Houston Independent School District's ASPIRE program: Estimated Effects of Receiving Financial Awards

Dr. Dara Shifrer, Houston Education Research Consortium, Kinder Institute for Urban Research, Rice University, dara.shifrer@rice.edu

Dr. Ruth López Turley, Houston Education Research Consortium, Kinder Institute for Urban Research, Rice University, turley@rice.edu

Dr. Holly Heard, Houston Education Research Consortium, Kinder Institute for Urban Research, Rice University, hheard@rice.edu

We are grateful to staff from the Houston Independent School District's Research and Accountability Department, particularly Dr. Laurie Zimmerman, Carla Stevens, and Luellen Jenkins-Bledsoe, for providing data for our use and for being available to answer questions throughout the course of this evaluation. In addition, statistical advice was provided by Dr. Rose Medeiros at Rice University, and this evaluation was reviewed externally by Dr. Steven Glazerman at Mathematica Policy Research, Inc. and Dr. Scott Imberman at Michigan State University, whose feedback was extremely helpful. Finally, the Houston Education Research Consortium is funded by the Laura and John Arnold Foundation. Any opinions expressed in this report reflect those of the authors and not necessarily those of the district, the reviewers, or the funding agency.

TABLE OF CONTENTS

List of Tables

List of Figures

Executive Summary

The Houston Independent School District (HISD) asked the Houston Education Research Consortium (HERC) at Rice University to conduct an independent evaluation of HISD's educator award program,[1] ASPIRE (Accelerating Student Progress: Increasing Results and Expectations), which pays out over $40 million each year to employees (Zimmerman et al. 2011). Award programs are a policy intervention aimed at increasing student achievement by rewarding educators financially. In order to examine whether ASPIRE effectively incentivized teachers, we would need to compare teachers who were eligible for ASPIRE to similar teachers who were not eligible. However, ASPIRE was implemented in virtually all HISD schools, and the vast majority of HISD employees were eligible for an award, so it is not possible to examine the effects of eligibility. Nonetheless, it is possible to examine the effects of receiving an award among those who are eligible, which is the scope of this evaluation.

Using data collected by HISD and the Texas Education Agency (TEA), we compared the outcomes of teachers who received awards to similar teachers in comparable schools who did not receive awards. Specifically, this evaluation estimated the effects of receiving an ASPIRE award for the 2009-10 school year (formally announced and paid out in January 2011) on teachers' retention by August 2011, the change in their attendance rates from 2009-10 to 2010-11, and their mean student test score gains from 2009-10 to 2010-11.[2]

Because this evaluation focuses on the estimated effect of receiving an award, rather than eligibility for an award, findings cannot be used to understand how teachers' outcomes might differ in the absence of an award program. Even so, if outcomes appear to improve more over time for award recipients than for those who did not receive awards, this suggests ASPIRE effectively motivates teachers through award receipt. Another contribution of this evaluation is its focus on the specific aspects of ASPIRE that are more or less effective and the differing effects of receiving an award for various subsets of teachers.

Receiving an Award Improves Teacher Productivity

Among core teachers who were eligible for the 2009-10 ASPIRE award program, the 2010-11 outcomes (retention, attendance, mean student achievement gain) of teachers who received a Strand I, II, and/or III award improved more than those of comparable teachers who did not receive an award.

Larger Awards are More Effective than Smaller Awards

The outcomes of core teachers who received a larger total ASPIRE award improved more than those of comparable core teachers who received a smaller total award. Furthermore, the relative amounts of the award (relative to their salary and relative to the mean award amount on their

[1] Policy interventions like these are often referred to as teacher merit pay programs or performance pay programs. We use "award program" throughout this evaluation to maintain consistency with HISD's terminology. Similarly, we often use the term "educator" rather than "teacher" because HISD's program targets a wide variety of school employees.

[2] The state of Texas transitioned from the Texas Assessment of Knowledge and Skills to the State of Texas Assessments of Academic Readiness in 2012, preventing the estimation of gain scores from 2010-11 to 2011-12.

campus) were more salient than the raw amount of the award. Because teachers' responses to their awards appear to be most dependent on award amounts relative to those of their colleagues, these findings support the idea that teachers are more responsive to competitive awards.

Teachers in Higher Need Schools Benefit Less

However, the outcomes of core teachers in higher need schools who received an award did not improve as much as those of comparable core teachers in lower need schools who received an award. Award programs may be less effective for teachers at higher need schools because of the possibility that the value-added methodology does not sufficiently account for the unique challenges of working in a higher need school.

