Connecting Mentoring to Student Achievement in Alaska: Results and Policy Implications

by
Barbara L. Adams
April 12, 2010
Presented at AERA May 4, 2010

ABSTRACT
Using hierarchical linear modeling, student standardized test scores are analyzed to determine the impact of mentoring first- and second-year teachers on their students' achievement. The contrasting group used for comparison consists of experienced teachers in matched schools, grade level, and content area. The study contains data from 300 teachers in grades 4-10 (196 treatment teachers and 104 in the contrasting group) serving over 6900 students in language arts, mathematics, and science from around the state of Alaska. The dataset is split into the three content areas that were tested, and students with only one teacher per content area are included in the study. Teacher, district, school, and student demographic information are taken into account. Results show that although mentoring new teachers did not bring the students' standardized scores of new teachers up to the same level as students in veteran classes, they are much closer than expected based on past research (statistically significant but very small effect sizes) for Reading, Writing, and Science. In the case of Mathematics, students in classrooms of mentored first- and second-year teachers perform the same as those in classrooms of veteran teachers. Thus, mentoring conducted through the Alaska Statewide Mentor Project shows promising results to start closing the achievement gap typically seen between the students of new and veteran teachers.

Barbara L. Adams
Research Lead, Alaska Statewide Mentor Project
K-12 Outreach, Office of Academic Affairs
University of Alaska
PO Box 755400
Fairbanks, AK 99775-5400
907.450.8410
barbara.adams@alaska.edu

Introduction

Argument for mentoring of new teachers

Still today, new teachers around the nation are given the most difficult teaching assignments, whether that means lowest performing students in the school, a wide variety of courses leading to a high number of preparations, disproportionate number of students with behavioral problems, or a lack of resources needed to teach (Moir, Barlin, Gless, & Miles, 2009). There seems to be this historical, unwritten rite of passage that when today's veteran teachers started in the profession they had to go through those hard times and so today's new teachers ought to as well. Many inner city schools as well as those with predominantly minority students, including Alaska Native / American Indian (AN/AI) students, have high rates of teacher turnover, thus recruiting more new teachers than their suburban counterparts proportionally (Guarino, Santibanez, & Daley, 2006; Darling-Hammond & Berry, 2006; Ingersoll, 2001).

In Alaska, this is certainly the case; the many logistical and educational challenges include a vast state with most of the districts accessible only by plane, a cross-cultural experience with 16 distinct Indigenous cultural and language systems, an academic achievement gap between rural and urban students, and a high turnover rate among new teachers. Historically, teacher retention rates in the rural schools average about 78% whereas in the urban Alaska schools (more similar to suburban communities in the lower 48) the historical retention rate is closer to 90%. When considering new teachers those retention rates drop down dramatically to about 67% for rural schools and 83% in urban schools. Overall, despite many efforts, the teacher retention rate has remained at a flat average of about 86% over the last ten years in Alaska (Hill & Hirshberg, 2008). Other characteristics of rural schools play a role in the low teacher retention rates in Alaska such as culture and language considerations, working conditions, remoteness or isolation, weather, and low retention of site administrators. Many of the rural village schools are predominantly mono-culture often with teachers from another culture. With a state university system producing only about 30% of the teaching force, it's guaranteed that at least 70% of the time the teachers are from a state other than Alaska (Hill, Hill, Hirshberg, & White, 2009).

For new teachers in these challenging situations, the first year is often more about "survival" both in the classroom and out, typically at the expense of student achievement. On this note, mentoring has been receiving national attention recently as programs seek to use experimental design and statistical methods on par with scientific procedures to analyze impacts of mentoring on student achievement, teacher retention, and teacher practice. Further, qualitative analyses continue to be conducted in hopes of understanding factors that improve teacher quality and professional development in the field. Meanwhile, more states, cities, and school districts are choosing to implement mentoring and in fact mandating participation for new teachers. Although the latest results published by Glazerman, Dolfin & et. al. (2008) and the second year study by Isenberg, Glazerman, & et. al. (2009) lack evidence of impact of mentoring on student achievement, teacher retention, and teacher practice, the study itself has been called into question and has spurred other researchers to step up to the plate and conduct more quantitative studies. As Strong indicates in his latest book on mentoring, although researchers are more certain about the approaches needed to link mentoring to student achievement, there is still a lack of studies in this area that provide any real evidence (Strong, 2008, p. 89).

Despite the current lack of student achievement research in the field of new teacher mentor programs, there have been studies that considered the relationship of teacher experience to student achievement. Often one argument made by those studying teacher turnover is that new teachers cost districts more money and produce little to no return on investment (Darling-Hammond, 2003). Further, Villar, Strong, & Fletcher (2007) found that although there is little relationship between teacher experience and student achievement, there is evidence that new teachers have lower student achievement. The relationship of increased teacher effectiveness and teacher experience is most pronounced in the first three years and then tends to fall off once teachers have about four years of experience (Villar, Strong, & Fletcher, 2007).

A recently completed doctoral dissertation at UAF concludes that the higher the teacher turnover the lower the percentage of 10th grade students scoring proficient on the mathematics portion of the Alaska Standards Based Assessment. Further, there is a high positive correlation between teacher turnover and districts serving Alaska Native students. Roehl conducted correlation analyses at a district level to analyze relationships between variables for teacher turnover, student proficiency level on math assessment, school size, percent of student population reported as receiving free or reduced lunch, and the percent of student population reported as Alaska Native (Roehl, 2010).

