utilization-focused feedback in relation to the study's goals and research questions. To ensure that ESE and the state legislature are informed of study progress and its findings on a timely basis, UMDI will prepare the deliverables specified below. The dates specified should be considered tentative, pending further discussion with ESE.

- January 2016 – Initial draft of evaluation and work plan, with final draft to be developed in collaboration with ESE regarding outstanding questions and joint decisions.
- February 2016 – Data sharing agreement requesting access to the ESE datasets needed to conduct the proposed analyses.
- March 2016 – Implementation of fidelity survey instruments.
- March 2016 – Implementation of program personnel phone interview instruments.
- June 2016 – Status report to ESE and the House and Senate Committees on Ways and Means. As described in the “Additional Guidelines and Background Information” document from the bid solicitation, the report will offer “the status and preliminary results of the evaluation and [describe] any obstacles encountered in access to data or other information that is negatively affecting the completion of the study.”
- December 2016 – Status report to ESE and the House and Senate Committees on Ways and Means. Same description as June 2016 status report.
- June 2017 – Comprehensive final report to ESE and the House and Senate Committees on Ways and Means describing methods and findings regarding program implementation, impacts, costs, benefits, limitations, and suggestions for future research.
- In addition, UMDI will communicate as frequently as needed with ESE by phone, by email, and in person to make decisions about the conduct of the evaluation, to share information, to provide evaluation feedback, and to resolve emerging issues or provide debriefings on sensitive findings. Prior to any scheduled project management calls, UMDI will provide a written agenda of items to be discussed, and a written follow-up that reflects decisions made during the call and next steps.

### Timeline of Academic Support Evaluation Activities, (1/1/16 – 6/30/2017)

**A = Development Begins; B = Date Finalized; C = Data Collection Complete**

Activity	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
Evaluation Plan	A	B																
Work Plan	A	B																
Analysis Plan						A	B											
Data Sharing Agreement	A	B																
Initial Data Review	A						B											
Implementation Fidelity Survey Instrument	A		B	C														
Program Personnel Phone Interview Instrument	A		B		C													
Data Analysis							A										B	
Check-in Meetings*	A	A	A	A	A	A	A	A		A		A		A		A		A
Status Report						A						A						
Final Report																		A

\*UMDI proposes to hold regular check-in meetings (every three weeks for the first seven months of the evaluation, and then every six weeks for the remainder of the project). In addition, UMDI will communicate as frequently as needed with ESE by phone, by email, and in person to make decisions about the conduct of the evaluation, to share information, to provide evaluation feedback, and to resolve emerging issues or provide debriefings on sensitive findings. Prior to any scheduled project management calls, UMDI will provide a written agenda of items to be discussed, and a written follow-up that reflects decisions made during the call and next steps.

## References

- Baum, Sandy, Jennifer Ma, Kathleen Payea. 2013. *Education Pays 2013: The Benefits of Higher Education for Individuals and Society*. Princeton, NJ: College Board.  
<https://trends.collegeboard.org/sites/default/files/education-pays-2013-full-report.pdf>.
- Belfield, Clive R., Henry M. Levin, Rachel Rosen. 2012. *The Economic Value of Opportunity Youth*. Washington, DC: Corporation for National and Community Service.  
[http://www.civicenterprises.net/MediaLibrary/Docs/econ\\_value\\_opportunity\\_youth.pdf](http://www.civicenterprises.net/MediaLibrary/Docs/econ_value_opportunity_youth.pdf).
- Bureau of Labor Statistics. 2014. *Earnings and Unemployment Rates by Educational Attainment*. Washington, DC: Author. [http://www.bls.gov/emp/ep\\_chart\\_001.htm](http://www.bls.gov/emp/ep_chart_001.htm).
- Carnevale, Anthony P., Stephen J. Rose, Ban Cheah. 2011. *The College Payoff: Education, Occupations, Lifetime Earnings*. Washington, DC: Georgetown University Center on Education and the Workforce.  
<https://www2.ed.gov/policy/highered/reg/hearulemaking/2011/collegepayoff.pdf>.
- Guo, Shenyang, Mark Fraser. 2010. *Propensity Score Analysis: Statistical Methods and Applications*. Thousand Oaks, CA: Sage.
- Massachusetts Department of Elementary and Secondary Education. 2014. *Report to the Legislature: MCAS Support Programs – Fiscal Year 2014*. Malden, MA: Massachusetts Department of Elementary and Secondary Education.  
<http://www.doe.mass.edu/research/reports/2014/05MCAS-ASPfy14.pdf>.
- Massachusetts Graduation and Dropout Prevention Recovery Commission Report. 2009. *Making the Connection: A Report of the Massachusetts Graduation and Dropout Prevention and Recovery Commission*. Malden, MA: Massachusetts Department of Elementary and Secondary Education. <http://www.doe.mass.edu/dropout/resources.html>.
- Patton, Michael Q. 1990. *Qualitative Evaluation and Research Methods (2nd ed.)*. Newbury Park, CA: Sage.
- Rubin, Donald B. 1997. Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine*, 127(8\_Part\_2), 757–763.