Hard-to-Staff Subjects: Math Teachers Respond More Positively but Special Education Teachers Respond Less Positively

The outcomes of core teachers of hard-to-staff subjects who received an award did not improve in the same ways as those of comparable core teachers not in hard-to-staff positions: Secondary level math teachers responded more positively, while special education teachers responded less positively. Like teachers in higher need schools, special education teachers may need different resources than financial incentives to address the challenges inherent to their position. Additional information gathering, through surveys perhaps, might improve our understanding of the different responses to award receipt for teachers of hard-to-staff subjects.

The Most Effective Awards Blend Competition, School-Wide Support, and Accountability Criteria

Strand I and III awards are determined entirely on the basis of campus level achievement, while Strand II awards are based on the achievement of more select groups of students. For this reason, we conceptualized Strand II awards as competitive, and Strand I and III awards as collaborative. The outcomes of core teachers were most improved by receipt of a Strand III award, improved to a lesser degree by receipt of a Strand II award, and were either unaffected or negatively impacted by receipt of a Strand I award. The strong impact of a Strand III award may result from fostering the development of school-wide initiatives, initiatives that enabled teachers to succeed, in ways that Strands I and II did not because of their focus on a more targeted set of academic goals (e.g., increased enrollment in Advanced Placement classes, math and reading test scores). Moreover, it was the only Strand to incorporate the statewide accountability criteria (e.g., school ratings from the Texas Education Agency) that are a dominant focus among school administrators. For most outcomes, core teachers benefited more from receipt of a Strand II award, if they worked on a campus in which receipt of Strand II awards was more prevalent. These schools may have contexts or structural features that enable teachers to capitalize on their personal strengths. Although these first two findings suggest teachers' preference for collaboration over competition, we also found that the outcomes of teachers who received a Strand II award on the basis of the achievement gains of their own students, or the students in their department, improved more than the outcomes of teachers who received a Strand II award on the basis of a larger group of students. This may indicate that teachers place greater value on awards that emphasize competition, or that teachers respond most positively to awards linked to outcomes they perceive to be within their control. Narrowing the focus of the award program to the most effective strands, Strands II and III, might simplify the design of the program, thereby increasing teacher understanding of and buy-in to the program.

Policy Implications

- Reduce the Number of Awards and Increase the Amounts
- Eliminate Strand I
- Provide Teachers with the Tools They Need to be Successful
- Improve the Identification of Quality Teachers
- Increase Non-Merit-Based Financial Incentives for Teachers in Hard-to-Staff Positions
- Obtain a More Informed View of ASPIRE through the Collection of Qualitative Data
- Clarify Desired Effect of ASPIRE on Teacher Retention