Describe ASMP mentoring intervention

To aid in addressing the teacher retention issue and thus the student achievement gap, the Alaska Statewide Mentor Project (ASMP) was created through a partnership with the Alaska Department of Education and Early Development (EED) and the University of Alaska (UA) system. The mission of ASMP is to make more effective teachers faster in order to provide all students with a quality teacher. The two goals are to increase teacher retention and to improve student achievement through mentoring new teachers.

In the same way that the education of students is challenging in Alaska, so are both the induction of new teachers and the professional development of mentors. The ASMP uses an intensive professional development model for mentors adapted from the New Teacher Center (NTC) located in Santa Cruz, California, to train and support experienced, veteran teachers to become effective mentors. This includes ongoing training both face to face and through distance-delivered technology, as well as a developed system of collaboration and support among mentors.

ASMP is built upon three philosophical components to the intervention model: full-release mentors, standards-driven project, and use of a formative assessment system. Full-release mentors are teachers who are out of the classroom on a full-time basis, employed as a mentor for their entire set of responsibilities. A standards-driven project uses standards at each level to ground the work in observable practices, relying less on subjectivity. ASMP uses standards for teachers, mentors, and the project as a whole. The formative assessment system provides tools that guide the conversation and provide documentation and data for the teacher, mentor, and the project. Together this intervention allows mentors to develop their own skills, provide more time, focus, and energy on new teachers; and to foster a district-wide and statewide perspective on education. This in turn allows many mentors to become professional leaders in their own

communities where they continue their careers with a renewed commitment to the education profession.

Due to the limitations of resources, ASMP chooses to mentor mostly first- and second-year teachers new to the profession in core content areas including elementary, special education, language arts, mathematics, science, and social studies. These teachers are called early career teaches (ECTs) and receive services for two years. Often times, most first-year teachers work on "survival skills" so that in their second year they can start to focus more on student learning. Through the professional teaching standards aligned with the Standards for Alaska's Teachers, mentors and ECTs focus on topics that affect the classroom, their students, and the profession of teaching. In this way, whether in survival mode or progress mode, ECTs have conversations that connect ultimately to the classroom and learning needs of their students. With this model in mind, it is hoped that a mentor's work with an ECT translates over to classroom assessments, both formative and summative.

The Alaska Statewide Mentor Project began in the 2004-2005 academic year (AY05) with 22 full-time mentors serving 334 early career teachers from around the state of Alaska. The model included mentors who were teachers either "on loan" from their districts or others such as recently retired contractors. During the first four years, research focused predominantly on ensuring the model was receptive to the needs of the early career teachers, the districts, and the mentors. Focus groups of mentors provided qualitative information to improve logistics, training, and communication for the project as a whole. Follow-up interviews were conducted with early career teachers during the summer to gather more detailed information on the benefits and challenges of the mentoring model and to better understand the effects of the induction. Online surveys were conducted each year in March to gather logistical, intervention, and perception data from early career teachers, mentors, and site administrators (Parker Webster & Whiteley, 2005; Parker Webster, 2006). Teacher retention information was gathered each year and verified by districts as well as through a partnership with the Institute of Social and Economic Research (ISER) at the University of Alaska Anchorage who access employment data from the Department of Labor and EED.

The typical implementation begins with recruiting experienced, expert teachers to become statewide mentors. Mentors live in their own communities around the state and come together in Fairbanks for training during eight academies—adapted from the NTC model—each academy lasting three days and staggered throughout mentors' two years with the project. Additionally, two days surrounding each academy are used for building the mentor learning community by training mentors on state initiatives, exploring computer applications and technology, sharing research updates, and gathering program data for constant project refinement. While the four academies in the first year tend to focus on learning how to use the formative assessment tools used for both guiding conversations as well as documenting work, the second-year set of four academies deepens mentors' understanding of the data and how to better facilitate learning on the teacher's part. While developing mentor skills, each ASMP mentor communicates weekly with all ECTs through email, phone, or Skype and visits them face to face once each month for about half a day. This is the equivalent face-to-face time of one hour a week, four weeks a month, as done in California. Mentors carry a caseload of about 15 ECTs who may be located at anywhere from 3 to 7 different sites (schools or villages) around the state. Often times, an ASMP

mentor has some ECTs located close to where they reside themselves as well as others who most often can only be visited by plane or, in a few cases, by road system. In between academies, mentors attend ongoing professional development three hours every two weeks through Elluminate Live, an online classroom environment that allows mentors to speak, chat, and collaborate on a shared whiteboard. Further, ASMP's master mentors are also certified NTC trainers who shadow and provide guidance and support to the other mentors. A few mentors remain in the project for more than two years, but the majority returns to their schools or take other leadership positions within education around the state.

By the 2007-2008 academic year (AY08), the project model described above was well established, districts welcomed mentors into their schools, increases in teacher retention were documented for those receiving services[1], and it was time to turn research efforts towards student achievement.