Sum, Andrew, Ishwar Khatiwada, Joseph McLaughlin, Sheila Palma. 2009. *The Consequences of Dropping Out of High School: Joblessness and Jailing for High School Dropouts and the High Cost for Taxpayers*. Boston, MA: Northeastern University Center for Labor Market Studies.

What Works Clearinghouse. 2014. *Procedures and Standards Handbook, Version 3.0*. Washington, DC: Institute of Educational Sciences.  
[http://ies.ed.gov/ncee/wwc/pdf/reference\\_resources/wwc\\_procedures\\_v3\\_0\\_standards\\_handbook.pdf](http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf).



Analysis of Longitudinal Data in  
Education Research Program

FEBRUARY 2016

---

Analysis of Longitudinal Data in Education Research Program

MARCH 2016

# **Massachusetts Educator Preparation and Licensure Year 1 Research Plan**

**James Cowan  
Dan Goldhaber  
Roddy Theobald**

# **Massachusetts Educator Preparation and Licensure Year 1 Research Plan**

**March 2016**

**James Cowan  
Dan Goldhaber  
Roddy Theobald**



AMERICAN INSTITUTES FOR RESEARCH®

1000 Thomas Jefferson Street NW  
Washington, DC 20007-3835  
202.403.5000

**[www.air.org](http://www.air.org)**

Copyright © 2016 American Institutes for Research. All rights reserved.

XXXX\_M0/YR



# Contents

	Page
Introduction.....	1
Research Plan.....	1
Research Questions.....	1
Research Methods.....	6
Timeline and Deliverables .....	9
References.....	9



# Introduction

The Massachusetts Department of Elementary and Secondary Education (ESE) has identified improving the quality of educator preparation as a key component of its strategic plan. This initiative is consistent with over a decade of educational research, which has consistently identified the classroom teacher as one of the most variable schooling factors affecting student achievement. And, as we describe in more detail below, an emerging body of evidence has further identified teacher preparation as a potentially important lever for improving the teaching workforce.

The Analysis of Longitudinal Data in Education Research Program (ALDER) at the American Institutes for Research has partnered with ESE to study teacher effectiveness and educator preparation programs in Massachusetts.<sup>2</sup> In this report, we outline the first year research plan for this collaboration. The first phase of this research will focus on providing an overview of the educator preparation system in Massachusetts. In particular, we will provide descriptive evidence of how teacher outcomes vary across license types, educator preparation institutions, and educator preparation programs in Massachusetts. ALDER and ESE have jointly identified three key workforce outcomes to guide the research: student achievement on the Massachusetts Comprehensive Assessment System (MCAS), summative performance ratings under the Massachusetts Standards for Effective Teaching Practice, and retention in the public school system. The research will culminate in a report providing empirical evidence answering the research questions defined below.

## Research Plan

### Research Questions

The research program described below grew out of a research proposal submitted jointly by ALDER and ESE to the Institute for Education Sciences in 2015. To define the specific questions upon which we plan to focus, ALDER staff reviewed the ESE Strategic Plan and proposed several potential topics drawn from the broader literature on teacher preparation. We then worked with ESE to refine the proposed research questions to best support ESE priorities. During the first year of this research grant, we will take a broad view of the preparation and licensure landscape in Massachusetts and provide some basic descriptive evidence on how teachers' workforce outcomes vary with the nature of their preservice preparation experience. We have identified three primary research questions:

1. What is the variation across licensure pathways, educator preparation institutions, and educator preparation programs in student achievement gains?

---

<sup>2</sup> Throughout this proposal, we refer to an institution authorized to recommend graduates for licensure, such as a university or organization offering alternative licensure programs, as a "preparation provider" or "institution" and to a specific course of study, such as "Elementary Teacher, Grades 1–6," as a "program." An institution may therefore include multiple programs.



2. What is the variation across licensure pathways, educator preparation institutions, and educator preparation programs in teacher evaluation results?
3. What is the variation across licensure pathways, educator preparation institutions, and educator preparation programs in teacher retention?

These three questions are designed to present a preliminary indication of the performance of program completers using a variety of important measures that have been directly linked to student learning. For each of these research questions, we will focus our analysis on three different descriptors of teachers' preservice preparation.

The first descriptor is a teacher's route into the profession, which we will classify based on the type of license a teacher possesses and whether the preparation program is an alternative program based outside a higher education institution. Massachusetts offers two alternative license types for teachers who have not completed a state-approved educator preparation program. The temporary license is valid for one year for teachers holding an out-of-state license but who have not completed the Massachusetts Tests for Educator Licensure (MTEL) requirements. The preliminary license, valid for five years, permits teachers who have completed the MTEL requirements but not an educator preparation program to teach in the state. In addition to the alternative license types, Massachusetts hosts a number of alternative preparation programs that allow teachers to enter the profession through programs hosted outside traditional higher education institutions. While all educator preparation programs are designed to meet state certification standards, there may be more variation in program design and curricular requirements between licensure pathways than among programs within the same pathway. Our first set of analyses will therefore compare teachers who enter through each of these alternative pathways to those who complete a traditional educator preparation program.

