10-4-2016

Kids vs. Adults: Using Observations and Student Surveys to Evaluate the Arkansas Teacher Corps

Elise Swanson
University of Arkansas, Fayetteville, eswanson@uark.edu

Gary W. Ritter
University of Arkansas, Fayetteville

Follow this and additional works at: https://scholarworks.uark.edu/edrepub

Part of the Educational Assessment, Evaluation, and Research Commons, Educational Leadership Commons, and the Other Educational Administration and Supervision Commons

WORKING PAPER SERIES

Kids vs. Adults:

Using Observations and Student Surveys to Evaluate the Arkansas Teacher Corps

Elise Swanson and Gary Ritter

4 October 2016

EDRE WP 2016-13

Abstract: This study is an evaluation of the Arkansas Teacher Corps, an alternative teacher certification program that places teachers in high needs schools in rural, southern Arkansas. This evaluation focuses on an intermediate goal of the organization—effective teaching practices— and uses a matching strategy to determine the effectiveness of Arkansas Teacher Corps Fellows. Data comes from third party observations and student surveys. ATC teachers are rated significantly higher on constructs of content knowledge, teacher-student relationships in class, and teacher-student relationships out of class by students. There are no significant differences between ATC and non-ATC teachers noted by observers or on other constructs measured by student surveys.

KEYWORDS: Alternative certification, high needs schools, teacher observations, student surveys, matching

I.   Introduction

No serious person interested in education disputes the importance of teachers in shaping students' learning trajectory. Teachers are schools' largest expense, and have the greatest impact on student achievement (Darling-Hammond 2000). Excellent teachers not only contribute to student learning, but also impact their students far beyond the classroom as they continue their education and live and work as adults (Chetty, Friedman, and Rockoff 2011). While everyone recognizes the importance of teachers and of having a high quality teacher in front of every student, schools have widely differing access to the supply of highly qualified teachers, with disadvantaged schools bearing the brunt of the limited labor pool (Ingersoll 2002). Teacher quality is essential for student growth, and unevenly distributed across schools. Thus, policymakers have been working to develop strategies to enhance teacher supply and quality in areas where it is lacking, such as low income urban and rural areas.

In response to these local challenges, states have devised strategies to increase the labor pool of teachers in these areas. One such strategy is alternative certification. Alternative certification programs are predicated on the assumption that individuals with a desire to teach can be effective in the classroom without having gone through a traditional teacher preparation program, which is generally thought of as a four-year undergraduate course of study with an unpaid student teaching experience as part of their preparation. Alternatively certified teachers typically have academic backgrounds in the subjects they teach, meaning they may actually have greater content knowledge than traditionally trained teachers. Further, the sole reliance on traditionally certified teachers limits the teacher labor pool only to those individuals who knew they wanted to teach when beginning their undergraduate career, or who can take four to five years out of the labor force to go back to school for this training. This places severe limitations

on the teacher supply pool, and does not necessarily ensure that only highly capable teachers are entering the classroom. Teacher labor shortages are often concentrated in particular subjects—notably math and science—and in disadvantaged areas, requiring schools serving disadvantaged students to make do with whomever is left after the affluent schools have had their pick of available applicants.

Not only do some argument that alternative certification is necessary because of labor shortages; still others believe that, regardless of teacher preparation, teachers experience the most development once they are in the classroom. Murnane and Phillips (1981) found that teacher effectiveness increases most dramatically during a teacher's first years in the classroom; similarly, Pigge (1978) found that most teachers believed their most useful development happened on the job (in King Rice 2003). This research demonstrates that the traditional four to five year preparation program may not be fully preparing teachers for the classroom while still creating an unnecessary barrier for entry into the field. Alternative certification programs accept individuals who meet program requirements, but who generally do not have an undergraduate degree in education. Alternative certification programs are thus able to attract professionals looking to switch careers, recent college graduates who majored in subjects other than education, and other non-traditional candidates. By expanding the teaching force, alternative certification programs are a means to provide greater choice—and higher quality choices—to districts facing teacher shortages and who would otherwise be forced to fill positions with a rotating cast of long-term substitutes or underqualified candidates they would have preferred not to hire. Such is the situation in rural, southern Arkansas, where schools face significant shortages of available teachers, and low-income students of color are disproportionately affected by this shortfall.

Teach for America (TFA) is the most widely recognized alternative certification program in the country. TFA is a highly competitive and selective alternative certification program which recruits and places CMs nationally in disadvantaged schools. By 2016 over 50,000 Corps Members (CMs) had taught in struggling schools across the country (Teach for America 2016). TFA recruits recent college graduates, as well as career-switchers, places CMs in a region, and arranges interviews between CMs and school districts. Corps Members are hired and paid directly by the district. CMs participate in a five week summer training institute, where they participate in development sessions focused on pedagogical techniques, classroom management strategies, content, and diversity, equity, and inclusiveness in education. TFA also partners with local districts to run school-wide summer school programs for students. CMs are responsible for teaching summer school classes in subjects and grade levels that roughly correspond to what they will be teaching at their placement school. Corps Members commit to two years with Teach for America, at the end of which they have a full teaching license in their placement state. Throughout their two year commitment, CMs are regularly observed and coached on their teaching in the classroom, and participate in ongoing professional development sessions facilitated by TFA.

A major criticism of TFA has been that it only requires a two year commitment, which critics see as an opportunity for ambitious college graduates to pad their resume with a two year service commitment before leaving the teaching profession for a more lucrative or competitive field. By encouraging high turnover, critics contend, TFA is hurting the schools and students it purports to serve, and ensures that disadvantaged students are perpetually taught by a rotating cast of inexperienced, uncommitted teachers (Labaree 2010). On a related note, critics argue that because TFA essentially imports college graduates to teach in high-needs area, it is encouraging

a sort of prolonged "voluntourism"(McGloin and Georgeou 2016), where CMs are not fully invested in their communities or appreciative of the strengths and assets of the communities, students, and families with whom they interact, and instead view themselves as outside saviors of children in need (Anderson 2013). These problematic mindsets, according to critics, both decreases retention among CMs who view teaching as a time-limited, feel-good exercise, and reduces communities' sense of agency and empowerment in overcoming the challenges they face.

In response to these criticisms, TFA has focused on improving retention among alumni and increasing programming for CMs revolving around community engagement, asset-based thinking, and diversity, equity, and inclusiveness to ensure that CMs are fully invested in their communities, see themselves as partners of those in the community in which they teach, and are empowering their students to make changes in their communities as well.

While there are vocal critics of TFA and other alternative certification, these criticisms are often not voiced by school leaders whose job it is to place high quality teachers in front of students each year. Indeed, many school leaders are supportive of such programs and programs based on the TFA model have been developed in a few areas across the country. For example, the New York City Teaching Fellows program was launched in 2000 to address teacher shortages in New York City public schools, with an emphasis on increasing the diversity of the teaching force. Similarly, the Mississippi Teacher Corps aims to staff high needs schools in Mississippi while enrolling its teachers in a master's program in Curriculum and Instruction. Because of the potential benefits of localized selective alternative programs in high needs contexts, the Arkansas Teacher Corps was developed to serve disadvantaged schools in rural Arkansas.

The Arkansas Teacher Corps (ATC) is an alternative licensure path operating within the state of Arkansas with the goal of providing high quality teachers to high-needs schools. ATC is modeled after TFA, but with two important features designed to address the major criticisms of TFA. First, ATC requires a three year commitment, and pays Fellows an additional $15,000 stipend over those three years in an effort to increase recruitment and long-term retention. Secondly, ATC recruits individuals who have specific ties to the Arkansas community and context—recent graduates from Arkansas universities, Arkansas natives who attended nearby schools, and professionals currently working in Arkansas who want to switch into teaching. These intentional design features should increase retention and increase Fellows' sense that they are working with their community to address education inequities, rather than promote the view that they are coming in from the outside with all the answers.

ATC first placed teachers in school in the 2013-14 school year. The recently completed 2015-16 school year was the program's third year of placing teachers in districts. In the 2015-16 school year, ATC Fellows taught in 21 high-needs schools in the state, teaching subjects as diverse as elementary art and high school chemistry. The program has not yet been rigorously evaluated on the effectiveness of its teachers, and is set to expand in the 2016-17 school year, with the number of incoming Fellows practically doubling the number of teachers from the three prior cohorts.

This evaluation is designed to determine the effectiveness of the first three cohorts of ATC Fellows and point to areas of improvement for the program. In a break from previous evaluations' reliance on student standardized test scores as a measure of effectiveness, we rely on classroom observations and student surveys, enabling us to capture important dimensions of teaching that may go uncaptured by standardized assessments; moreover, using this evaluation

strategy allows us to measure the effectiveness of teachers in subjects not connected to standardized state assessments, such as art and music and even such classes as high school calculus or 12th grade English. Thus, this evaluation strategy is also superior for practical reasons; because so many of the ATC teachers do not teach tested subjects, we would not have the statistical power needed to conduct a test-based value-added analysis of ATC Fellows' effectiveness.

The rest of this paper will proceed as follows: first, we will review the research on the effectiveness of alternative certification programs, building an evidence base supporting the proposition that the Arkansas Teacher Corps could be successful; next, we dive into the Arkansas context, establishing the need for the Arkansas Teacher and its underlying program theory of change. Moving forward from there, we explain our research design and sample, before discussing our results and concluding. The overarching goal of this paper is to address the following research questions:

1. How do ATC Fellows compare to their peers in using effective classroom practices, as measured by third part observers?

2. How effective do students perceive ATC Fellows to be relative to their peers on the following dimensions of teaching:

   - Overall learning
   - High Expectations
   - Content Knowledge
   - Preparation for Class
   - Relationships in Class
   - Relationships outside Class
   - Behavior Management
   - Class Engagement

II.   Prior Literature: How effective are alternative certification teachers?

Broadly speaking, alternative paths to licensure do not require individuals to have obtained a degree in education in order to earn a teaching certification. Alternative pathways generally place more emphasis on classroom experience and ongoing development, while traditional paths emphasize child development courses and pedagogical theory (Fraser, 2009). There is great variety in the design and reach of alternative licensure programs for teachers in the United States. In a national evaluation of traditional and alternative routes to certification, Constantine et al (2009) found that alternatively certified (AC) teachers received anywhere from 75-795 hours of instruction over the course of their program, while traditionally certified (TC) teachers received anywhere from 240-1,380 hours of instruction over the course of their program, highlighting both the overlap between alternative and traditional certification programs, and the variation between programs under each umbrella term.

Constantine et al also found that AC teachers were more likely to identify as Black, be older, and have children than TC teachers, but there were no differences between AC and TC teachers in terms of average SAT score, highest degree earned, or whether they were currently taking courses. In addition to the national 2009 review, there are several rigorous evaluations of the effectiveness of AC programs, mostly of selective AC pathways. Generally, evaluations of AC programs use randomized control trial or quasi-experimental design methods; we organize our review of the literature along these broad categories. In this section, we will summarize the evidence from the literature on the effectiveness of alternatively certified teachers[1], and discuss

---

[1]An initial search of EbscoHost using the terms "alt* cert* and education or teach*) returned 2,890 initial articles. We then limited the results to articles with full text available published in 1990 or later (Teach for America, the most widely-known selective alternative certification program, was started in 1989), reducing the number of articles to 1,373. Narrowing the search to focus on empirical studies by adding the search term "effect*" reduced the list to 350 articles. We then retained only include journal articles, academic journal articles, and reports found in education or economic databases (ERIC, Academic Search Complete, MasterFILE Premier, or EconLit). This reduced our

the methods used in prior studies as guidance for our own analysis. We will conclude this section by showing that our evaluation of the Arkansas Teacher Corps can make a valuable contribution to the overall literature.

In general, AC programs can be more or less selective, recruiting recent college graduates from highly selective universities as do Teach for America and the New York Teaching Fellows, or attracting career switchers or recruiting graduates from less-selective universities. Given the variation in both alternative certification (AC) and traditional certification (TC) programs, we would expect to see variation in the findings of research looking at the effectiveness of alternative certification programs. In line with this hypothesis, studies of selective programs have generally produced different results than studies of non-selective alternative certification programs. Selective alternative certification programs generally recruit high achieving individuals (measured by GPA, past test scores, leadership activities, and in-person interviews). Often, selective programs will also target recruitment efforts at prestigious universities to attract high achieving, ambitious individuals to the classroom.