BACKGROUND

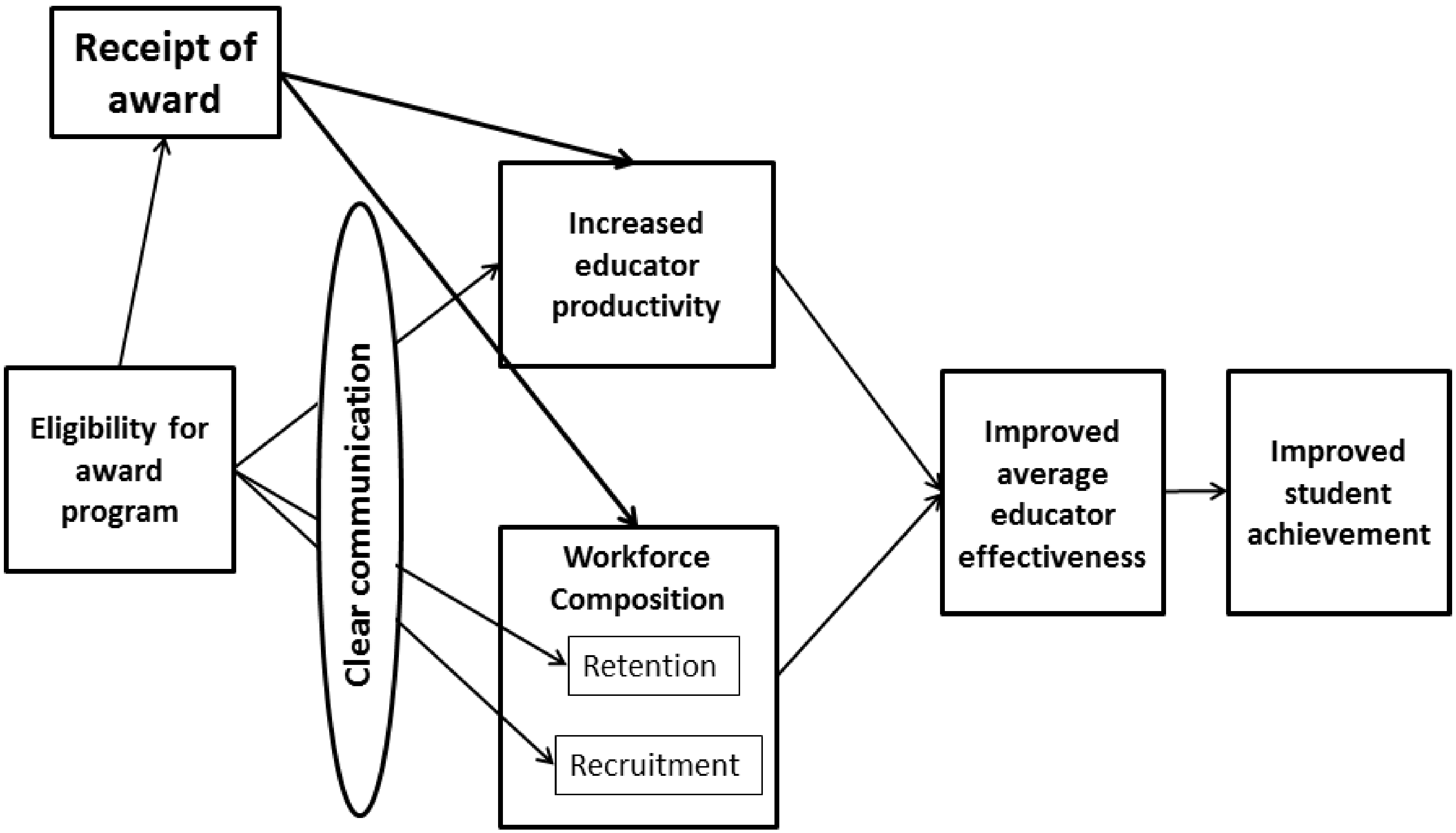
This section provides details on the theoretical history of educator award programs, the structure of the ASPIRE award program, previous evaluations of educator award programs, and our theoretical and methodological approach in this evaluation.

THEORETICAL BACKGROUND

Teacher award programs are a policy intervention aimed at improving student achievement, particularly at high need schools (Glazerman et al. 2011). They are motivated by persistent achievement gaps between status groups within the United States, and evidence that American students are not maintaining global competitiveness (Eberts, Hollenbeck and Stone 2002; Hanushek 2011). Interventions like these have grown in popularity with the recent federal emphases on accountability and the use of data-based evidence to improve teaching (Koedel and Betts 2009). They are also fundamentally tied to an economist perspective of schooling, labor, and human action (Storey 2000). Within this theoretical framework, student academic growth is contingent upon teacher effectiveness (Hanushek and Woessmann 2011; Rivkin, Hanushek and Kain 2005), with the lower average levels of achievement of students in high need schools particularly attributed to the lack of high quality teachers in such schools (Strunk and Zeehandelaar 2011). A central tenet of economic theory is that individuals are responsive to incentives, particularly financial incentives, and the structure of our public school system is criticized for providing no incentives for teachers to excel (Springer et al. 2010a). Proponents of award programs argue that the qualities upon which teachers' salaries are currently based (e.g., years of experience, postgraduate work) are not predictive of higher levels of student achievement (Aaronson, Barrow and Sander 2007; West and Mykerezi 2011). Proponents also point out that the undifferentiated pay structure, and barriers to firing less effective teachers within public schools, impede the recruitment and retention of quality teachers (Aslam and Kingdon 2011; Darling-Hammond 1996).