We will next focus on differences in educator effectiveness across educator preparation institutions within each licensure pathway. The purpose of this analysis is twofold. First, we plan to estimate individual contrasts in educator effectiveness and retention rates for particular institutions. Such contrasts help us to understand the variation between educator preparation programs. Second, we plan to quantify the variability in educator outcomes across institution by estimating the variance in institution effects. This latter measure provides a summary of the dispersion in effectiveness across the state and provides some context for whether the differences in outcomes we observe are either substantively or statistically important.

Finally, we will focus within institutions and estimate the variation in educator effectiveness across programs. Educator preparation institutions often offer programs for both undergraduate and graduate students and typically train teachers for a number of different teaching roles through separate programs. Within the same institution, different programs may have different curricula, admissions, or student teaching requirements and may therefore produce teachers with differing average outcomes. As with educator preparation institutions, our objective will be to both estimate contrasts for individual programs and to quantify the magnitude of the variability in outcomes across programs within a single institution. As preparation programs train teachers for differing roles, we plan to define the program in such a way that contrasts are made relative both to that specific program and to teachers trained for that specific role. Given that disaggregating results by program results in many fewer teacher observations per preparation

program, we caution that we are unlikely to be capable of reliably estimating program estimates for a large number of individual programs. However, for larger programs, such contrasts may allow ESE to identify institutions that produce effective teachers in particular high priority fields.

### ***Existing Evidence on Educator Preparation and Student Achievement***

Empirical research has consistently found that there is substantial variation in test score gains among teachers. These differences in educator effectiveness comprise one of the most variable schooling inputs in the research literature (Goldhaber et al., 1999; Rivkin et al., 2005). Measures of teacher effectiveness based on student standardized test scores additionally provide information about students' outcomes on a number of other outcomes, including future academic achievement, performance on higher-order academic tests, educational attainment, and earnings (Chetty et al., 2014b; Kane et al., 2013; Kinsler, 2012; Papay, 2011). Student achievement measures are also one component of the Massachusetts educator evaluation framework (Massachusetts Department of Elementary and Secondary Education, 2014).

Much of the evidence on the effectiveness of licensure requirements comes from comparisons of fully certified teachers to those with provisional or temporary credentials. The results from this literature are mixed and probably at least partially dependent policy context, including licensure requirements and the recruiting practices of alternative certification programs. Teach for America has probably been the most studied group of teachers with alternative certification, and both random assignment experiments (Clark et al., 2013; Glazerman et al., 2006) and observational studies (Boyd et al., 2006; Kane et al., 2008; Xu et al., 2011) suggest that they are at least as effective as regularly certified classroom teachers. Studies of other groups of alternative entry or lateral entry teachers have produced inconsistent results (Bastian and Henry, 2015; Boyd et al., 2006; Clotfelter et al., 2010; Kane et al., 2008; Papay et al., 2012; Sass, 2015), although a few random assignment studies have not found these groups to be statistically significantly less effective than traditionally certified teachers (Clark et al., 2013; Constantine, 2009). While the statistical significance, and sometimes direction, of these results differ, the differences in effectiveness are typically, although not uniformly, less than those between first-year and third-year teachers. While the policy context appears to matter for the results on licensure, the main point of commonality among these studies is that most of the variation in teacher quality is within, rather than between, categories of teacher preparation.

Researchers have begun to assess the influence of individual educator preparation institutions on student achievement in a number of settings. This research suggests that there are differences in average effectiveness across programs (Boyd et al., 2009; Goldhaber et al., 2013; Koedel et al., 2015; Mihaly et al., 2013; Henry et al., 2014). However, the magnitude of the variation differs across states and by research methodology and the substantive importance of these differences remains unclear. On the higher end, Boyd et al. (2009) estimate that the preparation institution explains about 15 percent of the variation in teacher value-added in New York City. At the other extreme, Koedel et al. (2015) find that preparation programs in Missouri explain only 0-3 percent of the variation in teacher effectiveness. Massachusetts provides an interesting contrast to many of these study sites. Most of these early studies limit their comparisons of the variability across institutions to only a single program type. Massachusetts, on the other hand, hosts both a large number of institutions and a wide variety of programs (U.S. Department of Education, 2013). A

study of teacher preparation in Massachusetts therefore potentially provides a greater range of variation in the nature of preservice training than has been the case in other states.

To address the first research question, we will assess the variation across licensure pathways, programs, and institutions in student achievement gains on the Massachusetts Comprehensive Assessment System (MCAS). Given the use of student test scores, our analysis will focus on math and reading test results for teachers in grades 4-8 and 10 and science test results for teachers in grades 5, 8, and 9/10 during the 2011-2015 school years. Importantly, the use of test scores as an outcome measure limits the sample to teachers in these tested grades and subjects only, which may not be representative of the teaching profession as a whole. We will further limit our analysis to teachers in their first six years of classroom teaching. We make this sample restriction for two reasons. First, educator experience is not tracked in the Education Personnel Information Management System (EPIMS), on which we will rely for information on teacher assignments. The returns to experience are quite high during the first few years of teaching and it is therefore important to control for teaching experience in order to avoid conflating differences in the likelihood of attrition across programs with differences in the average level of effectiveness among teachers initially hired into public schools in Massachusetts. Second, the effectiveness of program graduates may change over time and limiting the sample to recent graduates ensures that comparisons reflect the current operation of preparation programs.