A small student achievement study was conducted at the end of AY08 using a controlled quasi-experimental design between ASMP (mentored early career) teachers and non-mentored veteran teachers of fourth- and fifth-grade students in urban districts. The unit of analysis was gain in scale score on the Alaska Standards Based Assessments (SBAs) in Reading, Writing and Math from FY07 to FY08. The study included seven early career teachers (1-2 years of experience, averaging 1.16 years). The comparison group consisted of four veteran teachers (4-8 years of experience, averaging 6.03 years) from similar schools and districts as the ASMP teachers. The veteran teachers were asked to complete a short demographic form and the district provided student class lists linking students to teacher. The seven ASMP teachers in this study participated fully in the mentoring throughout that year, supplied demographic information to their mentors, and the mentors obtained class lists from the districts. Student scores were obtained from EED once supplied with the class lists. Preliminary teacher-level results (a conservative approach to analyzing this type of data with such small sample sizes), show students taught by mentored early career teachers achieving gain scores on SBAs similar to students taught by veteran teachers. Gains in Reading scores for students of ASMP teachers were 5.3 compared to 9.0 for veteran classrooms; Writing 2.1 vs. -1.0; Math -6.8 vs. -5.5.

In each case, the results are not statistically significant (all p-values >0.05, specifically 0.91, 0.14, 0.96), meaning that the small study found no difference in average classroom gain scores between mentored early career teachers and veteran teachers. The models produced results with R^2 values of 0.212, 0.392, and 0.113 respectively, showing that other variables beyond participation in ASMP and years of experience are needed to help describe the variation in data. Despite the limitations, the results of this small study were promising and provided ASMP with enough evidence to attempt a larger scale study linking mentoring of teachers to student achievement (Adams, 2008).

Purpose and Rationale for the Study

Given the low teacher retention rates in Alaska, the connection between new teachers and lower academic achievement throughout the nation, and the promising results from the small-scale study, a larger study was commissioned to further investigate the link between mentoring by

[1] ASMP, Research Summary 2004-2008 contains teacher retention updates and the description and results of the small exploratory student achievement study (Adams, 2008).

ASMP to Alaska students' achievement on standardized assessments. In essence, the null hypothesis is that mentoring early career teachers will close the gap between their students' standardized test scores and those of a contrasting group composed of veteran teachers.

Method

Participants

In the 2008-2009 academic year (AY09), ASMP trained 27 full-time mentors serving 434 early career teachers who were located in 37% of the schools (185 schools out of 506 total in the state) within 70% of the districts (38 districts out of the total 54 districts) in the state of Alaska. Districts choose to invite ASMP mentors into their schools to work with their early career teachers at no cost to the district.

The ASMP teachers in the study are located within 30 of the school districts who participated in AY09. Contrasting veteran teachers were recruited based on comparability to ASMP teachers using school characteristics, content area, and grade level on a district-by-district basis. Of the 434 early career teachers served, 196 satisfied the criteria for the student achievement study. The remaining teachers may not have been responsible for language arts, mathematics or science instruction; may have been teaching grades K-3 or grades 11-12, or may have been in districts unable to provide the class lists needed to group students with teachers for the HLM analysis. The distribution of teachers and students is presented here by demographic categories using the total dataset.

Gender: Males constitute 51.9% of the students, 47.7% are females with 0.4% missing data. At the teacher level 42.0% are male, 58.0% female.

Grade level: About 25% of the students fall into the elementary grades 4-6, 35% are considered junior high grades 7-8, and the remaining 40% are high school students in grades 9-10.

Special Education: There are 1208 special education students total (13.7% of the student pop), 106 (8.7%) are in special education treatment classrooms (early career teachers with an ASMP mentor), 762 (63.1%) are in treatment classrooms of early career teachers who are not special education and 340 (28.1%) are in veteran teacher classes who are not special education teachers. There are no special education teachers in the contrasting group. In total, there are only 1.8% of students in classrooms of special education teachers. This discrepancy could be for several reasons. Data of students for some special education teachers may not have been provided if the students were in other classrooms, having another teacher of record for the content areas.

School Location: There are 27% of teachers in urban districts, as defined by the state as the largest five districts, compared to 37% urban at the student level. Thus the majority of teachers are in rural schools, 73%, as well as the majority of students, 63%. Although the urban/rural category is used often, breaking down this category into school location shows a more revealing picture: urban, rural off the road system, rural hubs, bush schools (off road, out of hubs). Typically rural schools on the road system tend to have higher achievement than their more remote counterparts in bush Alaska. Also, in urban districts there are schools that are also more remote and thus tend to score more like rural schools. Using this new category, an equivalent percentage of students are located in bush 36% and urban 35.5% schools. Smaller numbers of

students are in rural hub villages, 23.2% and yet smaller numbers are in schools on the road system but not considered urban schools, 5.4%. However, at the teacher level, the majority are in the bush, 59%, thus small class sizes. There are 19% of teachers in urban which comprised 35% of students, thus showing large classes. There are 16% of teachers in rural hub villages and 5.6% on the road system not urban, about the same as students, so about average-sized classrooms considering those contained in this data set.

AACP Principals: Only 22.1% of the students are in schools with new principals in the AACP program, meaning those principals have an assigned principal coach. There may be more students in schools with new principals but they are not in the AACP program - they could be in other programs or they may not be in a program. At the teacher level, 21% of teachers are in schools of new principals who are in the AACP project. The similar finding between student and teacher level here shows that most of the AACP principals are in schools of average size within this dataset.

Teacher Years of Experience: The ASMP treatment teachers range from 1 to 2 years of experience with an average of 1.5 years. The contrasting veteran teachers range from 3 to 30 years of experience with an average of 12.2 years. A small number of teachers in the treatment and contrasting groups did not satisfy the criteria of teacher years of experience (for example, a treatment teacher with six years of experience or a veteran teacher with only one year of experience). Based on sensitivity analyses, the teachers and their associated students were removed from the data.