The conceptual model in Figure 1 shows that award programs are theorized to improve student achievement by improving average teacher effectiveness. Average teacher effectiveness is expected to increase as the possibility of receiving an award incentivizes teachers to increase their productivity (Currall et al. 2005; Milanowski 2003). This evaluation focuses on the possibility that receiving an award provides additional incentive beyond that of eligibility. In other words, among teachers eligible for the award program, we hypothesize that the outcomes of teachers who receive awards will improve more than the outcomes of teachers who do not receive awards. Award programs are also theorized to improve student achievement by increasing the quality of the workforce (Clabaugh 2009). Higher quality candidates will be attracted to schools with award programs, improving recruitment efforts (Lavy 2007; Milanowski 2003). Although few, if any, studies focus on the effect of receiving awards, many studies theorize that the workforce composition will also be improved by the attrition of lower performing teachers (i.e., teachers who do not receive awards) from the school or from the teaching profession (Jones and Hartney 2011). This aligns with our proposition that the outcomes of teachers who receive awards may improve more than the outcomes of teachers who do not receive awards.

Figure 1: Conceptual Model of Theorized Impact of Teacher Award Programs on Student Achievement



Note: This conceptual model is inspired by a figure from a report by Glazerman and colleagues (2011:7). These theorized relationships (with the exception of those relating to receipt of an award) align with the criteria the federal Teacher Incentive Fund suggests for its grantees. HISD's ASPIRE award program is also partially funded by TIF grants.

Opponents of award programs perceive their motivating theories as flawed. They argue that a business model is simply not applicable to schools and question whether teachers are truly motivated by money (Hanushek and Rivkin 2007; Morice and Murray 2003; Sandel 2012). They describe these programs' emphasis on competition, rather than collaboration and cooperation, as antithetical to the structure of schools and the general orientation of teachers (Andrews 2011; Goldhaber, DeArmond and DeBurgomaster 2008). Drawing from research that attributes academic achievement levels to differences across homes rather than differences across schools (Coleman 1990; Noguera 2003; Rothstein 2004), others argue that the assumption that students of low social status have lower levels of achievement because of their low quality teachers is simply a fallacy (Brick 2012; Hourigan 2011). Distinguishing between extrinsic and intrinsic motivation, opponents point to previous findings that rewards and punishments undermine intrinsic motivation and ultimately result only in short-term positive effects, if any (Benabou and Tirole 2003; Hulleman and Barron 2010; Ravitch 2012). Similarly, researchers forecast that financial incentives will demoralize teachers (Ballou 2001; Harkness and Schier 2011), erode morale and organizational trust (Murnane and Cohen 1986), and diminish the autonomy and professionalism of the teaching career (Ingersoll 2007; Storey 2002). From these perspectives, teacher award programs will never be effective because of their flawed theoretical origins.

Other researchers, not entirely discounting their potential, point out side effects of teacher award programs that may be counterproductive to good teaching. Teachers might focus on the emphases of the award program at the expense of other goals of public education (e.g., an emphasis on test-taking skills rather than experiential learning, or a de-emphasis of civic and social skills) (Jacob 2005; Propper and Wilson 2003; Ramirez 2011). The pressure that accompanies programs like these may lead teachers to cheat on tests, or to focus on the students with more potential for gain at the expense of other students (Kane and Staiger 2002a; Ryssdal 2011). Teachers may come to view students as a means to an award rather than building the relationships that are central to teaching/learning (Mahony, Menter and Hextall 2004). They may also compete rather than collaborate with colleagues (Benabou and Tirole 2003; Holt 2001) and develop more negative relationships with their principals (Clabaugh 2009). From these perspectives, the potential negative effects of teacher award programs may outweigh any benefits.

Researchers attribute the ineffectiveness of some award programs to poor design. The qualities of good teaching are complex and difficult to capture, and it is especially difficult to account for the confounding influence of the characteristics of schools and students when evaluating teachers (Caillier 2010). Crafting objective and reliable measures of teacher performance is a challenge (Carnoy et al. 2007; Gratz 2009; Ingvarson and Rowe 2008; Jacob and Lefgren 2008). Increasingly popular, value-added methodologies are designed to separate the effect of teachers from other influences on students' achievement (Ballou, Sanders and Wright 2004; Winters 2012), but the methodology also has its critics (Darling-Hammond et al. 2012; Kane and Staiger 2002b; Papay 2011). Others argue that award amounts are typically too low to really motivate teachers (Eberts et al. 2002; Fryer et al. 2012), or that the degree to which teachers are motivated depends on the degree to which they have control over the students whose achievement determines their award (e.g., entire campus vs. teachers' own students) (Imberman and Lovenheim 2012; Podgursky and Springer 2007).