One benefit of the value-added analyses proposed above is that there is relatively detailed consideration of the ability of such models to provide unbiased estimation of teacher effects and of the trade-offs inherent in alternative modeling decisions (e.g., Chetty et al., 2014a; Goldhaber et al., 2013; Ehlert et al., 2013). However, a number of states use an alternative specification of student learning, student growth percentiles, for descriptive measures of teacher performance and accountability purposes (Betebenner, 2009). Student growth percentiles situate students in the conditional distribution of current achievement given prior test scores and therefore provide a measure of growth that is defined relative to other peers with similar achievement histories. One potential drawback of student growth percentiles is that the specification of student achievement is more restrictive than standard value-added models that additionally control for student demographic characteristics, special learning needs, and the achievement of other students in the same classroom. These restrictions may reduce the ability of such models to adjust for differences across classrooms in the population of students served or to capture classmates' influences on students' own test scores. Analyses generally find that both types of models produce similar estimates of educator effectiveness, but also that the additional controls for student and classroom context are important for reducing bias from the matching of students and teachers (Chetty et al., 2014a; Goldhaber et al., 2013). However, there is little evidence on the consequences of how different types of value-added modeling procedures affect the evaluation of educator preparation programs. In order to provide such guidance, we will additionally conduct analyses of the effectiveness of educator preparation programs using the student growth percentile measures computed by the Massachusetts Department of Elementary and Secondary Education.

### ***Educator Preparation and Performance Evaluations***

As noted above, the analysis of student test score results necessarily limits the sample of teachers to those in tested grades and subjects. This group represents a fairly small proportion of all

teachers and disproportionately covers teachers with elementary education backgrounds. We will therefore continue the investigation of teacher effectiveness by analyzing teacher evaluation results, which cover all classroom teachers in Massachusetts.

Massachusetts introduced standards of effective teaching practice during the 2013-2014 school year. The standards are organized into four strands: curriculum, planning, and instruction; teaching all students; family and community engagement; and professional culture. Districts have developed systems of indicators to implement the standards. These frameworks include evidence from a number of sources, including observations of teachers and other artifacts of teaching practice. These indicators are aggregated across the four standards to form a summative performance rating. We will include in our evaluation the summative performance ratings for each of the four standards as well as the overall rating.

Although not specifically based on the Massachusetts standards, there is empirical evidence from several sources that individual indicators of the Massachusetts educator evaluation system are related to other measures of teacher effectiveness. Classroom observations of teacher practice predict student achievement gains and students' reports of classroom environment (Blazar, 2015; Grossman et al., 2013; Kane et al., 2011, 2013). Similarly, evaluations by administrators or mentors are predictive of test-based measures of effectiveness (Harris and Sass, 2014; Jacob and Lefgren, 2008). Finally, assessments like those offered by the National Board for Professional Teaching Standards, which include an evaluation of educators' ability to assess student needs and appropriately tailor instruction, identify teachers who are more effective at raising student test scores (Cantrell et al., 2008; Cowan and Goldhaber, 2015). Although we know much less about how these performance measures are linked to educator preparation, there is some emerging evidence that there are meaningful differences across preparation institutions (Campbell and Ronfeldt, 2015).

Our analysis of teacher performance evaluations will follow the same general plan as that of teacher contributions to student test scores, but will include a broader sample of teachers. As before, we will limit the analysis to teachers in their first six years of classroom teaching in order to measure teaching experience as accurately as possible and ensure that comparisons are made among recent program completers. As the evaluation data is available for the 2014 and 2015 school years, our analysis will focus on early career teachers who are still teaching during these two school years.

### ***Educator Preparation and Retention***

Teachers leave the profession at a higher rate during the first few years in the classroom than more senior teachers. According to the most recent Schools and Staffing Survey, about 7 percent of novice teachers (those with 1 to 3 years of experience) left the profession between 2012 and 2013 (Goldring et al., 2014). Teacher turnover may harm student achievement through at least two main channels. First, departing teachers are frequently replaced by novices, who are much less effective on average than teachers with even a few years of experience (Clotfelter et al., 2007; Rockoff, 2004). Second, turnover may disrupt collaborative relationships or otherwise harm school climate (Ronfeldt et al., 2013).

Researchers have begun to link elements of preservice preparation to teacher retention. High-quality practice teaching and methods coursework appear to improve retention once teachers enter the classroom (Ronfeldt, 2012; Ronfeldt et al., 2014). Retention rates also vary meaningfully across preparation institutions and licensure pathways (Goldhaber and Cowan, 2014; Kane et al., 2008). These findings suggest that improving the quality of preservice preparation may be one response to teacher shortages.

Our analysis of retention will focus on the likelihood that early career teachers leave public schools in Massachusetts. EPIMS provides an annual database of teacher assignments for the 2010-2015 school years. This database provides a sample of teachers who have taught for up to 6 years. We will therefore follow cohorts of teachers who complete educator preparation programs in Massachusetts between 2009 and 2014 and work in Massachusetts public schools during 2010-2015. We describe the research methodology in more detail below, but our analysis will focus on estimating differences among licensure pathways, educator preparation institutions, and programs in the probability that teachers leave public schools during each of the first six years of teaching.