In contrast, non-selective programs have low admission criteria, generally requiring only a bachelor's degree and a minimal undergraduate GPA. Non-selective programs do not select participants based on past evidence of their achievement, and typically do not focus their recruitment efforts on individuals who have found success in other areas, whether academic or professional. Constantine et al (2009) conducted a random assignment study in which students were randomly assigned to teachers in 63 schools across the country to examine differences in

---

search results to 166 titles. We read all 166 titles, saving 23 articles that seemed to be rigorous experimental or quasi-experimental evaluations of alternative certification programs. Abstracts were then read for relevance and methods. Relevant articles were then read in full and included if they were primary studies evaluating alternative certification programs with a valid comparison group).

effectiveness between traditionally certified teachers and alternatively certified teachers from non-selective programs, finding no significant differences in student achievement on math or literacy exams between the two groups of teachers. Sass (2011) used quasi-experimental methods, controlling for school-level fixed effects as well as teacher and student characteristics, to examine the relative effectiveness of alternatively certification teachers from selective and non-selective programs and traditionally certified teachers, finding null effects for alternatively certified teachers from non-selective programs and positive effects for alternatively certified teachers from selective programs.

Of course, the best known selective AC program is Teach for America, and there have been several rigorous studies of the effectiveness of this program. For example, Decker et al (2004), Glazerman et al (2006) and Chiang et al (2014) used experimental methods to evaluate TFA, and found positive effects in math but null effects in reading. Using the same dataset as Decker et al, Antecol et al (2013) confirmed the positive effects for TFA in math, but did not evaluate literacy outcomes. The preceding four studies exploited random assignment of students to teachers in multiple schools across the country to identify the impact of Teach for America on student achievement, lending confidence to their findings that the program was generally effective in improving math outcomes, and having no significant impact on reading outcomes.

While random assignment studies are the gold standard of social science research, rigorous quasi-experimental studies have also been conducted to evaluate selective alternative certification programs. Kane et al (2006) use value-added measures to evaluate both Teach for America and the New York City Teaching Fellows as compared to traditionally trained teachers with the same number of years of experience, finding positive effects for TFA teachers in math and negative effects for NYC Teaching Fellows in literacy. Boyd et al (2005) also used value-

added measures to evaluate Teach for America and the New York City Teaching Fellows as compared to traditionally trained teachers with the same level of experience and as compared to all traditionally trained teachers, but found negative effects for Teach for America in literacy and null effects in math, and negative effects for NYC Teaching Fellows in math and literacy.

Darling-Hammond et al (2005) and Raymond et al (2001) used fixed effects to evaluate Teach for America teachers relative to all other teachers and beginning teachers in the Houston School District. Darling-Hammond et al found positive effects of Teach for America teachers in math and null effects in literacy when students were assessed using the Texas state standardized assessment; however, they found negative effects of Teach for America teachers in both math and reading as measured by the SAT-9 and a Spanish language test for math and reading. In contrast, Raymond et al (2001) found positive effects of Teach for America teachers in math, and null effects in literacy; Raymond et al also concluded that the distribution of quality among Teach for America teachers was higher than among non-Teach for America teachers. Xu and Hannaway (2011) used student fixed effects to evaluate Teach for America high school teachers compared to all other North Carolina teachers, finding evidence of positive effects in reading, and null to positive effects in math. Penner (2014) compared TFA to non-TFA teachers within the same school in North Carolina and found evidence of positive effects on graduation for students assigned to TFA classrooms.

Finally, researchers have also detected program effects by using matching strategies. Laczo-Kerr et al (2002) matched alternatively certified teachers to certified teachers with similar experience to evaluate a mix of alternative certification programs, including Teach for America, and found negative effects in math and literacy for one year of their evaluation and null effects in the other year evaluated. Clark et al (2013) also utilized classroom-level matching across

multiple schools and states to evaluate Teach for America and the New York Teach Fellows relative to traditionally certified and non-selective alternatively certified teachers. Clark et al (2013) found positive math impacts for Teach for America teachers, and null effects for New York Teaching Fellows teachers in math.

Previous research on the effectiveness of alternative certification teachers typically examined math and literacy outcomes, and only focused on teachers in those subjects. Prior studies have found evidence of positive impacts of alternatively certified teachers from selective programs, and null effects for teachers from non-selective alternative certification program. Most of these evaluations of selective programs, such as Teach for America and the New York Teaching Fellows, occurred in urban areas. Fewer – if any – studies have been conducted of selective programs oriented towards serving high-needs schools in a rural context. This analysis fills that gap by evaluating a relatively new—three years old—program that places teachers in disadvantaged schools in rural southern Arkansas.

Moreover, ATC's design and context differs from TFA and other selective programs in important ways and thus the program merits an evaluation in its own right. In contrast to TFA's national scope, ATC is a highly localized program focused on rural disadvantaged schools. ATC's cost per Fellows is much lower than TFA's cost per CM, and a finding of ATC's effectiveness could provide a model for low-cost statewide AC programs. Further, the majority of past evaluations have focused on elementary and some middle school teachers, while many alternatively certified teachers work at the high school level; the vast majority of ATC teachers teach high school students In Arkansas, the Arkansas Department of Education (ADE) projected a 10.44% surplus of pre-K through 4th grade teachers and a 6.87% surplus of 5th-8th grade teachers in December 2015, but shortages in ten subject areas for high school teachers. There is

thus a clear need to evaluate the effectiveness of high school alternatively certified teachers in rural schools. This evaluation addresses that need.

III.   Arkansas Context

Arkansas is a mostly rural, mid-southern state with 476,049 K-12 students in 259 school districts in the 2015-16 school year. Although graduation rates are high—87% across the state as a whole—achievement remains low, with only 43% of students meeting the math readiness benchmark on the ACT Aspire assessment in 2014-15, and 68% of students meeting the English readiness benchmark on the ACT Aspire (ADE 2016a). Arkansas was also below the national average on the National Assessment of Educational Progress (NAEP) in fourth grade math and reading in 2015 (NCES 2016). Over 62% of the state's K-12 students qualified for free or reduced price lunch in 2015-16, and almost 40% of Arkansas students are minorities (ADE 2016g; f). The Arkansas Department of Education declared critical licensure shortage areas in 10 subject areas for the 2016-17 school year, including art, mathematics, and physical science (ADE 2016c). In addition to shortages in certain licensure areas, Arkansas public schools must also contend with the issue of teacher attrition. Between the 2006-07 and 2014-15 school years, between 6.9% and 23.6% of new teachers left the profession after just one year in the classroom, and 31.9% to 40.4% of teachers beginning in the 2006-07 through 2010-11 school years left within 5 years of entering the classroom (ADE 2016b).

There are currently five alternative routes to teacher licensure in the state: the Arkansas Professional Pathway to Educator Licensure (APPEL), Non-traditional MAT, MED, MTLL through Selected AR Colleges and Universities, Teach for America (TFA), Arkansas Teacher Corps (ATC), and a Provisional Professional Teaching License (ADE 2016e). Of these,

nontraditional certification programs offered through universities prepare the largest number of teachers. In the 2014-15 school year, 831 individuals were enrolled in alternative certification programs offered through universities; in that same year, APPEL enrolled 364 individuals, TFA had 110 Corps Members, and 37 Fellows were in ATC (ADE 2016b). In 2014-15, 618 teachers completed alternative certification programs, joining the 1,559 teachers who completed traditional certification programs. Interestingly, while the number of people enrolled in traditional certification programs decreased from 2010 to 2016 (from 7,067 to 2,053), the number of completers from traditional certification programs increased from 2011 to 2015 (from 1,470 to 2,177). From 2010 to 2016, the number of people enrolled in alternative certification programs rose from 1,188 to 1,342, and the number of nontraditional completers rose from 547 to 618 from 2011 to 2015 (ADE 2016b).

The Arkansas Teacher Corps is a selective alternative certification program that recruits high-achieving, committed individuals who have lived, worked, or studied in Arkansas to commit to teaching for three years in a high-needs school in the state. Only three years old, the Arkansas Teacher Corps has yet to be rigorously evaluated on whether the teachers it provides schools are effective in the classroom. The next section will delve deeper into the structure of the Arkansas Teacher Corps and the program's theory of change, while section five will detail our quasi-experimental methods of evaluation and data.

IV. Program Description and Theory of Change

The inaugural cohort of the Arkansas Teacher Corps began teaching in 2013. Twenty-one Fellows began teaching in their placement school in that year; 14 of those initial Fellows were still classroom teachers at their placement school at the time of our study. Twenty Fellows began

teaching in their placement school in 2014; of those, 12 were still classroom teachers at their placement school at the time of our study. In 2015, 22 Fellows began teaching at their placement school; 14 were still classroom teachers at their placement school when we conducted observations.
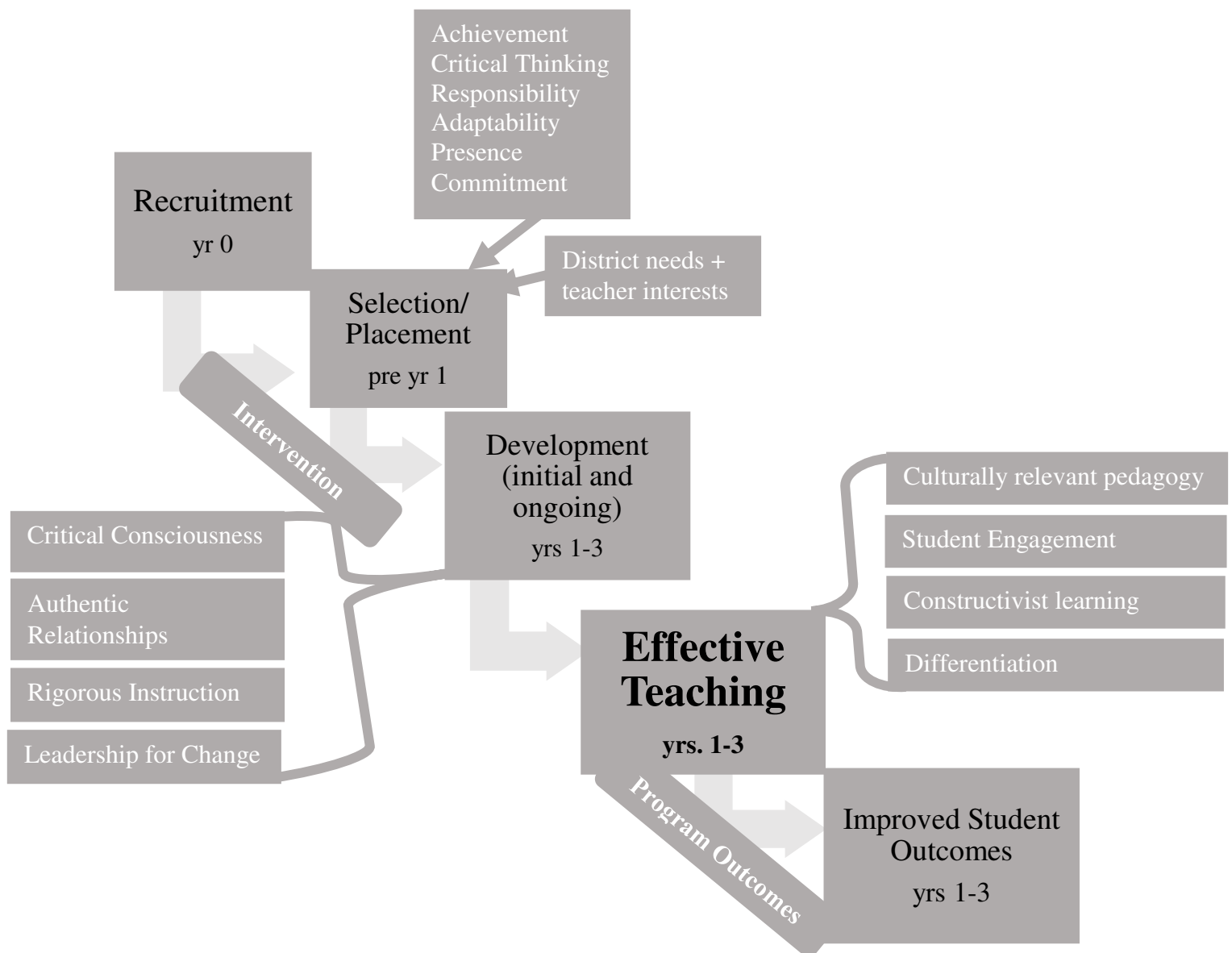
The Arkansas Teacher Corps was founded with the intent of placing highly qualified teachers in underserved schools in the Mississippi Delta region of southern Arkansas, and has gradually increased its focus to include schools in other high-needs areas of the state. The program rates applicants on academic achievement, critical thinking, responsibility, adaptability, commitment, and presence during a rigorous, multistep application process. The average ACT score among the first three cohorts of ATC Fellows was 26 (83rd percentile), and the average undergraduate GPA among Fellows was 3.37. ATC applicants are also required to submit a writing sample along with their application. The program requires a three-year commitment from Fellows and attempts to improve retention in the program and in the classroom by recruiting applicants who have specific connections to the state of Arkansas. Recruitment efforts are focused on university campuses in Arkansas and nearby out-of-state colleges in Tennessee, Louisiana, and Mississippi. In addition, recruitment efforts are also aimed at career changers— individuals who have been working the professional world and want to make the switch to teaching. As reference, CMs in the selective Teach for America program have an average 3.42 undergraduate GPA and are typically recruited from prestigious universities. ATC is much more selective than the Arkansas Professional Pathway for Educator Licensure (APPEL) program, which requires only a bachelor's degree and a 2.70 cumulative undergraduate GPA, and does not require applicants participate in a similarly rigorous application and interview process.