In addition, the effectiveness of even well-designed programs depends on the quality of their implementation. In a sea of ever-evolving new educational interventions, an effective award program must ensure teachers are aware of the program, understand the criteria by which they will be evaluated, and agree that the program is fair and worthwhile (Farrell and Morris 2009; Hanshaw 2004; Scott 2011; Wright 2003). Given the complexity and sophistication of the algorithms often used to make award determinations, this is a difficult task (Imberman and Lovenheim 2012). Some programs have additionally been criticized for not providing teachers with the tools to improve their performance (e.g., professional development, mentoring) (Hulleman and Barron 2010; Koppich 2005). Reviews of previous literature caution against entirely discounting teacher award programs on the basis of the flaws of a single program (Podgursky and Springer 2007; Podgursky and Springer 2011; Ritter and Jensen 2010). Although it is difficult to parse out whether programs fail because of flawed theory, poor design, or poor implementation, this evaluation focuses on capturing nuances like these that are especially relevant for policy reform.

ASPIRE Award Program

The U.S. Department of Education classified HISD as the eighth largest school district in the nation in 2008-09 (Sable, Plotts and Mitchell 2010). In 2010-11, HISD consisted of approximately 200,000 students, 12,000 teachers, and 300 schools (Department of Research and Accountability 2011). In 2010-11, 62% of the students in HISD were Hispanic, 26% were black, 8% were white, and 3% were Asian (Department of Research and Accountability 2011). Over 80% of the students in HISD were economically disadvantaged, with 11% qualifying for the reduced lunch program, 41% qualifying for the free lunch program, and 30% classified as living in poverty (i.e., more economically disadvantaged than the first two groups).[3] Around 30% of HISD's students in 2010-11 were categorized as Limited English Proficient (LEP), and 29% were enrolled in a bilingual or English as a Second Language (ESL) program (Department of Research and Accountability 2011). HISD had a 4-Year Completion Rate of 74.3% and a dropout rate of 12.6% for the 'Class of 2010' cohort (students who began high school in 2006-07) (Texas Education Agency 2011a). The prevalence of high need schools in HISD was likely a contributor to HISD's decision to implement an educator award program.

HISD has had an award program that included teachers since 2000-2001. The 2005-2006 school year was the first time awards were based on individual teacher performance as well as campus performance. HISD's award programs were originally funded through local sources, but the district's receipt of federal Teacher Incentive Fund (TIF) grants[4] and state district awards for teacher excellence grants enabled HISD to increase the maximum teacher award with each subsequent year [from $7,000 in 2005-2006 (Zimmerman et al. 2011) to $10,300 in 2009-2010 (see Appendix 2)]. With the incorporation of value-added methodology for the 2006-2007 school year, HISD's award program evolved into the ASPIRE award program. Since ASPIRE's first year, HISD has tweaked and modified the program to address issues that arose in feedback and consultation. The current ASPIRE program is designed to encourage teacher cooperation, align with the district's other school-improvement initiatives, use value-added data to reward teachers reliably and consistently, include core teachers at all grade levels, and address alignment of curriculum to tests on which awards are based (Zimmerman et al. 2011).

[3] These statistics estimated by the authors of this report.

[4] TIF grants are distributed with the intention of improving public education, particularly in high need schools, through award programs (Glazerman et al. 2011).

Table 1: 2009-10 Award Eligibility by ASPIRE Category

ASPIRE Category	Percentage of 2009-10 HISD employees[a]	Eligibility for 2009-10 ASPIRE Awards: Ineligible because: Attendance	Opted out	Eligible for Strands: I	I and III	I, II, and III
A: Self-contained core foundation teachers, gr. 3-6	4.6%	2.5%	0.00%	-	-	97.4%
B: Departmentalized core foundation teachers, gr. 3-8	11.7%	5.1%	0.24%	-	-	94.5%
C: Core foundation teachers, gr. 9-12	6.5%	6.6%	0.00%	-	-	93.3%
D: Core foundation teachers, gr. PK-2	13.7%	5.6%	0.06%	-	-	94.3%
E: Special education core foundation teachers, gr. 3-12	2.9%	7.8%	0.00%	-	-	92.1%
F: Elective/ancillary teachers	10.0%	9.9%	0.16%	-	89.8%	-
G: Instructional support staff	7.5%	8.7%	0.05%	-	90.5%	-
H: Teaching assistants	6.6%	15.3%	0.00%	-	84.6%	-
I: Operational support staff	13.7%	14.2%	0.00%	85.7%	-	-
J: Principals	1.1%	2.5%	0.00%	-	-	97.5%
K: Assistant principals	1.5%	3.9%	0.00%	-	-	96.1%