Procedure

Assignment

The research design did not include randomization as the population of interest forced assignment based on certain criteria. At the time ASMP was not in a position to be able to randomly assign early career teachers to receive mentoring or not receive mentoring, nor was that the intention of the project. With the high teacher turnover and struggling schools in rural Alaska, it was more desirable to first investigate a quasi-experimental design that employed a high level of matching to understand the difference between groups who were similar on many characteristics except for years of experience and the intervention. Thus, this study does not use a typical treatment and control design but a treatment intervention compared to a contrasting group. Despite this limitation, data were gathered in a rigorous manner that allowed for a high level of statistical analysis to still be used, hierarchical linear modeling.

Since a typical randomized controlled trial or quasi-experimental design is not feasible, the treatment group is defined as those ECTs participating in ASMP, teaching in reading, writing, mathematics and/or science grades 4-10, within districts who could provide the class lists. ASMP asked for volunteers of experienced teachers who were as similar as possible along those same traits and with characteristics described below within each district. Those recruited teachers form a “matched” group of experienced teachers to serve as the contrasting group. This design allows examination of whether the intervention enabled ECTs to achieve gains in students’ achievement comparable to the experienced teachers (the contrasting group), after controlling for other differences.

Contrasting teachers were recruited on a district basis using the following characteristics.

- Content areas of language arts, mathematics, and science. (Note: special education veteran teachers were not recruited.)
- Years of experience: recruited teachers in their third year or higher. (Note: since ASMP started in AY05, a second year teacher could have received mentoring through the project, remained teaching in the state, and would have been in their sixth year of teaching during this study. Taking this into account, data were linked from the project to the veteran teachers, identifying any who may have received ASMP mentoring in the past. Only 7 out of 104 (only 6.7%) veteran teachers were previously served by ASMP.)
- District or urban/rural or school location: recruited based on matching school type demographics identified by district personnel. For small districts without those teachers, the match was done across similar districts (for example, single site districts) upon acceptance of both districts

Intervention and Data Collection

Data gathered from school districts included teacher class lists for language arts (Reading and Writing), Mathematics and Science. The student information contained identification numbers needed to access their achievement data from the state's database, as well as demographic information such as gender, grade level, date of birth, and whether they were considered special education. The class lists were submitted to EED to obtain SBA data from 2008 and 2009 as well as a check on gender and special education classification.

Teachers were considered treatment or contrasting based on the criteria of whether they were early career teachers working with an ASMP mentor (treatment) or veteran teachers with three or more years of experience (contrasting). Teacher data were gathered through a short online demographic form and an incentive of 25,000 Alaska Airline miles were raffled off for the group of contrasting teachers completing the form. Teachers were also identified with a district code, and whether they taught in a school with a new principal who was receiving coaching from the Alaska Administrators' Coaching Project (AACP), a similar project designed for site administrators. The few teachers with less than three years of experience in the contrasting group were eliminated from the study; however, their raw data verified the findings from the literature that new teachers (receiving no intensive mentoring services) tend to have students performing much lower on standardized assessments than veteran teachers. There were about 40% missing data concerning the degree-granting institution for the contrasting teachers, and so that variable was eliminated. There are no missing data at the teacher level for the other variables used in these models.

The results of the matching process between ASMP teachers and the contrasting veteran teachers along with their subsequent student populations are shown in Table 1. The original criteria of recruiting within districts (or matching across similar small districts) provided roughly the same distribution of urban/rural districts as well as about the same school characteristics based on location with slightly higher percentage of rural off the road system schools (lower rural hub schools) in the treatment group as compared to the contrasting group. This is most likely due to ASMP abilities to serve fewer teachers in larger schools, which is the case with schools typically found in rural hubs. The proportion of ASMP teachers in each of the content areas is about 62-64% of the total population, with similar proportions of students, 62%-66%, showing the

recruitment of contrasting teachers fell short of producing a balanced sample. The years of experience, another measure of the treatment and contrasting groups based on design, show equal number of first- and second-year teachers in the treatment group and an average of about 12 years and a standard deviation of 7.5 years experience in the contrasting group. Below the darker line in the table are the results of the variables that were not used in recruitment but support that these groups are equivalent in many important variables except for their years of experience. There are about the same percentage of male and female teachers in the treatment and contrasting groups as well as in the related student groups. There are slightly more junior high students in the contrasting group than the treatment and less elementary. The percentage of students with individualized education plans in 2009 (IEP09), signifying special education services, is about the same between the groups, within a couple of percentages. ASMP teachers do have a higher proportion of new principals in the AACP program which aligns with the theory that retaining principals is similar to retaining teachers in the schools served by ASMP. The major difference between treatment and contrasting groups is the average scaled scores from 2008. In Reading, Writing and Math the difference is about 0.375 standard deviations. This also confirms that continued assumption that many beginning teachers are given the low performing students or are assigned to more difficult teaching situations.