Over the past four years, ATC has received 284 official requests from school leaders for

teachers in specific subjects and grades through a survey the program sends out to high-needs districts each year. In addition, school administrators may reach out directly to the program via phone calls, emails, and texts to request teachers, meaning the number of total requests ATC has received over the past four years is likely higher than the 284 official requests. In 2013-14, when ATC first began placing teaching teachers in disadvantaged schools, districts requested 1,696 waivers for teachers to teach without the proper certification. In 2014-15, districts filed 1,613 such licensure waiver requests with the Arkansas Department of Education. In 2015-16, 1,424 waiver requests were filed with the ADE. The Arkansas Teacher Corps emerged as a solution to a clear problem in the Arkansas educational landscape: a shortage of highly qualified teachers leading to limited personnel choices for principals, teachers teaching outside of their licensure, long-term substitutes leading classes, and low student achievement. By placing Fellows in disadvantaged schools in southern Arkansas, ATC is addressing an important need in the state and fulfilling an immediate goal of the program. However, the program's goal is broader than simply placing adults in the classroom. While a crucial goal of the program is to increase the supply of teachers to disadvantaged schools, the program is ultimately concerned with improving student outcomes. The underlying logic model of the organization is presented in Figure 1.

Figure 1: Arkansas Teacher Corps Logic Model

This logic model is complex, but each step follows easily from the last. The process

begins with recruitment, as described above. Program staff work to attract highly competent

individuals with a clear commitment to service and public education in Arkansas. Then,

applicants are put through a rigorous application process, and evaluated along six domains:

Achievement
Critical Thinking
Responsibility
Adaptability
Presence
Commitment

Recruitment
yr 0

District needs +
teacher interests

Selection/
Placement
pre yr 1

Intervention

Development
(initial and
ongoing)
yrs 1-3

Critical Consciousness

Authentic
Relationships

Rigorous Instruction

Leadership for Change

Culturally relevant pedagogy

Student Engagement

Constructivist learning

Differentiation

**Effective
Teaching**
yrs. 1-3

Program Outcomes

Improved Student
Outcomes
yrs 1-3

achievement, critical thinking, adaptability, responsibility, presence, and commitment through

their written, phone, and in-person applications and interviews. If accepted into the program,

Fellows work with ATC placement staff to interview at and be hired directly by districts. Fellows identify the subjects they are interested in teaching, and where they would be willing to teach within the state. Program staff use that information to connect Fellows with districts, based on district requests for teachers in particular subjects. These steps work to achieve ATC's proximal goal of addressing staffing shortages in southern Arkansas.

Next, ATC provides an initial 6-week training and teaching experience for Fellows, in which development staff work closely with Fellows to develop their instructional and classroom management skills, as well as focus on the program's social justice mission of reducing the achievement gap and equitably distributing teaching talent around the state. This development continues throughout Fellows' three years with the program. ATC development staff emphasize four goals of Fellows' development:

1. critical consciousness,

2. authentic and reciprocal relationships,

3. rigorous effective instruction, and

4. leadership for change agency.

To borrow from ATC's internal language, critical consciousness means that "Fellows will be self- and socially aware in social justice, diversity, and equity, recognizing and responding to prejudice, injustice, and discrimination." The goal of authentic and reciprocal relationships refers to relationships between Fellows and individuals in their school, geographic community, and professional communities. ATC envisions rigorous and effective instruction as rooted in "well-informed and ambitious student learning goals," especially for "students typically marginalized in the education system." Finally, ATC development staff emphasize that "Fellows will be an active agent *[sic]* of change by developing an empowered internal locus of control, by

establishing an authentic leadership presence, and by working with community stakeholders." It is clear that the program aims to train teachers who are more than deliverers of classroom instruction.

Through this development, the program expects to see effective teaching in the classroom, which is characterized by unit planning, data-driven planning, culturally-relevant pedagogy, student engagement, constructivist learning, and differentiation. This effective teaching will lead to improved student outcomes by ensuring that students are provided with rigorous instruction and engage deeply with the educational process.

The distal, or ultimate, goal of ATC is improved student learning experience and learning outcomes. In this evaluation, we focus on a more proximal goal: effective teaching by Fellows, as perceived by neutral observers and by students. As stated earlier, many ATC Fellows do not teach a tested subject, precluding the use of value-added measures for a significant number of our already-limited sample. More importantly from a conceptual standpoint, many of the dimensions of teaching we are able to capture by using observations are not measured on a standardized test. For example, using our instruments, we directly measure student-teacher relationships, an important aspect of teaching that provides students with a role model, mentor, and source of letters of recommendation for colleges and jobs. Each of these roles can have a significant impact on student outcomes, from instilling grit to opening doors to opportunity. Standardized tests can only indirectly measure this effect, assuming that such relationships motivate the student to work harder in class, learn more, and perform better on the exam.

ATC is an untested, new teacher preparation program, and effective teaching is an important component of its logic model, and absolutely critical for its ultimate goal of improving student outcomes; thus, focusing our evaluation on this aspect of the program is both warranted

and sensible. The next section discusses how we measured effective teaching.

V.   Methods and Data

The challenges to identifying program impacts in AC programs have been addressed in previous literature in one of three ways: random assignment of students to teachers, using a fixed effects model to control for school or student effects, or creating matched classroom pairs. In this paper, we use a teacher-level matching strategy in a school-level fixed effects framework. Several challenges arose when attempting to evaluate the effectiveness of Arkansas Teacher Corps (ATC) Fellows. First, the entire Corps consisted of 40 active Fellows, teaching in 21 schools across southern Arkansas. This small sample size limits our statistical power and increases the chances of a Type II error, where a true effect goes undetected. Second, ATC Fellows teach a wide variety of subjects and grade levels, many of which are untested. However, as we have previously discussed, the outcomes captured in classroom observations and student surveys address an important proximal goal of the organization: effectiveness of classroom practices, and therefore represent a valid means of assessing the effectiveness of ATC. This section details how we addressed the challenges of identification in our research design and analysis as well as the data we collected in the spring semester of the 2015-16 school year.

The Counterfactual: Within-School Matching Design

We used a matching design to identify the effectiveness of the Arkansas Teacher Corps teachers. Each ATC Fellow was matched with 1-2 teachers within the same school who taught the same/similar subject and grade. One interpretation is that this type of comparison group represents an upper bound of comparative teacher quality. That is, most often, principals request

ATC Fellows because they are unable to fill the position with <u>any</u> teacher. If Fellows were not in these positions, it is likely that principals would be forced to fill the position with a long-term substitute, a teacher with an emergency license, or an applicant they are similarly unenthusiastic about, and who would have been less effective than the typical teachers currently in the school's classrooms. Indeed, if principals were able to simply hire a "typical" teacher, they would likely not contact ATC in the first place. Thus, while we use the "typical" teacher as the counterfactual, we believe that this likely represents an upper bound estimate of the quality of the comparison teacher and accordingly, a lower bound estimate of the impact of ATC.

For each ATC Fellow in a school that agreed to participate in our study (3 schools opted out, excluding 5 ATC Fellows from our analytic sample), we asked the principal for two teachers within the same school who 1) taught the same subject as the Fellow; 2) taught the same grade as the Fellows; and 3) had approximately the same years of experience as the Fellow. Because the Arkansas Teacher Corps mainly places teachers in rural Arkansas, we were not able to find two exact matches for each Fellow, some of whom are the only teacher of their subject in the school. One principal would allow us only to observe one matched comparison teacher per Fellow, while another would only allow us to observe the Fellows, and no matched comparison teachers. Despite this, we were able to observe all 35 Fellows in participating schools (35) and at least one comparison teacher for the 33 matched Fellows. In all, we observed 101 teachers up to 3 times during a single semester. If a teacher was absent on the day of an observation, we attempted to substitute in another teacher in the school, again asking the principal for a recommended "typical" comparison teacher. If necessary, we scheduled a make-up observation for the absent teacher. We average each teacher's ratings across their observation rounds to account for the different number of observations per teacher. All of our analyses presented below use ordinary

least squares regressions, leveraging OLS' small sample properties to avoid bias resulting from our limited sample.

Matching teachers within the same school should reduce bias in our estimates; however, there may still be school-level effects, such as principal leadership, development opportunities, and community support, which impact teacher effectiveness in the schools. For this reason, our preferred specification includes school-level fixed effects.

Sample & Data

Our data are from two main sources: classroom observations (described above) and teacher surveys, through which we collected information on teacher demographics. Our analytic sample can thus only include teachers for whom we have survey and observation data. In total, we sent out 101 teacher surveys via email and fax, and reminded teachers to complete the surveys through their principals, emails, fax, and phone calls to the school office. Of the 101 surveys we sent out, 81 teachers completed the teacher survey, but 3 completed the survey twice, leaving us with 78 unique responses. Our analytic sample when controlling for experience is 78 observations; our analytic sample without controlling for experience is 100 observations.

Table 1 describes the ATC Fellows and their matched comparison teachers.

Table 1: Demographics, ATC and Comparison Teachers

	Observation Data		Student Surveys	
	ATC	Comparison	ATC	Comparison
Sample Characteristics				
Total Number of Teachers	35	66	39	59
Total Number of Schools	21	21	24	20
Teacher Characteristics				
Average Experience	1.1	7.0	1.0	6.4
% Female Teachers	55.0%	66.7%	52.9%	64.4%
School Characteristics				
Average Class Size	15.3	15.8	16.3	16.0
Subject Areas				
Arts	5 (14.3%)	12 (18.2%)	5 (12.8%)	7 (11.9%)
Social Sciences	16 (45.7%)	29 (43.9%)	17 (43.6%)	25 (42.4%)
Hard Sciences	14 (40.0%)	26 (39.4%)	17 (43.6%)	28 (47.5%)
Grade Level				
Elementary School	5 (14.3%)	11 (16.7%)	5 (12.8%)	5 (8.5%)
Middle School	9 (25.7%)	16 (24.2%)	10 (25.6%)	15 (25.4%)
High School	21 (60.0%)	39 (59.1%)	23 (59.0%)	31 (52.5%)

As shown in Table 1, most ATC Fellows teach either the social sciences— English Language Arts, social studies, foreign languages, or business— or hard sciences—math or science. Only five Fellows teach the arts—art, music, or theater. ATC Fellows are overwhelmingly teaching high school, with 21 teachers placed in 9-12th grade classrooms, 9 teachers placed in 6-8th grade classrooms, and 5 teachers placed in K-5 classrooms. The second column of Table 1 shows that, while the sample contained roughly 2 comparison teachers for each ATC teacher, we also had to substitute comparison teachers during some observations due to teacher absences or other classroom irregularities that prevented an observation; there were also ATC Fellows for whom we only had one comparison teacher, and two ATC Fellows (both teaching high school science) for whom we had no matched comparison teacher. About half of

all ATC Fellows are female, while almost 64% of comparison teachers are female; to account for this, we control for gender in all of our models below.

Outcome Measures

We focus in this evaluation on a proximal goal of ATC—teaching effectiveness as perceived by outside observers and students. The program has an explicit theory of what effective teaching looks like, with specific goals for what Fellows should be doing in the classroom with their students. ATC development staff describe effective teaching as encompassing unit planning, data-driven planning, culturally relevant pedagogy, student engagement, constructivist learning, and differentiation. With the exception of Fellows' planning process, we can observe whether and to what extent these practices are taking place in the classroom by observing actual teaching sessions. While recent studies have focused on value-added measures of teacher effectiveness or pairing value-added with classroom observations (as in the Measures of Effective Teaching reports released by the Gates Foundation), relying on observation data is not without precedent. Dewalt and Ball (1987) relied on classroom observations to examine the relationship between teacher preparation program and teaching competence, evaluating teachers on 12 dimensions of competence (in King Rice 2003).