Note: These analyses are based on the employees included in the 2009-10 ASPIRE award data file. The estimates in this table are very similar to those published by HISD (Zimmerman and Stevens 2011b). Minor differences are to be expected when different research teams conduct similar analyses using the same data, because even estimates as basic as these depend on a multitude of relatively subjective analytic decisions.

a - The percentages in this column do not sum to 100%, because the 19.6% of employees ineligible for ASPIRE because of job duties or campus are not included in this table.

Table 2: Receipt of 2009-10 Awards by ASPIRE Category

ASPIRE Category	Any ASPIRE award	Strand I	Strand II	Strand III	Attendance bonus
	Among those eligible, percent received:				
A: Self-contained core foundation teachers, gr. 3-6	96.1%	58.2%	84.0%	98.0%	14.4%
B: Departmentalized core foundation teachers, grades 3-8	93.5%	63.6%	60.4%	98.4%	10.9%
C: Core foundation teachers, grades 9-12	93.3%	73.1%	84.7%	97.0%	12.0%
D: Core foundation teachers, grades PK-2	93.4%	56.3%	61.0%	98.7%	11.5%
E: Special education core foundation teachers, grades 3-12	90.8%	66.8%	82.9%	96.3%	9.0%
F: Elective/ancillary teachers	88.7%	67.8%	-	97.9%	11.4%
G: Instructional support staff	85.5%	62.7%	-	93.8%	15.0%
H: Teaching assistants	75.7%	59.1%	-	77.8%	4.4%
I: Operational support staff	54.2%	63.3%	-	-	-
J: Principals	96.7%	54.1%	88.1%	98.1%	-
K: Assistant principals	94.6%	71.1%	93.3%	97.9%	-

Note: These analyses are based on the employees included in the 2009-10 ASPIRE award data file. The estimates in this table are very similar to those published by HISD (Zimmerman and Stevens 2011b). Minor differences are to be expected when different research teams conduct similar analyses using the same data, because even estimates as basic as these depend on a multitude of relatively subjective analytic decisions.

More broadly, the ASPIRE program also aligns with the theoretical underpinnings (as seen in Figure 1) and criteria for grantees associated with TIF grants. In 2010, TIF grantees were required to use their funds to provide differentiated levels of compensation for teachers and principals based on effectiveness, provide extra pay to teachers who take on additional responsibilities and assume leadership roles, and provide targeted professional development (Glazerman et al. 2011). TIF also encouraged grantees to provide incentives to retain and recruit effective teachers in high need schools and hard-to-staff subjects. Although eligibility for the ASPIRE award program depends on an employee's job category and duties, and their campus of employment, the vast majority of HISD's employees (core[5] and elective/ancillary teachers, principals, assistant principals, instructional support staff, teaching assistants, and operational support staff) are eligible to receive an award through the program. Employees can also become ineligible for the program through non-retention[6] or poor attendance rates throughout the school year. HISD employees can receive awards through a combination of three Strands of the ASPIRE program. HISD uses 11 ASPIRE categories to determine the Strands through which employees are eligible to receive awards, and the method by which award determinations are made within each Strand (see Table 1, more details available in Appendix 1).

Strand I awards are calculated in the same way for all HISD employees, but awards through Strands II and III are calculated in a multitude of ways, depending on employees' grade level and job functions. Award determinations through Strands I and III are based entirely on campus level achievement, while some Strand II awards are based on the achievement of more select groups of students. Strands I and II rely on a value added methodology,[7] which is thought to result in more fair and accurate determinations of educator effectiveness (Dillon 2010). Strand I awards are determined by campus level value-added scores across all grade levels and subjects, while Strand II awards are determined by the achievement gains of more select groups of students (more details to follow). Depending on employees' ASPIRE categories and grade levels, Strand III awards are based on each campus' Comparable Improvement[8] ranking from TEA; enrollment in Advanced Placement (AP), International Baccalaureate (IB) or dual credit courses; performance on AP/IB exams; TEA's rating of the campus; and/or the proportion of students who met the College Readiness Standard on the Writing Texas Assessment of Knowledge and Skills (TAKS). Across many of the various award criteria, employees whose schools are ranked in the top two quartiles across the district receive awards, with employees in the top quartile receiving a larger award amount than employees in the second quartile. There are also differences in maximum award amounts, depending on the employee's ASPIRE category. Table 2 shows the proportion of HISD employees who received an award through each Strand in 2009-10, depending on their ASPIRE category. Appendices 2 and 3 provide more details on ASPIRE categories and the 3 Strands of ASPIRE.[9] Average total awards for 2009-10 were around $3,000 for teachers, $6,500 for principals, $4,000 for assistant principals, and $650 for all other types of employees – see Table 3 for more details on average award amounts.