## Research Methods

### *Educator Preparation and Student Achievement*

Our analysis of the student achievement effects of teachers with various educator preparation backgrounds will focus on the Massachusetts Comprehensive Assessment System (MCAS) using value-added modeling. Value-added models assess the contributions of teachers to student learning by focusing on changes in achievement test scores from one year to the next. The focus on changes in student achievement, instead of on student achievement levels, allows researchers to isolate the effects of teachers from family background and other schooling factors that influence learning.

The primary value-added research design we employ can be summarized by the following regression equation:

$$Y_{ijst} = X_{ijst}\beta + P_{jt}\delta + \alpha_s + \epsilon_{ijst}. \quad (1)$$

In Eq. (1), the subscript  $i$  indexes students,  $j$  indexes teachers,  $s$  indexes school, and  $t$  indexes year. We regress student achievement on the MCAS for each student on a set of variables  $X$  that includes prior achievement, academic and demographic characteristics, and peer and school characteristics. The teacher characteristics  $P$  include descriptions of their preparation background (i.e., licensure pathway, preparation institution, program). We are interested in the teacher preparation variables  $P$  and their associated standard errors. In addition to these variables, we include school fixed effects  $\alpha_s$ , which ensure that comparisons of teachers from different programs are made only within schools. For high school students, these models additionally include controls for the class assignments of students to ensure that teachers are only compared relative to others teaching students in the same track. Similar specifications of the student achievement equation provide the foundation for much of the research literature on teacher credentials and preparation (Boyd et al., 2009; Goldhaber et al., 2013; Mihaly et al., 2013).

Estimation of the value-added model Eq. (1), like the other empirical models we will estimate for teacher evaluation results and retention, produce both a point estimate of a particular program's average teacher effectiveness and a standard error that describes the uncertainty in the estimate. The uncertainty results from the limited nature of the data. We observe only a limited number of teachers in a limited number of years per program. And in each year, as few as 20 students may inform the estimate of an individual teacher's effectiveness. Prior research has documented that teacher performance varies from year to year due to the small number of student observations available per teacher, changes in classroom chemistry, and changes over time in the effectiveness of individual teachers (Chetty et al., 2014a; Goldhaber and Hansen, 2013; McCaffrey et al., 2009). Typically, the resulting uncertainty in estimates of program effectiveness is quantified through the use of confidence intervals, which describe a range that would be expected to cover a certain percentage of replications of the estimation procedure from different samples. In the social sciences, researchers typically use a 95 percent confidence interval, which is expected to cover 95 percent of the estimates observed from repeated trials. Some accountability systems for educator preparation programs rely on the use of confidence intervals and identify programs whose confidence intervals exclude certain critical values, such as the mean of the program effects (Lincove et al., 2014). This approach is similar to hypothesis testing in social science research. However, the choice of a particular confidence interval for identification of exemplary or ineffective programs is also implicitly a choice for the sensitivity of the test. Selecting a higher level for the confidence interval reduces the likelihood of inadvertently identifying a program as ineffective, but necessarily increases the probability of failing to detect programs that, in fact, deviate from the mean. Given the importance of such decisions, we plan to present results for a range of confidence intervals and provide some insight into how such decisions affect the number of programs that are observed to differ from the mean program in Massachusetts.

Value-added models like Eq. (1) assume that classroom assignments do not vary systematically by unobserved factors affecting student achievement. Recent analyses of teacher turnover by Chetty et al. (2014a) and Jackson (2014) suggest that these models provide unbiased estimates of teachers' contributions to student achievement in the primary and secondary settings, respectively. Koedel et al. (2015) demonstrate that failing to account for the nested structure of student test score data produces misleading estimates of both the variability in effects across institutions and the precision of estimates of individual institution effects. In each of our analyses, we will be careful to account for the fact that repeated measurements of individual teachers' productivity do not provide independent observations on institution or program effects.

The model specified in Eq. (1) additionally allows us to test student growth percentiles as a measure of educator program effectiveness. By replacing student achievement  $Y$  with the student growth percentile and omitting the student covariates and school fixed effects, we can estimate the mean growth percentile by program and compare these estimates to the value-added results that emerge from the full estimation of Eq. (1).<sup>3</sup> Next, we can quantify the importance of the additional controls included in value-added models by assessing how the addition of the omitted student, classroom, and school variables in Eq. (1) affects the estimates of program effectiveness (Gelbach, 2016). Given that omitting these controls may be more consequential for programs whose teachers serve high-needs populations, this approach will allow us to assess how program

---

<sup>3</sup> The program median growth percentile can be computed similarly.

effectiveness estimates change with the addition of student variables for each individual program separately.