Effectiveness in teaching for this evaluation was measured using a classroom observation instrument based on the Arkansas Teacher Excellence and Support System, which in turn was based on Charlotte Danielson's Framework for Teaching. The Framework for Teaching was used in the Measuring Effective Teaching Project (MET Project) as one of five included observation protocols; that research only included Domains 2 and 3 (Classroom Environment and Instruction) in its evaluations of teachers. In the MET experimental study, researchers found that

teachers' scores on the Framework for Teaching were positively associated with student

achievement (Bill & Melinda Gates Foundation 2012). The report also recommended that

teacher observation scores be averaged across observations; we follow that practice here (Bill &

Melinda Gates Foundation 2012). By focusing specifically on classroom practices, we are able to

capture important nuances in teaching effectiveness that matter for students' experiences and

learning, and which are important goals for the program. We focus on two main aspects of

teaching for this analysis: classroom environment and instruction.

Observers

As noted by The New Teacher Project (2009), most teacher evaluations do not result in

differentiated ratings between teachers. To avoid any potential bias in ratings caused by teacher-

principal relationships, we hired 14 outside observers through the College of Education at the

University of Central Arkansas (UCA) to conduct observations for the study. Observers were

trained on the observation instrument and given the chance to offer suggestions for improvement,

based on their experiences training teacher candidates at UCA and their own professional

backgrounds. We conducted three rounds of observations, coordinating with school

administrators to plan the observations. Teachers were informed about the study, but did not

know the specific dates or times of the observations in advance. We coordinated with schools to

avoid disruptions to the normal teaching day, such as interim or end-of-year testing, assemblies,

and field trips. Because we did not notify teachers in advance of the observations, we did find

some classrooms engaged in class-specific testing, hosting guest speakers, or other irregularities

that limited our observers' ability to evaluate the teacher. There were also other unforeseen

irregularities, such as power outages and flash flooding, which caused our observations to be rescheduled. When necessary, we returned on a different date to make up the observation.

<u>Measuring Effectiveness—Classroom Observations</u>

1. Classroom Environment

Our observation instrument, available in Appendix C, captures information about teaching effectiveness in the domains of Classroom Environment and Instruction. The observation instrument was developed through a review of the Arkansas Teacher Excellence and Support System (TESS), which is based on the Danielson Framework for Teaching (FFT). Detailed descriptions of each potential rating (unsatisfactory to basic) for each subdomain generated potential student and teacher actions that stand as evidence of teachers' proficiency in each area. These descriptions were used to generate a checklist of teacher and student actions that observers looked for during 15 minute segments of the class period. Observers took note as to whether each action took place or not (or was not applicable to the situation) in each 15 minute segment, remaining in the class for the entire period. During a calibration day with our observers, all faculty at the College of Education at the University of Central Arkansas, the teacher and student actions were refined and observers calibrated to correctly and consistently notice and label teacher and student actions, as well give comparable ratings across teachers. In measuring classroom environment, observers made note of such details as whether students were in the appropriate place in accordance with teacher instructions or a clear procedure; whether there are clear indications that the teacher knew his or her students (evidenced by use of names, personalized motivators, relevant examples in explaining content), and whether students appeared to be on task and engaged the majority of the time.

At the end of the period, observers gave teachers a 1-4 (unsatisfactory to distinguished) rating on the five FFT subdomains of classroom environment: creating an environment of respect and rapport, establishing a culture for learning, managing classroom procedures, managing student behavior, and organizing physical space.[2] These ratings were then averaged together to create a composite classroom environment rating for each specific teacher for each specific observation. Each teacher thus received three distinct ratings in classroom environment over the course of the spring semester (one from each observation); these ratings were averaged together to create one single overall rating in classroom environment for each teacher. Ratings were then standardized and are reported in standard deviations below.

2. Instruction

Our observers also rated teachers on the FFT domain of Instruction. The same procedure was followed for developing the specific teacher and student actions noted by observers during each classroom observation, developing a list from the detailed descriptions of ratings for each subdomain, then revising and calibrating the observation instrument with observers. In measuring instruction, observers made note of such practices as whether the teacher explained the purpose of each activity, whether teachers used content-specific vocabulary, and whether students were working in pairs or small groups. At the end of the period, observers again gave teachers a 1-4 rating on the five FFT subdomains of instruction: communicating with students,

---

[2] These domains are the same as those covered in the Arkansas Teacher Excellence Support System (TESS), which the state has adopted for the purpose of teacher evaluation. These domains are thus highly policy relevant and considered important in the teaching process.

using questioning and discussion techniques, engaging students in learning, using assessment in instruction, and demonstrating flexibility and responsiveness.

Calculating Teacher Ratings

We create one measure of effectiveness from the two parts of our observation instrument to identify the impact of ATC on teacher effectiveness. First, we use the ratings that observers gave on each of the ten subdomains of the Danielson Framework for Teaching, and on our observation instrument.[3] We average those scores to create an overall rating for each teacher in each observation, and then combine each teacher's three ratings into one overall average rating. By doing so, we have 13 ratings per teacher—one overall average rating, an average classroom environment rating, an average instruction rating, and an average rating in each of the five subdomains of each classroom environment and instruction. Second, we also construct a rating for each teacher for each observation in classroom environment, instruction, and overall based on the specific teacher and student actions recorded by observers during the course of each observation. These practices (detailed in the observation instrument found in Appendix C) are tied specifically to the Framework for Teaching domains, and are averaged for each observed class period for each teacher. We then scale up the average frequency count score (between 0 and 1) onto the same scale as the observer ratings for each domain (1 to 4). Using the observer ratings includes the observers' subjective sense of how the classroom felt and qualitative information about how teacher and students interacted. In other cases, using observable facts may be more informative than relying on observers' potentially subjective ratings. We average

---

[3] The subdomains are: 1) respect and rapport; 2) culture of learning; 3) classroom routines and procedures; 4) behavior management; 5) physical environment; 6) communicate with students; 7) rigorous questions; 8) student engagement; 9) using assessment in instruction; and 10) responsiveness to students

each teacher's subjective (observer-given) and objective (based on frequency of practices) ratings in our analyses.[4] This allows us to utilize all information collected during each observation, while mitigating the shortcomings of either individual rating method. We standardize these ratings so that all results are reported in standard deviation units.

Measuring Effectiveness—Students' Perceptions

We surveyed all students of all ATC and comparison teachers in the spring of 2016. Students did not know that this survey was conducted as part of an evaluation of the Arkansas Teacher Corps, and were not compensated for completing the survey. We surveyed students in class, and teachers were not responsible for administering the survey to their students. We might expect students to give inflated ratings to all their teachers because they have built relationships with their teachers throughout the year and do not want to seem to insult them to outsiders. However, these inflated ratings should be given to both ATC and comparison teachers. Thus, while the ratings given by observers may be closer to a true measure of a teacher's effectiveness related to teaching practices, student ratings can measure differences between teachers on other important dimensions, such as building strong relationships with students and generating student engagement and enthusiasm among students. This also represents an important source of information on teacher classroom effectiveness.

Surveys consisted of 41 closed-response items, each answered on a 4-point Likert scale. Questions were adapted from the Panorama student survey, a nationally validated survey

---

[4] Observer-given ratings and frequency counts of teacher practices are highly correlated; the full correlation matrix is available in Appendix B.

instrument designed to measure students' perceptions of their teachers.[5] Questions were added

and modified to include comparative measures of teacher quality (e.g. my current teacher

compare to my other teachers). Students were also asked three open-ended questions about their

teacher. The full student survey is presented in Appendix D. We use 40 questions[6] to measure

eight constructs through the student survey. We collapse student survey results by teacher, so

that each teacher has one score for each construct measured to avoid unfair weighting of teacher

scores based on how many students are enrolled in each teacher's classes. This also removes any

artificial statistical precision potentially lent to the model by the large N of student surveys.

Table 2 summarizes these constructs, their internal reliability as measured by Cronbach's alpha,

and sample items from the student survey. We also include our prediction of how ATC teachers

will do on each construct, based on prior evaluations of alternatively certified teachers and

ATC's logic model.

---

[5] The Panorama Student Survey resulted from a collaboration between the Harvard Graduate School of Education and Panorama Education. More information can be found at http://www.panoramaed.com/panorama-student-survey
[6] One item (question 41) was excluded during the process of validating the constructs. Designed to be part of the preparation for class construct, student responses indicated that the question was confusing and did not contribute to the construct's validity.

Construct	# Items	Alpha	Sample Item	Predicted Effect
Overall learning	5	0.964	Overall, how much have you learned from this teacher about <SUBJECT>?	Null to Positive
High expectations	8	0.951	How often does this teacher encourage you to do your best?	Positive
Content knowledge	4	0.949	How often is your teacher able to answer your questions regarding <SUBJECT>?	Positive
Preparation for class	2	0.882	How prepared is your teacher for class?	Null to Negative
Relationships in class	6	0.969	If you walked into class upset, how concerned would your teacher be?	Positive
Relationships outside class	5	0.946	How approachable is your teacher outside of class?	Null to Negative
Behavior management	4	0.969	How often do students behave well in class?	Negative
Class engagement	6	0.970	Overall, how interesting does this teacher make what you are learning in class?	Null to Positive
			Total Surveys Administered	*7,265*

We measure eight constructs in the student survey: overall learning, high expectations, content knowledge, preparation for class, relationships in class, relationships outside of class, behavior management, and class engagement. Each construct has a high Cronbach's alpha, indicating its reliability. Previous research on alternative certification teachers leads us to believe that ATC teachers will have a positive effect in content knowledge, and a negative effect in behavior management. The criticism of TFA that its Corps Members are simply "voluntourists," without real connections to the communities in which they work, lead us to think that ATC Fellows will have null to negative effects in relationships outside of class. ATC's focus on high expectations and student relationships leads us to believe that ATC teachers will have a positive effect in high expectations and relationships in class. The availability of support for ATC teachers and ongoing development would lead us to believe that ATC teachers would be highly prepared for class, but their relative lack of experience could also put them at a disadvantage in

creating lesson plans and having materials for class, as they do not have materials from prior years to draw on—we thus hypothesize that there will be a null effect in this area. ATC development stresses student-centric teaching practices, which should increase ATC's effectiveness in class engagement; however, their counterparts have generally been trained for four to five years in pedagogy and teaching strategy, so we believe we will find a null effect on engagement as well. Of course, the ATC logic model explicitly calls for greater student learning as a result of ATC teachers being in the classroom; however, because of the mix of hypotheses in the seven components of a successful classroom, we believe this effect will be null to positive.

There are reasons to prefer both the observer ratings and student survey results in this analysis. While our observers observed each teacher two-three times throughout the 2016 spring semester, collecting hundreds of hours of information[7] on teacher practices and effectiveness, students ostensibly attend class every day and have access to more information about their teacher's day-to-day practices than do our intermittent observers. However, while our observers were highly trained and experienced in working with teacher evaluation and calibrated on our observation instrument, students each bring their own unique perspectives and criteria of teacher quality to their assessments of their teachers. Indeed, the constructs measured by the student survey have high internal validity, as shown by their high Cronbach's alphas, meaning that students are evaluating real aspects of teaching that observers are not able to measure. This indicates that observers and students are capturing different, if related, aspects of teaching.[8] Therefore, we do believe that information collected through student surveys are a useful way of

---

[7] In total, our observers rated 201 hours of class time during 268 classroom observations, each of which lasted an average of 45 minutes.

[8] In general, student survey constructs and observer outcomes are weakly correlated, although there are strong correlations between the three observed outcomes and the eight survey constructs. The full correlation matrix is shown in Appendix B.

triangulating teacher effectiveness; while they do not provide the whole (or the most objective)

picture of teacher effectiveness, they do add important insight into the impact of ATC teachers

on student experiences of school and ultimately, therefore, on students' futures.