---

[5] HISD classifies reading, math, language arts, science, and social studies as core subjects (Zimmerman and Stevens 2011).

[6] Employees who are terminated or who leave in the middle of the year become ineligible for an award. Employees who complete the year and leave the district in good standing retain their eligibility to receive an award.

[7] HISD contracts with EVAAS (Dr. William Sanders' Education Value-Added Assessment System) to obtain value-added scores for campuses and teachers.

[8] Comparable Improvement is a measure that shows how student performance on the reading/English Language Arts and mathematics TAKS at a given school has changed from one year to the next, and then compares that change to that of the 40 schools in the state that are demographically most similar to the school of interest. (Texas Education Agency, http://ritter.tea.state.tx.us/perfreport/ci/2011/index.html).

[9] HISD's ASPIRE website, www.houstonisd.org/ASPIRE, also offers other details on the program.

Table 3: Average 2009-10 Award Amounts by ASPIRE Category

	Among those who won an ASPIRE award, total amount including attendance bonus (if any):		
ASPIRE Category	Min	Max	Mean
A: Self-contained core foundation teachers, gr. 3-6	$100	$11,330	$4,157
B: Departmentalized core foundation teachers, grades 3-8	$100	$11,110	$4,164
C: Core foundation teachers, grades 9-12	$200	$10,670	$4,432
D: Core foundation teachers, grades PK-2	$100	$7,260	$2,737
E: Special education core foundation teachers, grades 3-12	$100	$7,040	$2,827
F: Elective/ancillary teachers	$100	$3,410	$1,594
G: Instructional support staff	$44	$1,870	$813
H: Teaching assistants	$25	$1,155	$544
I: Operational support staff	$150	$750	$564
J: Principals	$200	$15,530	$6,300
K: Assistant principals	$100	$7,765	$4,036

Note: These analyses are based on the employees included in the 2009-10 ASPIRE award data file. The estimates in this table are very similar to those published by HISD (Zimmerman and Stevens 2011b). Minor differences are to be expected when different research teams conduct similar analyses using the same data, because even estimates as basic as these depend on a multitude of relatively subjective analytic decisions.

PREVIOUS EVALUATIONS OF EDUCATOR AWARD PROGRAMS

The mixed findings from previous evaluations leave it largely unclear whether educator award programs are effective. To facilitate a digestible synthesis, we present a summary of results from previous evaluations in Table 4. We do not include studies conducted on programs outside of the U.S., because these findings are likely not generalizable to the U.S. [because of differences across countries in the structure of schools, the economy, and the workforce (Rothstein 2012)]. Ritter, Maranto and Buck (2009) note that position and theoretical pieces abound because of this topic's prevalence in contemporary politics; Table 4 only includes pieces that utilized data. A sizeable proportion of these evaluations were published by think tanks, or are "working papers" – we include these, although the extent to which they were formally peer reviewed is unclear.

In general, Table 4 shows that changes in the composition of the teacher workforce (which includes retention and recruitment outcomes), and students' scores on standardized tests, are the most prevalent outcomes of interest. Some evaluations also focused on the effect of award programs on teachers' behaviors, attitudes, attendance rates, or work environment (the 'Other' category). Other evaluations focused on the effect of award programs on students' attendance, behaviors, test taking, course taking, grades, course failures, graduation, or dropout (the 'Other' category for students). Eight of twelve evaluations found a positive estimated effect on teacher workforce composition, recruitment, or retention for at least some teachers or in some contexts; three of the remaining four found no effect, and one found a negative effect. Six of ten evaluations found some positive effect of award programs on some other teacher outcome, whereas the other four were equally divided between finding null or negative effects. Thirteen of eighteen evaluations found that award programs have positive effects on students' test scores, but in virtually all of these, the positive effects were only evident for some students, some tests, or in certain contexts. Four of eighteen evaluations found some negative effects on students' test scores. The five evaluations that looked at other student outcomes had very mixed findings. In general, this suggests that award programs merit further study, with a particular focus on the aspects of these programs that are more effective than others.