### ***Educator Preparation and Performance Evaluations***

The analysis of educator effectiveness on district evaluations will follow a similar approach to the analysis of student achievement gains. We will therefore estimate regression models similar to Eq. (1) where we replace the student achievement outcomes with teachers' summative performance rating, but retain the student and classroom variables and school fixed effects. In order to account for the discrete nature of the performance rating data, we will estimate both linear models, as in the student achievement case, and generalized nonlinear extensions of Eq. (1). Emulating the covariate adjustment approach in Eq. (1) reduces the impact of two sources of potential bias in the estimation of teacher preparation effects. First, there is some evidence that teacher evaluations are correlated with student characteristics, which may indicate that some teachers receive more favorable evaluations because of the students they teach (Whitehurst et al., 2014). Second, even if classroom composition effects are unimportant, districts may employ different standards when assigning summative ratings. For instance, a district may assign fewer teachers to the exemplary performance rating than other districts because of its application of the standards and not because it has a less effective pool of teachers. If some institutions send teachers disproportionately to districts with more or less stringent evaluation standards, then comparisons of teachers across districts may misstate the influence of educator preparation. Consequently, graduates of certain programs may receive lower average performance ratings because they are more likely to work in particular districts and not because they are less effective teachers. As is the case with the student achievement models, including school effects removes mean differences in evaluation standards across districts and ensures that comparisons are only made with other teachers in the same schools.

As with the results on educator effectiveness derived from value-added modeling and student growth percentiles, we plan to provide estimates of the variability in program effectiveness as well as estimates of individual program effects and their uncertainty. We will replicate the analysis of the student achievement data using a variety of decisions about the proper confidence interval for assessing whether individual program estimates are statistically distinguishable from the mean among all programs.

### ***Educator Preparation and Retention***

The final research question asks whether there are differences in teacher retention across licensure pathways and educator preparation programs. Our analysis will focus on the probability that a teacher will depart the Massachusetts public school system in a given year. In particular, we will estimate duration models of teachers' careers in public schools that include the indicators of teachers' preparation included in Eq. (1). We can summarize this approach with the following empirical model:

$$\Pr(Quit_{jst} = 1 \mid T \geq t) = f(Exp_{jt}\lambda + X_{jst}\beta + P_{jt}\delta). \quad (2)$$

In Eq. (2),  $j$  indicates teacher,  $s$  indicates school, and  $t$  indicates the year of service. We estimate the likelihood that a teacher quits in year  $t$  given that she has not yet left the school system; that is, her total tenure in Massachusetts public schools ( $T$ ) is at least  $t$  years long. As before, we

include teacher and school characteristics ( $X$ ) and indicators of teacher preparation ( $P$ ). Models like Eq. (2) have been used to study the career pathways of teachers with different preparation backgrounds and the influence of a number of school and district characteristics on teacher retention (Clotfelter et al., 2008; Goldhaber and Cowan, 2014; Goldhaber et al., 2011; Imazeki, 2005; Kane et al. 2008). The model in Eq. (2) implicitly assumes that school-level factors omitted from  $X$  that affect teacher attrition are not related to educators' preparation background. As this may not be a reasonable assumption, we will follow the approach of Goldhaber and Cowan (2014), who relax the assumption that school-level unobservables are unrelated to the composition of the teaching staff.

## Timeline and Deliverables

During the first year of this collaboration with ESE, ALDER staff will conduct research into the three questions identified in this research plan. The ultimate objective of the research activity is a written report containing empirical evidence on the variability in teacher workforce outcomes across licensure pathways, educator preparation institutions, and educator preparation programs in Massachusetts.

In Exhibit 1, we provide a timeline of the key research activities. We anticipate that data cleaning and preparation will proceed through May 2016. Following the completion of the data cleaning activities, ALDER researchers will proceed to the main analysis. We expect to provide preliminary results to ESE staff by September 1, 2016 and complete the written report on educator preparation effectiveness by November 1, 2016.

**Exhibit 1. Timeline of Research Activities**

Task Name		Start	Finish
<b>Analysis of Teacher Preparation Programs and Institutions</b>			
1.1	Obtain and clean Massachusetts data for analysis of state preparation programs	12/01/15	05/01/16
1.2	Progress Check-In with ESE: Data		03/01/16
1.3	Complete analysis of the variation in teacher effectiveness across state preparation programs	05/01/16	09/01/16
1.4	Progress Check-In with ESE: Preliminary results		09/01/16
1.5	Finalize results and draft report on program effectiveness		11/01/16

Following the completion of the year 1 research activities, ALDER staff will work with ESE to identify additional research on alternative licensure or preparation models for the second year of the grant.

## References

Bastian, K. C., & Henry, G. T. (2015). Teachers without Borders: Consequences of Teacher Labor Force Mobility. *Educational Evaluation and Policy Analysis*, 37(2), 163–183.