Estimating Impacts of ATC—Building a Model

We use multivariate OLS regression techniques to determine the effect of ATC on

teacher effectiveness in classroom environment, instruction, and overall. Our general model is:

$$y_i = \beta_0 + \gamma_1 atc_i + \boldsymbol{\beta x_i} + \boldsymbol{\delta c_i} + \varepsilon$$

Our outcomes are our observer domains: overall effectiveness, classroom environment, and

instruction; and our student constructs: overall learning, high expectations, content knowledge,

preparation for class, relationships in class, relationships outside of class, behavior management,

and class engagement. Our coefficient of interest is γ_1, which estimates the average difference in

effectiveness between ATC Fellows and their matched counterparts. We include a variety of

teacher characteristics in vector $\boldsymbol{x_i}$, including gender, subject taught, average class size, and

experience (which we log to allow for diminishing returns). We also include school level effects,

$\boldsymbol{c_i}$, to control for school-level factors that may systematically impact teacher quality, such as a

supportive administration, instructional coaches, schoolwide discipline policies, and community

support. For observer outcomes, we show results with and without controlling for teacher

experience; for student outcomes we only report results while controlling for experience, but

results without controlling for experience are available in the appendix. Our preferred model

includes controls for experience. Although part of the treatment of ATC is having an

inexperienced teacher, the proper counterfactual to an ATC teacher is having a novice non-ATC

teacher. Thus, while it is instructive to see how ATC teachers compare to all other teachers, the program effect of ATC is better measured when experience is held equal.

VI.   Results

Observation Results

We begin by looking at teachers' effectiveness as measured by observers. Table 3 shows the results of OLS regressions analyzing the difference between ATC and non-ATC teachers. The left-most panel examines overall effectiveness, the middle panel shows our results on classroom environment, and the right-most panel shows our results when looking at instruction.

We saw earlier that our sample was unbalanced on gender, and therefore we control for gender in all models. We also include subject area controls, recognizing that what may be successful in a math classroom may not be as relevant or helpful in an English classroom. We also include average class size, to account for any potential systematic differences between teachers with varying class sizes. Finally, we include school fixed effects, removing the influence of between-school variation in teacher quality. The second model in each panel (columns 2, 5, and 8) is our preferred model, in which we include a control for experience. The literature suggests that teachers are generally less effective in year one than in later years, but effectiveness does not continue to increase with experience systematically after the first few years of teaching. We log each teacher's total experience to reflect these diminishing returns.

Table 3: Observer Ratings[9]

	Overall Rating		Classroom Environment		Instruction	
	(1)	(2)	(3)	(4)	(5)	(6)
ATC	0.042	0.087	-0.080	0.042	0.165	0.130
	(0.321)	(0.418)	(0.301)	(0.408)	(0.331)	(0.413)
Hard Sciences	-0.414	-0.821	-0.431	-0.757	-0.367	-0.824
	(0.418)	(0.622)	(0.382)	(0.621)	(0.445)	(0.590)
Social Sciences	-0.328	-0.521	-0.270	-0.546	-0.362	-0.458
	(0.354)	(0.424)	(0.292)	(0.434)	(0.406)	(0.401)
Female Teacher	0.167	-0.009	0.131	-0.040	0.181	0.012
	(0.249)	(0.494)	(0.264)	(0.474)	(0.228)	(0.490)
Avg Class Size	0.012	0.036	0.022	0.036	0.001	0.033
	(0.036)	(0.048)	(0.029)	(0.045)	(0.042)	(0.049)
Log(Experience)		0.123		0.167		0.072
		(0.141)		(0.158)		(0.126)
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	100	78	100	78	100	78
Adjusted R-squared	0.025	0.057	0.023	0.048	0.024	0.048

Standard errors clustered by school   *** p<0.01, ** p<0.05, * p<0.1

Table 3 first demonstrates the low explanatory power we have in these models—we have low adjusted R-squared values in all models. The problem is exacerbated in the models in which we do not control for teacher experience, validating our preferred model which does include experience. However, even after controlling for experience our explanatory power remains low, indicating that there are other, unaccounted factors that have a significant impact on observed teacher effectiveness. The second striking feature of Table 3 is the lack of statistical significance across models. We have a limited sample size—78 teacher-level observations when including experience—and therefore do not have the statistical power to detect an effect on a noisy outcome, including observed teacher effectiveness.

---

[9] In an alternate specification, we used interaction effects to determine if ATC teachers differ in effectiveness based on their subject area. None of the interaction effects were significant. Results available upon request.

Focusing first on our preferred model the left-most panel, column (2), we see that ATC teachers are rated slightly more positively (0.087 standard deviations) by observers on overall effectiveness than their counterparts, but the result is not statistically significant. Arts teachers are rated more highly than hard sciences or social sciences teachers by observers, but the difference is not significance and applies to both ATC and non-ATC teachers. Teacher effectiveness seems to increase slightly (0.036 standard deviations) as class size increases, but again the effect is not statistically significant. There are positive returns to experience, but by taking the log of teachers' years of experience, we account for diminishing returns to experience in the model. Our observers found no significant differences between ATC and non-ATC teachers in observed overall effectiveness.

The middle panel of Table 3 examines ATC effectiveness in classroom environment. Column (4) is our preferred model, in which we control for experience. In column (4), ATC teachers are nominally more effective in classroom environment than their counterparts (0.042 standard deviations), but the difference is not statistically significant. Hard sciences teachers are rated as nominally less effective in classroom environment than are arts teachers (by 0.757 standard deviations), but the difference is not statistically significant. Social sciences teachers are also nominally less effective in classroom environment than arts teachers (by 0.546 standard deviations), but the effect is not statistically significant. Female teachers are rated as slightly less effective than their male counterparts (0.040 standard deviations), but the difference is not statistically significant. Average class size has a small positive, but insignificant impact on observed teacher effectiveness in classroom environment. Experience again has positive, but diminishing, returns to effectiveness. The average ATC teacher is in their second year of teaching, so controlling for experience in this model may be controlling away part of the ATC

treatment effect. However, we do not know if the observed positive ATC effect in classroom environment without controlling for experience would persist as ATC teachers gain experience past their fifth year in the classroom, or if diminishing returns to experience would set in for ATC teachers in a similar pattern as is observed for non-ATC teachers. In 2015-16, the most senior ATC Fellows were their third year in the classroom, so we will need to track ATC alum as they continue to teach in order to observe the full effect of experience on ATC development.

The final panel of Table 3 analyzes ATC teachers' observed instructional effectiveness. Column (6) is our preferred model, as it controls for experience, which column (5) omits. In column (6), ATC teachers are an estimated 0.130 standard deviations more effective than their peers, but the effect is not statistically significant. Arts teachers are rated more favorably by observers than either hard sciences teachers or social sciences teachers, but the differences are not statistically significant. Female teachers are rated as essentially the same as male teachers in instruction. There is again a small, insignificant positive relationship between average class size and effectiveness (0.033 standard deviations), but the estimate is not statistically significant. Experience again has a positive, but diminishing and insignificant, impact on effectiveness.

The observers found slight positive effects for ATC in overall effectiveness, classroom environment, and instructional effectiveness, but the estimates were not statistically significant. In sum, third party observers were not able to detect a meaningful difference in observable classroom effectiveness between ATC and non-ATC teachers. We turn next to student perceptions of their teachers' effectiveness.

Student Survey Results

We have eight outcomes captured by student surveys: overall learning, high expectations, content knowledge, preparation for class, relationships in class, relationships outside of class, behavior management, and class engagement. Table 4 presents the results from our preferred model, in which we control for teacher experience as well as teacher gender, subject area, average class size, and school effects.[10] Our variable of interest is whether or not the teacher was trained through the ATC program.

Table 4: Student Survey Results

	Overall Learning	High Expectations	Content Knowledge	Preparation for Class	Relationships in Class	Relationships outside Class	Behavior Management	Class Engagement
ATC	0.164	0.290	0.462*	0.267	0.468	0.428	0.455*	0.288
	(0.327)	(0.317)	(0.263)	(0.300)	(0.334)	(0.324)	(0.260)	(0.333)
Hard Sciences	-0.624	-0.616	-0.333	-0.320	-0.309	-0.129	-0.113	-0.517
	(0.441)	(0.520)	(0.428)	(0.470)	(0.386)	(0.430)	(0.455)	(0.424)
Social Sciences	-0.291	-0.262	-0.110	0.036	-0.313	-0.123	-0.237	-0.476
	(0.548)	(0.541)	(0.366)	(0.462)	(0.562)	(0.561)	(0.409)	(0.523)
Female Teacher	-0.069	-0.052	0.023	-0.143	0.113	0.083	0.187	-0.098
	(0.447)	(0.450)	(0.398)	(0.452)	(0.390)	(0.411)	(0.346)	(0.403)
Avg Class Size	-0.062	-0.043	-0.033	-0.020	-0.089	-0.085	-0.066*	-0.082
	(0.086)	(0.066)	(0.047)	(0.056)	(0.064)	(0.058)	(0.036)	(0.068)
Log(Exper)	-0.228	-0.166	-0.080	-0.079	-0.101	-0.122	0.094	-0.221
	(0.151)	(0.159)	(0.145)	(0.156)	(0.152)	(0.154)	(0.118)	(0.165)
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	79	79	79	79	79	79	79	79
Adjusted R^2	-0.019	0.003	0.148	-0.081	0.085	0.087	0.190	0.086

Standard errors clustered at the school level   *** $p<0.01$, ** $p<0.05$, * $p<0.1$

[10] For the sake of brevity, we do not include results of models that do not control for experience, or of ATC Fellows by subject area. Results of analyses that do not control for experience are available in Appendix A. Without controlling for experience, ATC Fellows are significantly more effective in relationships in class, relationships out of class, and class engagement. There are no differences in effectiveness among ATC teachers by subject area. Results are available upon request.

We were also able to expand our sample size slightly (from 78 to 79) because some principals who would not allow us to observe teachers did agree to let us administer surveys to students, provided the teachers and students consented to the survey. However, our sample remains limited by the availability of information on teachers' years of experience and the small universe of ATC teachers we are analyzing.

Before running our analyses, we predicted that ATC teachers would be significantly more effective than their counterparts in setting high expectations for students, content knowledge, and relationships in class. Our results bear out these predictions for content knowledge. Students estimate that ATC teachers are 0.462 standard deviations more effective in content knowledge than non-ATC teachers, an effect that is significant at the 90% confidence level. Content knowledge was measured on a 1-4 scale, where a score of 4 indicates that students perceived their teachers to have a high level of content knowledge. The mean score found for content knowledge was 3.26, indicating that in general students believe their teachers possess fairly high levels of content knowledge. Despite this high bar, 65.0% of ATC teachers were rated as having above average levels of content knowledge, while only 48.9% of comparison teachers were rated as having above average levels of content knowledge. Among ATC teachers, the mean content knowledge rating was 3.33, while among comparison teachers the mean content knowledge rating was 3.20.

We predicted that ATC teachers would be less effective at behavior management, as they have not had the same training as non-ATC teachers in student teaching or other opportunities to learn from and practice behavior management techniques. However, ATC teachers are an estimated 0.455 standard deviations more effective than non-ATC teachers in behavior

management, and the effect is statistically significant. The average rating in behavior management was 2.99 for all teachers; 60% of ATC teachers were above average in behavior management, while only 47% of comparison teachers were above average in behavior management. Among ATC Fellows, the average rating for behavior management was 3.05, while among comparison teachers the average rating for behavior management was 2.92.   In practical terms, this means that on average students slightly more than agree that their ATC teacher is effective at behavior management, while on average students are slightly disagree that their non-ATC teacher is effective at behavior management.[11]

We predicted that ATC teachers would be more effective in developing relationships with students inside the classroom because of ATC's commitment to developing critical consciousness within its teachers, and emphasizing the importance of relationships during its teacher training sessions. Students estimated that ATC teachers were 0.468 standard deviations better at developing relationships with students in class, but the relationship was not statistically significant. We also predicted that ATC teachers would be rated significantly more effective than non-ATC teachers in setting high expectations. Although ATC teachers are an estimated 0.290 standard deviations better at setting high expectations than non-ATC teachers, the effect is not statistically significant.

Because of competing factors that could make ATC teachers more or less effective in overall learning and class engagement, we predicted that a null to positive effect would be found for ATC teachers in these two constructs. This is what we observe in the data. ATC teachers are an estimated 0.164 standard deviations better in improving students' overall learning, but the result is not statistically significant. ATC teachers are an estimated 0.288 standard deviations

---

[11] Each construct was measured on a 4-point scale: 1- strongly disagree; 2- disagree; 3- agree; 4- strongly agree.

better at engaging students throughout class than on-ATC teachers, but again the result is not statistically significant.