Table 1: Results of matching teachers during recruitment

	Treatment: ASMP Teachers	***Contrasting: Veteran Teachers***
Percent of teachers in a rural school district	74.5% (146/196)	71.2% (74/104)
School Location:		
Urban	18.9% (37/196)	19.2% (20/104)
Rural on the road system	4.6% (9/196)	5.8% (6/104)
Rural hub	13.8% (27/196)	21.2% (22/104)
Rural off the road system	62.8% (123/196)	53.8% (56/104)
Content Area:		
Reading	2621 students, 144 teachers	1380 students, 82 teachers
Writing	2618 students, 144 teachers	1388 students, 82 teachers
Mathematics	2267 students, 130 teachers	1387 students, 76 teachers
Science	2650 students, 120 teachers	1387 students, 74 teachers
Years of Experience	Mean: 1.5 years SD: 0.5 years	Mean: 12.32 years SD: 7.49 years
Teacher Gender	56.6% female (111/196)	59.6% female (62/104)
Student Gender[2]	47.2% female ± 0.95%	48.3% female ± 0.5%

[2] Student gender distributions varied slightly over the three content area datasets providing the mean with error estimates.

	Treatment: ASMP Teachers	***Contrasting: Veteran Teachers***
Grade Level[3]		
Elementary, grades 4-6	41.5%	39.7%
Junior High, grades 7-8	27.5%	32.4%
High School, grades 9-10	31.0%	27.9%
Student % with IEP in 2009		
Language Arts	16.5% (491/2780)	14.4% (206/1432)
Mathematics	16.3% (394/2415)	13.4% (195/1456)
Science	13.8% (384/2774)	14.0% (202/1438)
Percent of Teachers in a school with principal in AACP	24.5% (48/196)	14.4 % (15/104)
RSS08	Mean: 328.10 SD: 47.40 N: 194 teachers	Mean: 346.07 SD: 48.12 N: 104 teachers
WSS08	Mean: 308.58 SD: 52.10 N: 194 teachers	Mean: 327.97 SD: 51.78 N: 104 teachers
MSS08	Mean: 305.79 SD: 48.80 N: 194 teachers	Mean: 323.97 SD: 48.51 N: 104 teachers

The student outcome data consist of scaled scores from 2009 for Reading (RSS09), Writing (WSS09), Mathematics (MSS09), and Science (SciSS09). Covariates of the students' scaled scores from 2008 (RSS08, WSS08, MSS08) were used in each model. At the student level, there is about 5.1% missing outcome (RSS09, WSS09) and 6.3% missing pre-test (RSS08, WSS08) student data for the Reading and Writing scaled scores. According to Puma, Olsen, Bell, & Prince (2009), these are the lower limits of what is usually missing and thus implementing the method of dropping missing data produces typically low bias for the impact estimate and low bias for the standard error of the impact estimate. Similarly, there is about 5.6% missing outcome (MSS09) and 7.1% missing pre-test (MSS08) for the Mathematics scaled scores. And for the Science data, the outcome variable, SciSS09 has 8.3% missing data and the covariates from 2008 (RSS08, MSS08) have about 7.5% missing (note that it is a different set of students, those associated with science teachers and so the value is not the same as the language arts or mathematics datasets). This is still considered low and thus dropping cases with missing data is an appropriate method. Further, in all cases there is no difference in the rate of missing data based on treatment or contrasting groups that would introduce bias. There was negligible missing demographic data for students, less than 1%.

[3] Distributions of grade level varied across the content area datasets and are averaged here. For students in contrasting veteran classes, the elementary distribution ranged from 38.7% Math to 41.3% LA; junior high ranged from 25.2% Science to 38.7% Math, and high school ranged from 22.5% Math to 35.6% Science. For the students in the treatment ASMP classes, the elementary ranged from 38.3% LA to 44.2% Math, junior high 16.3% Math to 34.1% LA, and for high school 25.8% Science to 39.5% Math.

Measures

The state of Alaska has created and administered the Alaska Standards Based Assessments since Spring 2005. The assessments are given to students in grades 3-10 on content areas of Reading, Writing, Mathematics. In Spring 2008 the first round of Science SBAs were also administered to students in grades 4, 8, and 10 only. EED computes scaled scores from raw scores for each test at each grade level creating a common standard score used for proficiency measurements. The scores range from 100 to 600 and the cut-off for proficiency is 300 for each test. The scores and proficiency levels were validated through a process involving teacher and administrator input in the early stages. Although the tests are not vertically aligned, EED states, "Thus, a student who receives a scale score of 300 at each grade is making progress from grade to grade that exactly equals the difference in the standards for Proficient across those two grades" (EED Technical Report, p. 53). Since the scaled scores at each grade level indicate the level of the students' performance relative to the standards for that grade, the data collection allowed for grouping of all scaled scores across grade levels. This assumption was tested by analyzing the dataset for the 2009 scaled scores and the 2008 scaled scores in Reading, Writing, and Mathematics independently showing that the distributions across grade levels followed the same patterns.

Analytic Approach: HLM

Four separate null hypotheses were tested all following the same format. If the mentoring intervention is successful, students of ASMP mentored early career teachers will score similar to students of contrasting veteran teachers on the Alaska Standards Based Assessments (SBA) taking into account students' scores from the previous year. Thus, the intervention will be considered effective if the difference between the treatment and contrasting teachers in students' achievement scores is not statistically significant.

Data were entered, organized, coded, and cleaned using the statistical software SPSS and then imported into HLM Software for modeling. The HLM text by Raudenbush and Bryk (2002) was also used as a reference.