- Betebenner, D. (2009). Norm- and Criterion-Referenced Student Growth. *Educational Measurement: Issues and Practice*, 28(4), 42–51.
- Blazar, D. (2015). Effective Teaching in Elementary Mathematics: Identifying Classroom Practices that Support Student Achievement. *Economics of Education Review*, 48, 16–29.
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). How Changes in Entry Requirements Alter the Teacher Workforce and Affect Student Achievement. *Education Finance and Policy*, 1(2), 176–216.
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2008). *Who Leaves? Teacher Attrition and Student Achievement* (NBER Working Paper No. 14022). Cambridge, MA: National Bureau of Economic Research.
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher Preparation and Student Achievement. *Educational Evaluation and Policy Analysis*, 31(4), 416–440.
- Campbell, S. & Ronfeldt, M. (2015, November). Evaluating Teacher Preparation Using Observational Data. Paper Presented at the 37th Annual Fall Research Conference of the Association for Public Policy Analysis and Management, Miami, FL.
- Cantrell, S., Fullerton, J., Kane, T. J., & Staiger, D. O. (2008). *National Board Certification and Teacher Effectiveness: Evidence from a Random Assignment Experiment* (NBER Working Paper No. 14022). Cambridge, MA: National Bureau of Economic Research.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *The American Economic Review*, 104(9), 2593–2632.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *The American Economic Review*, 104(9), 2633–2679.
- Clark, M. A., Chiang, H. S., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). *The Effectiveness of Secondary Math Teachers from Teach For America and the Teaching Fellows Programs* (No. NCEE 2013-4015). Washington, D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Clotfelter, C. T., Glennie, E., Ladd, H., & Vigdor, J. (2008). Would Higher Salaries Keep Teachers in High-Poverty Schools? Evidence from a Policy Intervention in North Carolina. *Journal of Public Economics*, 92(5-6), 1352–1370.
- Clotfelter, C. T., Ladd, H., & Vigdor, J. (2007). Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects. *Economics of Education Review*, 26(6), 673–682.
- Clotfelter, C. T., Ladd, H., & Vigdor, J. (2010). Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects. *The Journal of Human Resources*, 45(3), 655–681.

- Cowan, J., & Goldhaber, D. (2015). *National Board Certification and Teacher Effectiveness: Evidence from Washington* (CEDR Working Paper No. 2015-3). Seattle, WA: Center for Education Data And Research.
- Ehlert, M., Koedel, C., Parsons, E., & Podgursky, M. J. (2013). The Sensitivity of Value-Added Estimates to Specification Adjustments: Evidence from School- and Teacher-Level Models in Missouri. *Statistics and Public Policy*, 1(1), 19–27.
- Gelbach, J.B. (2016). When Do Covariates Matter? And which Ones, and How Much? *Journal of Labor Economics*, 34(2), 509–543.
- Glazerman, S., Mayer, D., & Decker, P. (2006). Alternative Routes To Teaching: The Impacts Of Teach For America On Student Achievement And Other Outcomes. *Journal Of Policy Analysis And Management*, 25(1), 75–96.
- Goldhaber, D. D., Brewer, D. J., & Anderson, D. J. (1999). A Three-Way Error Components Analysis of Educational Productivity. *Education Economics*, 7(3), 199–208.
- Goldhaber, D., & Cowan, J. (2014). Excavating The Teacher Pipeline: Teacher Preparation Programs And Teacher Attrition. *Journal of Teacher Education*, 65(5), 449–462.
- Goldhaber, D., Gross, B., & Player, D. (2011). Teacher Career Paths, Teacher Quality, and Persistence in the Classroom: Are Public Schools Keeping Their Best? *Journal of Policy Analysis And Management*, 30(1), 57–87.
- Goldhaber, D., & Hansen, M. (2013). Is it Just a Bad Class? Assessing the Long-term Stability of Estimated Teacher Performance. *Economica*, 80(319), 589–612.
- Goldhaber, D., Liddle, S., & Theobald, R. (2013). The Gateway to the Profession: Assessing Teacher Preparation Programs Based on Student Achievement. *Economics of Education Review*, 34, 29–44.
- Goldhaber, D., Walch, J., & Gabele, B. (2013). Does the Model Matter? Exploring the Relationship between Different Student Achievement-based Teacher Assessments. *Statistics and Public Policy*, 1(1), 28–39.
- Goldring, R., Taie, S., & Riddles, M. (2014). *Teacher Attrition and Mobility: Results From The 2012-13 Teacher Follow-Up Survey* (No. NCES 2014-077). Washington, D.C.: National Center For Education Statistics, U.S. Department Of Education.
- Grossman, P., Loeb, S., Cohen, J., & Wyckoff, J. (2013). Measure for Measure: The Relationship between Measures of Instructional Practice in Middle School English Language Arts and Teachers' Value-Added Scores. *American Journal of Education*, 119(3), 445–470.
- Harris, D. N., & Sass, T. R. (2014). Skills, Productivity and the Evaluation of Teacher Performance. *Economics Of Education Review*, 40, 183–204.
- Henry, G. T., Purtell, K. M., Bastian, K. C., Fortner, C. K., Thompson, C. L., Campbell, S. L., & Patterson, K. M. (2014). The Effects of Teacher Entry Portals on Student Achievement. *Journal of Teacher Education*, 65(1), 7–23.