We predicted that ATC teachers would be as or less effective than their counterparts in developing relationships with students outside of class, since ATC Fellows are not from the communities in which they are teaching, and therefore might feel less compelled to attend afterschool functions, or linger at the school after hours, particularly if they commute in from another city in Arkansas. However, we estimate that ATC teachers were 0.428 standard deviations more effective at developing relationships with students outside of class. While this result is not statistically significant, it provides suggestive evidence refuting the notion that alternatively certified teachers are simply "voluntourists" who lack a true commitment to the area in which they serve.

We also predicted that ATC teachers would be as or less prepared for class than non-ATC teachers, as they may have fewer resources from past years or connections in the teaching world to ask for suggestions in planning. However, we see a null positive effect for ATC teachers on preparation for class as measured by students. ATC teachers are an estimated 0.267 standard deviations better at being prepared for class, but the difference is not statistically significant.

Consistent with the findings from our observers, hard sciences and social sciences teachers are nominally less effective than arts teachers in all eight constructs captured by the student surveys, although the differences are not statistically significant. Again consistent with the observers' findings, students find small differences in effectiveness between male and female teachers, and the difference is never statistically significant. In contrast to observers, who documented a slight positive and insignificant relationship between average class size and

effectiveness, students document a small negative relationship between average class size in all eight measures of effectiveness. This difference is significant for behavior management—a one student increase in average class size is associated with a 0.066 standard deviation decrease in effectiveness in behavior management. Students do not feel that larger class sizes are beneficial for teacher effectiveness in any of these areas. Students do not perceive the same returns to experience that observers noted, but again the estimates are not statistically significant.

Overall, students consistently rate ATC teachers are more effective than non-ATC teachers, and significantly so in content knowledge and behavior management. Our prediction that ATC teachers would be more effective in content knowledge was correct. We predicted that ATC teachers would be null to negative in behavior management, and instead found that they were significantly more effective according to the students in the classrooms. However, we also predicted that ATC teachers would be significantly more effective in setting high expectations and building relationships in class, and found null to positive results instead. We predicted that we would find a null to positive effect for ATC teachers in overall learning and class engagement, which we did see in the results of our analysis. We predicted that ATC teachers would be null to negative in preparation for class; our result was null to positive. ATC teachers are fulfilling the goals of the logic model in content knowledge, and are either fulfilling or well on their way to fulfilling the goals of the program in expectations, preparation for class, behavior management, and student engagement.

VII.   Conclusion

Before embarking on this evaluation, we considered what sort of outcomes should be viewed as positive or negative for the Arkansas Teacher Corps, a nearly brand new alternative

cortication program serving only economically challenged schools and districts. The first half of the evaluation was based on multiple formal observations from trained faculty from an accredited College of Education within the state of Arkansas. On this measure, an optimistic advocate for the new program might have hoped that ATC teachers would surpass the typical peer teachers in the district, but a more realistic observer might suggest a good outcome for ATC would be if the observers couldn't tell the difference between the alternatively-certified teachers and their typical peers. And, with our relatively small sample of teachers, we indeed found in this analysis that ATC teachers were indistinguishable from non-ATC teachers to outside observers on measures of overall effectiveness, classroom environment, and instruction.

Nevertheless, while the small sample size and low statistical power ensured that any observed nominal differences were not statistically significant, the good news for ATC was that most of the observed differences were in favor of ATC teachers. Observers rated ATC teachers as nominally more effective on overall effectiveness, classroom environment, and instruction.

With the large number of student surveys collected and the explicit ATC focus on relationship building, one might have been more optimistic about the potential for this young program to inspire meaningful differences in student responses. Indeed, the results from the student section of our evaluation do suggest that ATC teachers do have the potential to improve the educational experiences for students in low-income schools. Students felt consistently more positive about ATC teachers than non-ATC teachers, rating ATC teachers as nominally more effective than their counterparts in boosting overall learning, setting high expectations, being prepared for class, building relationships with students in and out of class, and engaging students in class. Students also rated ATC teachers as significantly more effective in possessing content knowledge and managing student behavior. Table 5 summarizes these findings, and depicts the

relationship between the observer domains and student survey constructs. All effects are in standard deviations.

Table 5: Summary of Findings

Outcome Measure	Estimated Effect	Significant?
Overall Indicators		
Overall Effectiveness (Observers)	0.087	No
Overall Learning (Students)	164	No
Classroom Environment		
Classroom Environment (Observers)	0.042	No
High Expectations (Students)	0.290	No
Relationships in Class (Students)	0.468	No
Behavior Management (Students)	0.455	Yes
Instruction		
Instruction (Observers)	0.130	No
Content Knowledge (Students)	0.462	Yes
Class Engagement (Students)	0.288	No
Other Outcomes		
Preparation for Class (Students)	0.267	No
Relationships outside of Class (Students)	0.428	No

As reiterated in Table 5, ATC teachers are found to be significantly more effective in content knowledge and behavior management. ATC teachers are also found to be nominally more effective in preparing for class, engaging students in learning, instructional practices, behavior management, setting high expectations, fostering overall learning, creating a classroom environment conducive to learning, and overall observed effectiveness.

The ATC model relies on the organization's ability to recruit individuals highly knowledgeable about their subject and with a commitment to teaching in a struggling school to ensure that all students, regardless of their background, has access to a high quality teacher. These results suggest that ATC recruitment and content development processes are working to ensure that they are providing schools with teachers who are more knowledgeable about their subject than the schools would otherwise be able to find. ATC's focus on critical consciousness and developing authentic relationships with students and communities also seems to be working

from the students' perspectives, who feel they have better relationships with ATC teachers than non-ATC teachers both inside and outside the classroom. Observers also note that ATC teachers are more frequently engaging in positive classroom environment actions than non-ATC teachers. On the whole, ATC is providing schools with teachers who are just as good as or slightly better than the non-ATC teachers they have also chosen to hire, indicating that the three-year old program with its abridged teacher training program is at least as effective as traditional four to five year training programs.

One plausible interpretation of our results is that they represent a lower bound of the effectiveness of ATC teachers, because the comparison group in this analysis is made up of "typical" teachers working in the schools with the ATC Fellows. Recall that school leaders seek ATC Fellows when they cannot fill teaching positions through their normal recruitment and application processes. Thus, the comparison teachers we use in this analysis may well be more qualified and capable than the true counterfactual teachers for ATC Fellows—those applicants who districts would have been forced to hire if they had had no other options.

The Arkansas Teacher Corps was created in response to a perceived need in southern Arkansas for qualified, motivated teachers. The program attempted to provide high-needs schools with a limited pool of teacher candidates a source of qualified, effective teachers. While the quantitative evidence on the effectiveness of ATC to provide high-quality teachers to schools and students in the state is inconclusive, the organization is meeting one of its stated goals: to provide teachers to understaffed schools in need. As the three-year old program continues to grow and collect data on the effectiveness of its teachers, these and future evaluations of student outcomes should guide continued improvement within the program.

References

Anderson, A. (2013). Teach for America and Symbolic Violence: A Bourdieuian Analysis of Education's Next Quick-Fix. *The Urban Review, 45*(5), 684-700. doi:http://0-dx.doi.org.library.uark.edu/10.1007/s11256-013-0241-x

Antecol, H., Eren, O., & Ozbeklik, S. (2013). The Effect of Teach for America on the Distribution of Student Achievement in Primary School. IZA Discussion Paper # 7296.

Arkansas Department of Education (2016a). 2015-2016 ACT Aspire Preliminary State Results. Retrieved from http://www.arkansased.gov/divisions/learning-services/student-assessment/test-scores/year?y=2016

Arkansas Department of Education (2016b). Arkansas Educator Preparation Performance Report. Retrieved from https://adedata.arkansas.gov/eppr/docs/State/StatewideReportApril2016.pdf

Arkansas Department of Education (2015c). Critical Academic Licensure Shortage Areas 2015-2016 School Year. Retrieved from http://www.arkansased.gov/public/userfiles/HR_and_Educator_Effectiveness/HR_Educator_Licensure/15-16_shortage_areas_flattened_text.pdf

Arkansas Department of Education (2016d). Critical Academic Licensure Shortage Areas for the 2016-2017 School Year. Retrieved from http://www.arkansased.gov/public/userfiles/HR_and_Educator_Effectiveness/HR_Educator_Licensure/16-17_Critical_Shortage_Areas-72337.pdf

Arkansas Department of Education (2016e). Routes to Educator Licensure. Retrieved from http://www.arkansased.gov/divisions/human-resources-educator-effectiveness-and-licensure/educator-licensure-unit/routes-to-educator-licensure

Arkansas Department of Education (2016f). Statewide Information System Reports: Statewide Enrollment by Race. Retrieved from https://adedata.arkansas.gov/statewide/State/EnrollmentByRace.aspx

Arkansas Department of Education (2016g). Statewide Information System Reports: Statewide Free/Reduced/Paid Lunch Counts. Retrieved from https://adedata.arkansas.gov/statewide/State/FreeReducedPaidLunch.aspx

Bill and Melinda Gates Foundation (2012). Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains. Measures of Effective Teaching Project. Retrieved from http://k12education.gatesfoundation.org/wp-content/uploads/2016/06/MET_Gathering_Feedback_for_Teaching_Summary1.pdf

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.

Chetty, R., Friedman, J., and Rockoff, J. (2011). *The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood.* National Bureau of Economic Research Working Paper 17699. Retrieved from http://www.rajchetty.com/chettyfiles/value_added.pdf

Chiang, H., Clark, M., McConnell, S. (2014). *Supplying Disadvantaged Schools with Effective Teachers: Experimental Evidence on Secondary Math Teachers from Teach for America.* Mathematica Policy Research.

Clark, Melissa A., Hanley S. Chiang, Tim Silva, Sheena McConnell, Kathy Sonnenfeld, Anastasia Erbe, and Michael Puma. (2013). *The Effectiveness of Secondary Math Teachers from Teach For America and the Teaching Fellows Programs* (NCEE 2013-

4015). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Constantine, J., Player D., Silva, T., Hallgren, K., Grider, M., and Deke, J. (2009). *An Evaluation of Teachers Trained Through Different Routes to Certification, Final Report* (NCEE 2009- 4043). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Darling-Hammond, L. (2005). Does Teacher Preparation Matter? Evidence about Teacher Certification, Teach for America, and Teacher Effectiveness. *Education Policy Analysis Archives, 13(42)*. Retrieved from http://files.eric.ed.gov/fulltext/EJ846746.pdf

Decker, P. T., Mayer, D. P., & Glazerman, S. (2004). The effects of Teach for America on students: Findings from a national evaluation. University of Wisconsin--Madison, Institute for Research on Poverty.

Glazerman, S., Mayer, D., and 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.

Ingersoll, R. (2002). *Out-of-Field Teaching, Educational Inequality, and the Organization of Schools: An Exploratory Analysis*. Center for the Study of Teaching and Policy. Retrieved from https://depts.washington.edu/ctpmail/PDFs/OutOfField-RI-01-2002.pdf

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.

King Rice, J. (2003). Teacher Quality: Understanding the effectiveness of teacher attributes. Economic Policy Institute: Washington, D.C.

Krieg, J., Theobald, R., and Goldhaber, D. (2009). A Foot in the Door: Exploring the Role of Student Teaching Assignments in Teachers' Initial Job Placements. CEDR Working Paper 2015-2.1. University of Washington, Seattle, WA. Retrieved from http://www.cedr.us/papers/working/CEDR%20WP%202015_2.1.pdf

Labaree, D. (2010). Teach for America and Teacher Ed: Heads They win, Tails We Lose. *Journal of Teacher Education, 61*(1-2), 48-55. doi: 10.1177/022487109347317.

Laczko-Kerr, I., & Berliner, D. C. (2002). The Effectiveness of" Teach for America" and Other Under-certified Teachers on Student Academic Achievement: A Case of Harmful Public Policy. *Education Policy Analysis Archives, 10*(37), 1-53.

McGloin, C. and Georgeou, N. (2016). 'Looks good on your CV': The sociology of voluntourism recruitment in higher education. *Journal of Sociology, 52*(2), 403-417. Doi: 10.1177/144078331452416.

NCES (2016). State Profiles: Arkansas. Nation's Report Card. Retrieved from http://nces.ed.gov/nationsreportcard/states/

Penner, K. (2014). Does Teach for America Have Long-Term Impacts? *Society for Research on Educational Effectiveness*.

Raymond, M., Fletcher, S. H., & Luque, J. (2001). Teach for America: An evaluation of teacher differences and student outcomes in Houston, Texas. CREDO, Stanford University.