To address the null hypotheses, four separate models were conducted using outcomes of scaled scores on (a) Reading and (b) Writing using only teachers who were assigned to teach language arts, (c) Mathematics using only teachers assigned to mathematics classes, and (d) Science using only teachers assigned to teach science. Often at the elementary level the same teacher may belong to each of the three datasets and thus may be contained within each of the four models. The Benjamini-Hochberg adjustment is applied at the end to take into account multiple comparisons using the same dataset.

The final HLM model[4] has the following properties:

[4] Additional HLM analyses were performed in an attempt to create the best model possible that represented the design and data well. The district variable was recategorized in two ways: urban and rural and by four school locations (urban, rural on the road system, rural hub off the road system, or bush - rural off the road). It was determined that with matching of contrasting teachers done for recruitment at the district level, the district variable was most indicative of the nature of the design. Further, about 7% of the student data in each case were assigned to more than one teacher. To address this, each student was recoded with a teacher 1 and teacher 2 identifier. Running a cross-classification HLM model placed more weight on those students with a single teacher and did not seem to represent the structure of the data well, thus the decision to remove all students assigned to multiple teachers.

1. Removes all students in multiple classrooms, so the level-1 population is only for students with one teacher for language arts (mathematics or science), ranging from grades 4-10, from the participating districts.
2. Includes controlling for the following level-1 variables: student gender, student grade category (elementary, junior high, or high school) and if the student had an IEP in 2009.
3. Includes a level-1 covariate of the students' corresponding SBA score from 2008 (Reading, Writing, or Mathematics). Note that science assessments are only given in 4th, 8th and 10th grades, so no 2008 Science scores were available as covariates; rather, Reading and Mathematics from 2008 were used once found to be highly correlated with Science outcomes ($r = 0.79$ and 0.72 respectively).
4. Controls for the following level-2 variables:
    a. teacher gender
    b. whether the teacher is special education certified
    c. whether the teacher is in a school where there is a new principal enrolled in the AACP
    d. school district: Reading and Writing models included 29 districts with one falling out from missing data, the Mathematics model included 29 districts with a different one falling out from missing data, and the Science model included 24 districts with six districts falling out from missing data

Results

Overall Impact

Reading

- Controlling for school district, teacher gender, special education certification, and principal participation in AACP at the teacher-level and gender, grade category, special education classification at the student-level and the student's Reading scaled score from 2008, there is a statistically significant difference between the treatment and contrasting groups on the Reading scaled score for 2009 ($p = 0.037$). In fact, the ASMP teachers have average student Reading scaled scores about 4.7 points lower than students in the contrasting group of veteran teachers.
- The difference between average Reading scaled scores of students within classrooms of ASMP teachers and the contrasting veteran teachers produces an effect size of 0.06 (4.7 / 73.36 = 0.06), which is very small as determined by Cohen's rule of thumb (Gliner & Morgan, 2000, p. 178). Even when using a teacher-level standard deviation to calculate the effect size, the results remain very small: 4.7 /46.93 = 0.10.
- Practically, on the standard-based assessments designed for Alaska as found in the *Spring 2006 Alaska Standards Based Assessments (SBAs) Operational and Field Test Technical Report* (page 53), scoring at 300 for a scaled score is considered proficient for each grade level. The tests all have approximately a 75 point standard deviation, so to reach an effect size of any meaning—even a small one such as 0.20—the average difference needs to be at least 75*0.20=15 points. Here a difference of nearly 5 points, though statistically significant, is still small.

Writing

The results are summarized for all four models in Table 2. The model for Writing produced similar results as for Reading. The ASMP teachers have average student writing scaled scores about 5.5 points lower than students in the contrasting group of veteran teachers, which is statistically significant ($p = 0.038$), but a small effect size of 0.07 at the student level and 0.12 at the teacher-level.

Mathematics

The model for Mathematics also produced similar results. The ASMP teachers have average student mathematics scaled scores about 7.0 points lower than students in the contrasting group of veteran teachers, which is statistically significant ($p = 0.023$), but again a small effect size of 0.06 at the student level and 0.12 at the teacher-level.

Science

The model for Science also produced slightly similar results. The ASMP teachers have average student science scaled scores about 8.2 points lower than students in the contrasting group of veteran teachers, which is statistically significant ($p = 0.023$), but a small effect size of 0.10 at the student level and 0.17 at the teacher-level.

Table 2: Summary of Results from HLM Models[5]

	Reading	**Writing**	**Mathematics**	**Science**
outcome variable	RSS09	WSS09	MSS09	SciSS09
difference in scores between ASMP and contrasting group	-4.7	-5.5	-7.0	-8.2
p-value	0.037	0.038	0.023	0.023
student-level effect size	0.06 4.7 / 73.36	0.07 5.5 / 74.95	0.09 7.03/76.32	0.11 8.2/ 76.56
teacher-level effect size	0.10 4.7 /46.93	0.12 5.5 / 47.81	0.15 7.03 /47.34	0.17 8.2 /49.67

Benjamini-Hochberg Adjustment

Due to multiple comparisons in the student achievement domain, the Benjamini-Hochberg adjustment was applied to the results. The procedure starts by ordering the null hypotheses in terms of the smallest p-value to the largest. The criterion tests if the p-value is smaller than increments of a quarter of the alpha-level (since there are four hypotheses). Since $p=0.023$ for both Mathematics and Science, which is not smaller than $0.05/4=0.0125$, then at least one of the hypotheses is no longer significant. Upon analysis three of the four results remain statistically significantly different, but the fourth result, either Science or Mathematics, does not (Benjamini & Hochberg, 1995). The criterion does not address handling of tied p-value scores and so the choice is arbitrary. It seems with the Mathematics results showing a smaller difference related to a smaller effect size, it is more logical to state that with the adjustment the Mathematics scores of

[5] Covariates were used in each model: RSS08 for Reading, WSS08 for Writing, MSS08 for Mathematics and RSS08 and MSS08 for Science.

students in the ASMP treatment teacher classes is not statistically significantly different from those in the contrasting veteran teacher classes.