- Imazeki, J. (2005). Teacher Salaries and Teacher Attrition. *Economics Of Education Review*, 24(4), 431–449.
- Jacob, B. A., & Lefgren, L. (2008). Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education. *Journal of Labor Economics*, 26(1), 101–135.
- Jackson, C. K. (2014). Teacher Quality at the High School Level: The Importance of Accounting for Tracks. *Journal of Labor Economics*, 32(4), 645–684.
- Kane, T. J., Mccaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have We Identified Effective Teachers?* Seattle, WA: Bill And Melinda Gates Foundation.
- Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What Does Certification Tell Us about Teacher Effectiveness? Evidence from New York City. *Economics of Education Review*, 27(6), 615–631.
- Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying Effective Classroom Practices Using Student Achievement Data. *The Journal of Human Resources*, 46(3), 587–613.
- Kinsler, J. (2012). Beyond Levels and Growth: Estimating Teacher Value-Added and Its Persistence. *The Journal Of Human Resources*, 47(3), 722–753.
- Koedel, C., Parsons, E., Podgursky, M., & Ehlert, M. (2015). Teacher Preparation Programs and Teacher Quality: Are There Real Differences across Programs? *Education Finance And Policy*, 10(4), 508–534.
- Lincove, J. A., Osborne, C., Dillon, A., & Mills, N. (2014). The Politics and Statistics of Value-Added Modeling for Accountability of Teacher Preparation Programs. *Journal of Teacher Education*, 65(1), 24–38.
- Massachusetts Department Of Elementary And Secondary Education. (2014). *Educator Evaluation Data: Student Growth Percentiles, Race/Ethnicity, Gender, and Professional Teaching Status*. Malden, MA: Massachusetts Department of Elementary and Secondary Education.
- McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The Intertemporal Variability of Teacher Effect Estimates. *Education Finance and Policy*, 4(4), 572–606.
- Mihaly, K., McCaffrey, D., Sass, T. R., & Lockwood, J. R. (2013). Where You Come from Or Where You Go? Distinguishing Between School Quality and the Effectiveness of Teacher Preparation Program Graduates. *Education Finance and Policy*, 8(4), 459–493.
- Papay, J. P. (2011). Different Tests, Different Answers: The Stability of Teacher Value-Added Estimates across Outcome Measures. *American Educational Research Journal*, 48(1), 163–193.
- Papay, J. P., West, M. R., Fullerton, J. B., & Kane, T. J. (2012). Does an Urban Teacher Residency Increase Student Achievement? Early Evidence from Boston. *Educational Evaluation and Policy Analysis*, 34(4), 413–434.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2), 417–458.

- Rockoff, J. E. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *The American Economic Review*, 94(2), 247–252.
- Ronfeldt, M. (2012). Where Should Student Teachers Learn to Teach?: Effects of Field Placement School Characteristics on Teacher Retention and Effectiveness. *Educational Evaluation and Policy Analysis*, 34(1), 3–26.
- Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How Teacher Turnover Harms Student Achievement. *American Educational Research Journal*, 50(1), 4–36.
- Ronfeldt, M., Schwartz, N., & Jacob, B. A. (2014). Does Pre-Service Preparation Matter? Examining an Old Question in New Ways. *Teachers College Record*, 116(10), 1–46.
- Sass, T. R. (2015). Licensure and Worker Quality: A Comparison of Alternative Routes to Teaching. *Journal Of Law And Economics*, 58(1), 1–35.
- Sass, T. R., Hannaway, J., Xu, Z., Figlio, D. N., & Feng, L. (2012). Value Added of Teachers in High-Poverty Schools and Lower Poverty Schools. *Journal of Urban Economics*, 72(2-3), 104–122.
- U.S. Department Of Education. (2013). *Preparing and Credentialing the Nation's Teachers: The Secretary's Ninth Report on Teacher Quality*. Washington, D.C.: Office Of Postsecondary Education, U.S. Department of Education.
- Whitehurst, G. J., Chingos, M. M., & Lindquist, K. M. (2014). *Evaluating Teachers with Classroom Observations*. Washington, D.C.: Brown Center on Education Policy, Brookings Institution.
- Xu, Z., Hannaway, J., & Taylor, C. (2011). Making a Difference? The Effects of Teach For America in High School. *Journal of Policy Analysis and Management*, 30(3), 447–469.



## ABOUT AMERICAN INSTITUTES FOR RESEARCH

Established in 1946, with headquarters in Washington, D.C., American Institutes for Research (AIR) is an independent, nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance both domestically and internationally. As one of the largest behavioral and social science research organizations in the world, AIR is committed to empowering communities and institutions with innovative solutions to the most critical challenges in education, health, workforce, and international development.



AMERICAN INSTITUTES FOR RESEARCH®

1000 Thomas Jefferson Street NW  
Washington, DC 20007-3835  
202.403.5000

[www.air.org](http://www.air.org)

## LOCATIONS

### Domestic

Washington, D.C.  
Atlanta, GA  
Austin, TX  
Baltimore, MD  
Cayce, SC  
Chapel Hill, NC  
Chicago, IL  
Columbus, OH  
Frederick, MD  
Honolulu, HI  
Indianapolis, IN  
Metairie, LA  
Naperville, IL  
New York, NY  
Rockville, MD  
Sacramento, CA  
San Mateo, CA  
Waltham, MA

### International

Egypt  
Honduras  
Ivory Coast  
Kyrgyzstan  
Liberia  
Tajikistan  
Zambia

*Making Research Relevant*