Sass, T. (2011). *Certification Requirements and Teacher Quality: A Comparison of Alternative Routes to Teaching, Working Paper (CALDER 2011-64).* National Center for Analysis of Longitudinal Data in Education Research.

Suell, J.L. and Piotrowski, C. (2006). Efficacy of Alternative Certification Programs: A Study of the Florida Model. *Education, 127*(2), 310-315.

Teach for America (2016). Our History. Retrieved from https://www.teachforamerica.org/about-us/our-story/our-history

The New Teacher Project (2009). The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness. Retrieved from http://tntp.org/assets/documents/TheWidgetEffect_2nd_ed.pdf

Xu, Z., Hannaway, J., and 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.

Zelazek, J. R., Williams, W.W., McAdams, C., Palmer, K, and Central Missouri State Univ, W. (2001). Teacher Education Follow-Up Study 2001: A Summary of First and Second Year Teachers, and their Employers with Respect to the State of Missouri Standards.

Zumwalt, K. and Craig, E. (2009). Teachers' Characteristics: Research on the Demographic Profile. In M. Cochran-Smith and K.M. Zeichner (Eds.), *Studying Teacher Education: The Report of the AERA Panel on Research and Teacher Education* (111-156). New Jersey: Lawrence Erlbaum Associates, Inc.

Appendix A: Student Survey Results from Alternative Models

Table A1: Student Survey Results without Controlling for Teacher Experience

	Overall Learning	High Expectations	Content Knowledge	Preparation for Class	Relationships in Class	Relationships outside Class	Behavior Management	Class Engagement
ATC	0.284	0.370	0.372	0.320	0.547**	0.510**	0.340	0.418*
	(0.203)	(0.231)	(0.227)	(0.219)	(0.230)	(0.206)	(0.203)	(0.218)
Hard Sciences	-0.325	-0.287	0.011	-0.244	-0.111	0.081	0.073	-0.288
	(0.453)	(0.519)	(0.440)	(0.358)	(0.444)	(0.451)	(0.428)	(0.423)
Social Sciences	-0.067	0.026	-0.018	0.144	0.104	0.220	0.023	-0.188
	(0.374)	(0.412)	(0.314)	(0.271)	(0.425)	(0.367)	(0.293)	(0.344)
Female Teacher	0.141	0.085	0.122	0.122	0.248	0.236	0.361	0.048
	(0.392)	(0.378)	(0.393)	(0.381)	(0.354)	(0.341)	(0.350)	(0.359)
Avg Class Size	-0.052	-0.028	-0.037	-0.001	-0.056	-0.065	-0.054	-0.063
	(0.066)	(0.054)	(0.043)	(0.049)	(0.058)	(0.049)	(0.039)	(0.062)
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	98	98	98	98	98	98	98	98
Adjusted R^2	0.036	0.046	0.127	0.026	0.091	0.135	0.151	0.086

Standard errors clustered by school   *** p<0.01, ** p<0.05, * p<0.1

Appendix B: Correlation Matrices

Table B1: Correlations of Observer-Given Ratings and Frequency Count Based Ratings

		Practices			Subjective		
		Overall	Class Enviro	Instruct	Overall	Class Enviro	Instruct
Subjective	Overall	1.00					
	Class Enviro	0.97	1.00				
	Instruct	0.96	0.87	1.00			
Practices	Overall	0.82	0.77	0.80	1.00		
	Class Enviro	0.82	0.81	0.78	0.94	1.00	
	Instruct	0.71	0.64	0.73	0.94	0.76	1.00
	N	*101*	*101*	*101*	*101*	*101*	*101*

Table B2: Observer and Student Survey Outcome Correlations

	1.Ob Rating	2. Ob Class Enviro	3. Ob Instruct	4. Overall Learning	5. High Expect	6. Content Know	7. Prep	8. Relate In Class	9. Relate Out class	10. Behave	11. Engage
1	1.00										
2	0.95	1.00									
3	0.95	0.80	1.00								
4	0.21	0.25	0.15	1.00							
5	0.29	0.31	0.25	0.93	1.00						
6	0.17	0.20	0.12	0.81	0.81	1.00					
7	0.25	0.30	0.16	0.90	0.90	0.82	1.00				
8	0.13	0.12	0.12	0.82	0.87	0.75	0.80	1.00			
9	0.16	0.16	0.14	0.86	0.90	0.74	0.80	0.96	1.00		
10	0.16	0.22	0.09	0.76	0.76	0.82	0.79	0.83	0.80	1.00	
11	0.17	0.18	0.14	0.93	0.92	0.79	0.85	0.90	0.92	0.79	1.00

Appendix C: Observation Instrument
ATC Teacher Effectiveness Study—Observation Form
Observations should last the entire class period
Adapted from the AR Teacher Excellence Support System Formative Evaluation form developed by Charlotte Danielson

Check here if class was out of room at time of observation: _____
Check here if there was a substitute teacher: _____
Check here if students were testing _____

Date: _____   **Time In/Out**: _____/_____   **Total minutes_____**

School: _____   **Teacher**: _____   **Grade**: _____

Number of Students Present: _____

Subject of Lesson: (circle) Math Science English/LA Soc. Studies History Other

Topic of Lesson: _____

Note: If classes change during observation, do not continue on the same sheet

Activity Observed *Circle whether the action was observed, not observed, or not applicable in each time block*	Block 1	Block 2	Block 3
Classroom Environment—Student Actions			
1 Students put classmates down	Y N	Y N	Y N
2 Students have meaningful discussion with each other	Y N N/A	Y N N/A	Y N N/A
3 Students encourage each other	Y N N/A	Y N N/A	Y N N/A
4 Students are in the appropriate place (at desk, getting material, coming to board, etc.) in accordance with teacher instructions or clear procedure	Y N	Y N	Y N
5 Students' voice levels are appropriate to the activity (silent if teacher is talking, discussing with partner(s) if assigned, presenting effectively if asked)	Y N	Y N	Y N
6 Students enter/leave classroom without disruption, if necessary (clear procedure for bathroom, tardiness, etc.)	Y N N/A	Y N N/A	Y N N/A
7 Students transition between activities without disruption	Y N N/A	Y N N/A	Y N N/A
8 Students appeared to be on task and engaged the majority of the time	Y N	Y N	Y N
Classroom Environment—Teacher Actions			

9	Teacher addresses student(s) failing to meet expectations appropriately (behavior, place, volume, activity)	Y N N/A	Y N N/A	Y N N/A
10	Teacher knows students (evidence by use of names, personalized motivators, relevant examples in explaining content)	Y N	Y N	Y N
11	Teacher puts students down	Y N N/A	Y N N/A	Y N N/A
12	Teacher answers student questions	Y N N/A	Y N N/A	Y N N/A
13	Desk arrangement facilitates learning activity (facing front if teacher is lecturing, groups if working in groups, etc.)	Y N N/A	Y N N/A	Y N N/A
14	Technology is in use to achieve lesson objective	Y N N/A	Y N N/A	Y N N/A
15	Manipulatives used to achieve lesson objective	Y N N/A	Y N N/A	Y N N/A
16	Materials for activity or task are accessible to all students	Y N	Y N	Y N
17	Teacher gives clear instructions for how to transition between activities	Y N N/A	Y N N/A	Y N N/A
	Instruction—Student Actions			
18	Students ask questions relevant to the lesson's content	Y N	Y N	Y N
19	Number of students contributing to class/small group discussion: (please provide a count in each box)			
20	Students have choice in what they will be doing	Y N N/A	Y N N/A	Y N N/A
21	Students are working independently	Y N N/A	Y N N/A	Y N N/A
	Instruction—Teacher Actions			
22	Teacher explains purpose of each activity	Y N N/A	Y N N/A	Y N N/A
	No instruction- leave next 3 rows blank	Y N	Y N	Y N
23	Direct instruction	Y N	Y N	Y N N/A
24	Students working in pairs or small groups	Y N	Y N	Y N N/A
25	Student-led full class discussion	Y N	Y N	Y N
26	Teacher uses content-specific vocabulary	Y N	Y N	Y N
27	Teacher uses accessible and appropriate Tier 2 vocabulary (not content specific, but widely used)	Y N	Y N	Y N
	No questions asked—leave next 5 rows blank	Y N	Y N	Y N

28	Teacher's instructions prompt students to recall information	Y N	Y N	Y N
29	Teacher's instructions prompt students to explain, summarize, infer, or discuss	Y N	Y N	Y N
30	Teacher's instructions prompt students to apply information in a new situation	Y N	Y N	Y N
31	Teacher's instructions prompt students to analyze or evaluate content	Y N	Y N	Y N
32	Teacher's instructions prompt students to create something based on learned content	Y N	Y N	Y N
33	Instruction is individualized to meet the needs of the students	Y N	Y N	Y N
34	Teacher modifies the lesson if necessary	Y N	Y N	Y N
35	Teacher gives formal (pre-made) formative assessment	Y N	Y N	Y N
36	Teacher gives informal formative assessment (CFU)	Y N	Y N	Y N
37	Teacher gives feedback to formative assessment	Y N N/A	Y N N/A	Y N N/A
Other Observations				
38	Teacher makes error when explaining content	Y N	Y N	Y N
39	Students can state the purpose of the lesson (ask one student when it will not disrupt the lesson)	Y N N/A		
	Classroom environment has relevant or motivating decorations (student work, data tracker, posters, word wall, content reminders, clearly marked turn trays, etc.)	Y N		

Additional Notes (more space on back if needed):

SUMMARY	TOTAL OBSERVATION
(Please circle/highlight your rating of this teacher in each domain, and provide a short explanation of your rating)	

What is the level of respect and rapport in the classroom?		Q's 1, 3, 10, 11	
Unsatisfactory—regular interruptions between teacher and students and among students; interactions are inappropriate, insensitive, or negative. Teacher does not deal with disrespectful behavior.	Basic—inconsistent interactions between teacher and students and among students; teacher may display favoritism or dislike of a particular student. Teacher attempts to respond to disrespectful behavior, but is not always successful.	Proficient—teacher-student interactions are friendly and demonstrate caring and respect. Interactions are appropriate, polite, and respectful among students. Teacher consistently and successfully responds to disrespectful behavior among students.	Distinguished—interactions between teacher and individual students are highly respectful, warm, and caring. Teacher is sensitive to individual student identities. High levels of civility between all members of the class.

Please explain:

What is the classroom's culture of learning?		Q's 3, 8, 12	
Unsatisfactory—lack of teacher or student commitment to learning; little or no student energy or investment in tasks. Teacher holds low to medium expectations of students; might hold high expectations for one or two students.	Basic—little teacher or student commitment to student; class seems to be going through the motions rather than energetically committing to learning. High expectations held only for the few students who seem to have a natural aptitude for the subject.	Proficient—learning is valued by all, with high expectations for all students. Classroom is a cognitively busy place where it is understand that hard work leads to success. Students see themselves as learners and support each other in class.	Distinguished—classroom is a cognitively vibrant place, with a shared belief in the importance of learning. Teacher holds high expectations for all students; students respond by consistently working hard, initiating self-improvement, making revisions, adding detail, and helping peers.

Please explain:

To what extent do classroom routines and procedures effectively contribute to learning?		Q's 6, 7, 17	
Unsatisfactory—much instructional time is lost to inefficient routines and procedures. Little or no evidence of teacher managing groups, transitions, or materials. Little or no evidence that students know or follow established routines.	Basic—some instructional time is lost to partially effective classroom routines and procedures. Teacher's management of instructional groups, transitions, and materials is inconsistent. Students require regular guidance and prompting to follow established routines.	Proficient—little loss of instructional time due to effective classroom routines and procedures. Teacher's management of instructional groups and materials are consistently successful. Students follow established classroom routines with minimal guidance and prompting.	Distinguished—instructional time is maximized due to efficient classroom routines and procedures. Students contribute to management of instructional groups, transitions, and materials. Routines are clearly well understood and may be initiated by students.

Please explain:

To what extent is student behavior well-managed?		Q's 4, 5, 9	
Unsatisfactory—no apparent established standards of conduct, and little or no teacher monitoring of student behavior. Students	Basic—established standards of conduct, but inconsistent implementation. Teacher tries to monitor student behavior and	Proficient—student behavior is generally appropriate; clear established standards of conduct. Teacher consistently responds to student misbehavior in a	Distinguished—student behavior is entirely appropriate, and actively monitor their own and their peers' behavior against established standards of content. Teacher's monitoring of student

challenge the standards of conduct. Response to students' misbehavior is repressive or disrespectful.	respond to student misbehavior, but results are uneven.	proportionate and respectful manner.	behavior is subtle and preventative; response to student misbehavior is sensitive and respectful.