Subgroups

The research design was not set up to test the difference between subgroups based on any of the teacher- or student-level variables, so whether they are significant in this model only supplies motivation for an exploratory analysis. Breaking out data by district would violate the agreement for research. The data is too unbalanced to look at special education students or teachers compared to the others. Another variable of interest, AACP, is also too unbalanced to proceed with that type of an analysis. For all four models, the teacher gender is not significant and thus an analysis may not provide much information. The student gender is statistically significant in the Writing and Science models and could be of interest for future exploratory analysis.

Discussion

General summary

There is a statistically significant difference between Reading, Writing, and Science scores of students in early career teachers' classes and those in contrasting veteran teachers' classes. This is true for standardized scaled scores once controlled for student demographics, teacher demographics, and student scaled scores from the previous year. For the Mathematics scaled scores there is no statistically significant difference between students in classrooms of early career teachers and those with veteran teachers, once adjustments were made for multiple comparisons.

The effect sizes of the difference in scores for Reading, Writing and Science is much smaller than expected from the literature review and evidenced by small subsamples of first-year teachers receiving no mentoring. Further, in a study conducted by Rockoff (2004) he states, "I also find evidence that teaching experience significantly raises student test scores, particularly in reading subject areas. Reading test scores differ by approximately 0.17 standard deviations on average between beginning teachers and teachers with ten or more years of experience" (p. 248). Rockoff analyzed teacher quality, and one characteristic being years of experience, and its relationship to student achievement through a meta-analysis approach varying across years for individual teachers. The effect sizes found within this study for those differences that were statistically significant were a fraction of what Rockoff found in his analysis (for example, 0.07 standard deviations for Writing compared to 0.17). Even the teacher-level effect sizes are slightly less than those found by Rockoff, especially when comparing across Reading scores, here 0.10 compared to 0.17.

Even after adjusting for multiple comparisons, three out of four differences are statistically significant, which means that the intervention was not completely successful in eliminating the gap. However, with effect size differences between ASMP mentored teachers and experienced teachers smaller than differences found previously between new and experienced teachers, ASMP mentoring of first- and second-year teachers shows promise for closing the achievement gap commonly experienced by students of beginning teachers. For a quasi-experimental design without randomization, this rigorous study is strengthened by using state standardized assessments that carry high levels of internal and external validity, by having a small amount of

missing student data, and by recruiting a contrasting group that was similar to the treatment group in multiple ways (excluding pre-test scores).

Limitations

From a statistical point of view, this study does not answer the question, "Does mentoring new teachers work?" To answer this in a definitive manner, a true comparison control group of new teachers who do not receive mentoring, especially in the rural school districts of Alaska, is needed. Due to the persistent achievement gap between Alaska's rural and urban school districts and the long-term low teacher retention in rural schools, ASMP and the State of Alaska are not willing to withhold mentoring from any of those districts, schools, or teachers who request it. Further, without random selection from the larger population of teachers within the state, this study does not generalize beyond the group involved. If it could be shown that the teachers in this study are comparable to the larger population of teachers, then it may be possible that this study is likely to be a good guide to the potential effects of mentoring ECTs statewide. However, at this time, access to the necessary data to conduct such a comparison is limited.

Conclusion

The Alaska Statewide Mentor Project is the only fully funded, non-mandated, statewide induction program in the nation. This means that the state of Alaska and the University of Alaska supply all funding for the project, requiring no financial obligations on the part of the school districts. Resources received through that allocation allow only 55% of early career teachers in rural school districts and 10% in urban districts to receive services from ASMP. New teachers in rural districts served by ASMP who are *not* in core content areas are already *not* receiving services. To this end, ASMP continues to look for funding that would allow all first- and second-year teachers, new to the profession, in both rural and urban districts to be mentored. Although ASMP has improved teacher retention within the small subsample served, extending services to include not only all first- and second-year teachers new to the profession but also experienced teachers new to Alaska might increase the teacher retention rate for the state. The ultimate goal is for the impact of mentoring on teacher retention to continue to positively impact student achievement for all Alaskan students.

In order to focus on student achievement, a full randomized controlled trial should be conducted. Currently, ASMP serves so few urban teachers despite the bulk of new teachers being hired by urban districts. Thus, in urban districts, it might be feasible to use random assignment to determine which new teachers receive mentoring. Although teacher turnover tends to be much lower in these regions, student achievement issues remain a focus for most of the districts. Alaska has five larger districts that are considered "urban," but most truly tend to be more aligned with suburban situations in the lower 48 states. These five districts—Anchorage, Mat-Su Borough, Fairbanks North Star Borough, Kenai Peninsula Borough, and Juneau School District—encompass a variety of school situations ranging from typically meeting AYP to struggling to meet AYP for the last five years. Further, school sizes vary from quite small to the largest in the state. In contrast to many of the rural village schools where the students are mono-culture, often with a teacher from another cultural background, some of these urban districts have over 80 different cultural and language groups.