Please explain:

How conducive to student learning is the physical environment?		Q's 13, 14, 15, 16	
Unsatisfactory—physical environment is unsafe or inaccessible to many students. Poor alignment between arrangement of furniture and resources with lesson activities.	Basic—classroom is safe, and essential learning is accessible to most students. Teacher's use of physical resources is moderately effective; partially effective attempts are made to modify the physical arrangement to suit learning activities.	Proficient—classroom is safe and learning is accessible for all students; physical arrangement is appropriate for each learning activity.	Distinguished—classroom is safe and learning is accessible to all students, including those with special needs. Physical arrangement is appropriate to each learning activity, and students contribute to use or adaptation of the physical environment to advance learning.

Please explain:

How well does the teacher communicate with students?		Q's 17, 22, 26, 27, 33, 38	
Unsatisfactory—purpose of lesson, directions, and procedures are unclear and confusing to students. Teacher's explanation of content contains major errors. Teacher's grammar, syntax, or inappropriate use of vocabulary leaves students confused.	Basic—teacher's attempt to explain instructional purpose has limited success. Directions and procedures must be re-explained after initially causing confusion. Explanation of content may contain minor errors. Vocabulary is limited or not fully appropriate for students' ages or backgrounds. Students are not invited to engage with explanation of content.	Proficient—instructional purpose is clearly communicated, including its place in the course's overall sequence of content. Explanation of content is scaffolded, clear, accurate, and connects with students' experiences and knowledge. Students are invited to engage with explanation of content. No errors in grammar or syntax; vocabulary is appropriate for students.	Distinguished—purpose of lesson is linked to student interests; directions and procedures are clear and anticipate student misunderstanding. Explanation of content is thorough and clear, with appropriate scaffolding and connections to student interests. Students help extend content and explain material to peers. Teacher's language is expressive and expands students' vocabulary.

Please explain:

To what extent does the teacher push students through rigorous questioning?		Q's 18, 28, 29, 30, 31, 32	
Unsatisfactory—questions are not cognitively challenging, and mostly on the low end of Bloom's taxonomy. Teacher mediates all questions and answer, with a few students dominating the discussion.	Basic—single path of inquiry, with pre-determined answers, or teacher attempts to ask higher-order questions, but only a few students are involved in the exchange. Teacher attempts to engage all students in the discussion with uneven results.	Proficient—teacher focuses on higher order questions, with some lower-level questions added in. Students have a genuine discussion, with appropriate wait time after questions, and minimal participation by the teacher. Teacher successfully engages most students in the discussion.	Distinguished—variety of questions and prompts are used to challenge students, advance high-level thinking, and promote meta-cognition. Students formulate questions, initiate topics, and contribute to discussion without prompting. Students ensure that all voices are heard in the discussion.

Please explain:

To what extent are students engaged in learning?		Q's 19, 20, 21, 23, 24, 25, 33, 34	
Unsatisfactory—learning tasks and activities are poorly aligned with instructional outcomes, or require only rote responses. The pace of the lesson is too slow or too rushed, leaving few students intellectually engaged or interested.	Basic—learning or prompts are partially aligned with instructional outcomes, but require only minimal thinking. Most students are passive or compliant. The pacing of the lesson may not provide students the time needed to be intellectually engaged.	Proficient—learning tasks and activities are aligned with instructional outcomes and are designed to challenge student thinking. Teacher scaffolding keeps students intellectually engaged with important and challenging content. Pacing of lesson is appropriate. Learning is relevant for all students.	Distinguished—virtually all students are intellectually engaged with challenging content. Learning tasks are well designed and scaffolded to align with instructional outcomes. Students explore important content, and initiate inquiry of content. Pacing of lesson allows students to engage meaningfully with content, reflect on learning, and consolidate understanding. Students may have some choice in how they complete tasks, and assist each other when appropriate.

Please explain:

To what extent does the teacher effectively use assessment in instruction?		Q's 34, 35, 36, 37	
Unsatisfactory—little or no assessment or monitoring of student learning; no or poor feedback. Students do not appear to know assessment criteria and do not engage in self-assessment.	Basic—sporadic assessment during instruction and some monitoring of progress of learning. Feedback to students is general, with students only partially aware of the assessment criteria. Questions/prompts/assessments are rarely used to diagnose evidence of learning.	Proficient—assessment is regularly used during instruction, through teacher and student monitoring of progress of learning. Students receive accurate, specific feedback that advances learning. Students are aware of assessment criteria, and engage in some self-assessment. Questions/prompts/assessments are used to diagnose evidence of learning.	Distinguished—assessment is fully integrated into instruction. Students are aware of and may have contributed to the assessment criteria. Students self-assess and receive feedback from each other and from the teacher. Feedback is accurate, specific, and advances learning. Questions/ prompts/assessments are regularly used to diagnose evidence of learning by individual students.

Please explain:

To what extent is the teacher flexible and responsive to student needs?		Q's 33, 34, 37	
Unsatisfactory—teacher adheres to instruction plan despite evidence of poor student understanding or lack of interest. Teacher ignores student questions, and blames students or their home environment for any difficulty in learning.	Basic—teacher attempts to modify the lesson when needed and respond to student questions and interests, with moderate success. Teacher accepts responsibility for student success, but has a limited repertoire of instructional strategies.	Proficient—teacher promotes successful learning of all students, making minor adjustments as needed to instructional plan, and accommodating student questions, needs, and interests. Teacher persists in seeking approaches for students struggling in the lesson, drawing on a broad repertoire of instructional strategies.	Distinguished—teacher makes use of spontaneous events, student interest, and other opportunities to enhance learning. Teacher successfully adjusts and differentiates instruction to address individual student misunderstandings. Teacher persists in seeking effective approaches for students who need help, using an extensive repertoire of

			instructional strategies and soliciting additional resources.
Please explain:			
How would you rate this teacher overall?			
Unsatisfactory—teacher is ineffective in creating a productive learning experiences for students. Students do not appear to have intellectually benefitted from going through the lesson.	Basic—teacher attempts to provide a meaningful learning experience for students, with limited effectiveness. Some students appear to have intellectually benefitted from the lesson.	Proficient—teacher provided a productive learning experience for most students. Most students have intellectually benefitted from the lesson, and know how the course is helping them grow as learners.	Distinguished—teacher provided a meaningful learning experience for virtually all students. Students can clearly articulate what they are gaining from the course and from that particular lesson.
Please explain:			

Appendix D: Student Survey Template[12]

1. Overall, how much have you learned from this teacher about <SUBJECT>?	Almost nothing ○	A little bit ○	Quite a bit ○	A tremendous amount ○
2. For this class, how clearly does this teacher present the information that you need to learn?	Not at all clearly ○	Slightly clearly ○	Quite clearly ○	Extremely clearly ○
3. How often does this teacher give you feedback that helps you learn (for example: comments or grading on assignments or projects)?	Almost never ○	Once in a while ○	Often ○	Almost always ○
4. How often does this teacher require everyone to participate in class?	Almost never ○	Once in a while ○	Often ○	Almost always ○
5. How much have you learned from this teacher compared to your other teachers this year?	A lot less ○	Less ○	More ○	A lot more ○
6. This teacher gives me feedback that helps me learn ___ often than my other teachers this year.	A lot less ○	Less ○	More ○	A lot more ○
7. This teacher requires everyone to participate in class_____ than my other teachers this year.	A lot less often ○	Less often ○	More often ○	A lot more often ○
8. How knowledgeable is your teacher about <SUBJECT>?	Not at all knowledgeable ○	A little bit knowledgeable ○	Quite knowledgeable ○	Extremely knowledgeable ○
9. How often is your teacher able to answer your questions regarding <SUBJECT>?	Almost never ○	Once in a while ○	Often ○	Almost always ○
10. This teacher knows _____ about <SUBJECT> than my previous <SUBJECT> teacher?	A lot less ○	Less ○	More ○	A lot more ○

---

[12] Surveys were adjusted for subject and grade level for each class

11. My teacher is able to answer my questions about \<SUBJECT\> ____ often than my previous \<SUBJECT\> teachers.	A lot less ○	Less ○	More ○	A lot more ○
12. How fair are the rules for the students in this class?	Very unfair ○	Slightly unfair ○	Slightly fair ○	Very fair ○
13. On most days, how pleasant is your teacher's mood?	Very unpleasant ○	Slightly unpleasant ○	Slightly pleasant ○	Very pleasant ○
14. How often do students behave well in this class?	Almost never ○	Once in a while ○	Often ○	Almost always ○
15. The rules in this class are ____ than my other teachers' rules this year.	Much more unfair ○	Slightly more unfair ○	Slightly more fair ○	Much more fair ○
16. My teachers' mood is _____ compared to my other teachers this year.	Much more unpleasant ○	Slightly more unpleasant ○	Slightly more pleasant ○	Much more pleasant ○
17. The students in this class are ____ than the students in my other classes this year.	A lot less behaved ○	Less behaved ○	More behaved ○	A lot more behaved ○
18. How often does this teacher encourage you to do your best?	Almost never ○	Once in a while ○	Often ○	Almost always ○
19. Overall, how high are this teacher's expectations of you?	Not high at all ○	Slightly high ○	Quite high ○	Extremely high ○
20. This teacher encourages me to do my best _____ than my other teachers this year.	A lot less ○	Less ○	More ○	A lot more ○
21. This teacher's expectations are ____ than my other teachers' expectations.	Much lower ○	Lower ○	Higher ○	Much higher ○
22. Compared to your other teachers in \<SUBJECT\>, the work in this class is ____ challenging.	A lot less ○	Less ○	More ○	A lot more ○
23. In this class, how much do you participate?	Not at all ○	A little bit ○	Quite a bit ○	A tremendous amount ○
24. Overall, how interested are you in this class?	Not at all interested ○	A little bit interested ○	Quite interested ○	Extremely interested ○
25. Overall, how interesting does this teacher make what you are learning in this class?	Not at all interesting ○	A little bit interesting ○	Quite interesting ○	Extremely interesting ○
26. Compared to your other classes this year, how much do you participate in this class?	A lot less ○	Less ○	More ○	A lot more ○
27. This teacher makes what we are learning _____ than my other teachers in this \<SUBJECT\>.	A lot less interesting ○	Less interesting ○	More interesting ○	A lot more interesting ○
28. Compared to my other teachers in this \<SUBJECT\> this teacher makes learning this subject___ to learn.	A lot less fun ○	Less fun ○	More fun ○	A lot more fun ○

29. How interested is this teacher in what you do outside of class?	Not at all interested ○	A little bit interested ○	Quite interested ○	Extremely interested ○
30. If you walked into class upset, how concerned would your teacher be?	Not at all concerned ○	A little bit concerned ○	Quite concerned ○	Extremely concerned ○
31. How approachable is your teacher outside of class?	Not at all approachable ○	A little bit approachable ○	Quite approachable ○	Extremely approachable ○
32. This teacher is ____ interested in what I do outside of class than my other teachers this year.	A lot less interested ○	Less interested ○	More interested ○	A lot more interested ○
33. My teacher would be _____ if I was upset than my other teachers this year.	A lot less concerned ○	Less concerned ○	More concerned ○	A lot more concerned ○
34. My teacher is ____ approachable outside of class than my other teachers this year.	A lot less ○	Less ○	More ○	A lot more ○
35. My teacher cares ___ about me than my other teachers this year.	A lot less ○	Less ○	More ○	A lot more ○
36. My teacher believes in me ___ than my other teachers this year.	A lot less ○	Less ○	More ○	A lot more ○
37. How willing is this teacher to take time outside of class to help you?	Not at all willing ○	Slightly willing ○	Quite willing ○	Extremely willing ○
38. How prepared is your teacher for class?	Not at all prepared ○	Slightly prepared ○	Quite prepared ○	Extremely prepared ○
39. This teacher is _____ to help me outside of class than my other teachers this year.	A lot less willing ○	Less willing ○	More willing ○	A lot more willing ○
40. My teacher is ____ for class than my other teachers this year.	A lot less prepared ○	Less prepared ○	More prepared ○	A lot more prepared ○
41. How often do you have classes when this teacher does not engage with you (for example, show videos or give you worksheets to do on your own)?	Almost never ○	Once in a while ○	Often ○	Almost always ○