Mentoring through ASMP is a very promising intervention. With 73% of the teachers in this study serving in rural school districts within Alaska, the results from this study are impressive. Consider the situations in which most of these first- and second-year teachers find themselves. The majority choose to move to a rural, often times remote, location in Alaska where access to the village may be by plane or boat only (and this is true for some schools in the urban districts as well). Wherever they are located, these early career teachers experience extreme weather situations such as, temperatures around -40° F, limited sunlight, eight months of snow and winter or possibly horizontal winds and rains for extended periods of time. Most are in culturally different villages from their own background. Among the Alaska Native villages transitional language issues run the gamut from little to no native language to broken English to broken English and broken native language to fully functioning bilingualism. Districts struggle with high teacher and administrator turnover, ongoing curriculum changes, and struggling school boards. Many of the schools are on plans of improvement under the No Child Left Behind Act. These early career teachers are placed into the most challenging schools, communities, and classrooms. If they survive in the profession, they gain skills at the expense of the students the first few years. They then shift into classroom situations such that the makeup of the students is often times less challenging due to, for example, parental requests, negotiated agreements that allow seniority "benefits," and having a role in determining class lists. In light of this, it is easy to see how many of the early career teachers in Alaska actually begin their careers in the most difficult educational settings within the country. If this study took place with teachers in well supported situations in which factors existed that typically bolstered student achievement, it may be that the results would seem minimal. However, given the circumstances of the teachers in this study, the results do start to answer the question "Does mentoring make a difference?" The results here coupled with the less than ideal situations in which new teachers in rural Alaska and their students find themselves leaves one to believe that mentoring new teachers is making advancements in closing the achievement gap between students of new teachers and those with veteran teachers. These results give a clear indication that the question continues to be worth pursuing.

REFERENCES

Adams, B.L. (2008). Research summary 2004-2008. *Alaska Statewide Mentor Project, Internal Report*.

Alaska Department of Education and Early Development. (2006). *Alaska standards based assessments (SBAs) operational and field test technical report*.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1), 289-300.

Darling-Hammond, L. (2003, May). Keeping good teachers: why it matters, what leaders can do. *Educational Leadership, 60*(8), 6-13.

Darling-Hammond, L., & Berry, B. (2006, November). Highly qualified teachers for all. *Educational Leadership*, 14-20.

Glazerman, S., Dolfin, S., Bleeker, M., Johnson, A., Isenberg, E., Lugo-Gil, J., Grider, M., Britton, E., & Ali, M. (2008). Impacts of comprehensive teacher induction: results from the first year of a randomized controlled study. *Report to U.S. Department of Education*, NCEE 2009-4034

Gliner, J.A., & Morgan, G.A. (2000). *Research methods in applied settings: an integrated approach to design and analysis*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc..

Guarino, C.M., Santibanez, L., & Daley, G.A. (2006). Teacher recruitment and retention: a review of the recent empirical literature. *Review of Educational Research*, *76*(2), 173-208.

Hill, A., & Hirshberg, D. (2008). Turnover among Alaska teachers: is it changing? *Institute of Social and Economic Research, University of Alaska Anchorage*. Research Summary No. 69.

Hill, M., Hill, A., Hirshberg, D., & White, G. (2009). Alaska's university for Alaska's schools. *Report to the State of Alaska*, SB241.

Ingersoll, R.M. (2001). Rejoinder: misunderstanding the problem of out-of-field teaching. *Educational Researcher*, *30*, 21-22.

Isenberg, E., Glazerman, S., Bleeker, M., Johnson, A., Lugo-Gil, J., Grider, Dolfin, S.,M., Britton, E., & Ali, M. (2009). Impacts of comprehensive teacher induction: results from the second year of a randomized controlled study. *Report to U.S. Department of Education*, NCEE 2009-4072.

Moir, E., Barlin, D., Gless, J., & Miles, J. (2009). *New teacher mentoring: hopes and promise for improving teacher effectiveness*. Cambridge, MA: Harvard Education Press.

Parker Webster, J. (2006). Alaska statewide mentor project NTC 2006 online induction survey report. *Internal Report*.

Parker Webster, J., & Whiteley, G. (2005). Alaska statewide mentor project research report: initial implementation year 2004-2005. *Internal Report*.

Puma, M.J., Olsen, R.B., Bell, S.H., & Prince, C. (2009). What to Do when data are missing in group randomized controlled trials. *Report to U.S. Department of Education*, NCEE 2009-0049.

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: applications and data analysis methods. Second edition*. Thousand Oaks, CA: Sage Publications, Ltd.

Rockoff, J.E. (2004). The Impact of individual teachers on student achievement: evidence from panel data. *American Economic Review*, *94*(2), 247-252.

Roehl, R.F. (2010). *Correlation between teacher turnover rates in the state of Alaska and standardized test scores in the area of Mathematics on the Standards Based Assessment/ High School Qualifying Exam*. Fairbanks, AK: Ph.D. Dissertation.

Strong, Michael. (2008). *Effective teacher induction and mentoring: assessing the evidence*. Teachers College Press.

Villar, A., Strong, M., & Fletcher, S. (2007). The Relation between years of teaching experience and student achievement: evidence from five years of value-added English and Mathematics data from one school district. *NTC Working Paper*.