Table 4, Part 1 of 2: Summary of Findings from Previous Evaluations of Educator Award Programs

Program Name	Location	Teachers: Workforce Composition[a]	Teachers: Other[b]	Students: Test Scores	Students: Other[c]	Citation
ASPIRE	TX - Houston			Positive or no effect		Imberman & Lovenheim 2012
Career Ladder Evaluation System	TN			Positive for some		Dee and Keys 2004
District Awards for Teacher Excellence	TX	Positive for some	Positive for some	Positive for some		Springer et al. 2010b
Governor's Educator Excellence Grant	TX	Positive or no effect	Mixed	Mixed		Springer et al. 2009b
Mission Possible	NC	Positive	Negative	Positive		Bayonas 2010
Project on Incentives in Teaching	TN - Nashville			No effect for most		Springer et al. 2010a
School-Wide Performance Bonus Program	NY - New York City		No effect	Negative or no effect	No effect	Fryer 2011
		No effect		Negative or no effect	Positive for some	Goodman and Turner 2010
			No effect	No effect		Marsh et al. 2011
				Positive, negative, or no effect		Goodman and Turner 2011
Teacher Advancement Program	AZ			Positive for some		Schacter and Thum 2005
	IL - Chicago	Positive		No effect		Glazerman et al. 2009
	IL - Chicago	No effect		No effect		Glazerman and Seifullah 2010
	USA - 2 states			Positive or mixed		Springer et al. 2008
	USA	Positive	Positive	Positive or no effect		Solmon et al. 2007

Table 4, Part 2 of 2: Summary of Findings from Previous Evaluations of Educator Award Programs

Program Name	Location	Teachers		Students		Citation
		Workforce Composition[a]	Other[b]	Test Scores	Other[c]	
Texas Educator Excellence Grant	TX	No effect	Positive	No effect		Springer et al. 2009a
no name	GA		Positive	Positive		McCollum 2001
no name	IL - Chicago Heights			Positive for some		Fryer et al. 2012
no name	MI	Positive			Negative or no effect	Eberts et al. 2002
no name	NC			Positive		Jinnai 2012
no name	NC			Negative or no effect	Positive or no effect	Smith and Mickelson 2000
no name	SC			Positive		Cooper and Cohn 1997
no name	TX - Dallas	Positive		Positive for some	Positive	Ladd 1999
various	NC	Negative for some				Guarino et al. 2011
various	USA		Positive or negative			Kelley 1999
various (NELS)	USA			Positive		Figlio and Kenny 2007
various (Schools and Staffing Survey)	USA		Negative or no effect			Belfield and Heywood 2008
		Positive				Jones and Hartney 2011

a - 'Workforce composition' includes retention and recruitment outcomes.

b - 'Other' outcomes for teachers include behaviors, attitudes, attendance, and work environment.

c - 'Other' outcomes for students include attendance, behavior, test taking, course taking, grades, course failures, graduation, and dropout.

METHODOLOGICAL APPROACH

In this section, we discuss our methodological approach in the context of the methodology of previous evaluations of educator award programs. Educator award programs are expected to result in aggregate level changes over time, including, but not limited to, improved student achievement levels and a higher quality workforce. Without a reference group, such as a similar district without an ASPIRE program, we cannot attribute aggregate level changes (i.e., district wide changes) to ASPIRE with any level of confidence as these changes could be attributable to a wide variety of forces beyond the ASPIRE program (e.g., the economy, other programs within the district[10]). Only two previous evaluations did not employ a control group in their analyses [see (Bayonas 2010; Kelley 1999)]. Although some evaluations compared changes over time in their target group to national data or data from other sources [see (Ladd 1999; Smith and Mickelson 2000)], the distinctiveness of HISD and differences in measures across data sources make appropriate matches nearly impossible. [11]

Randomized trials are considered the best method by which to construct a control and treatment group with similar characteristics (Neuman 2009). By focusing on programs designed to accommodate evaluation, i.e., programs implemented among a random selection of schools or teachers (Podgursky and Springer 2007), the vast majority of previous evaluations of educator award programs have compared changes over time between eligible and ineligible but otherwise similar teachers (Dee and Keys 2004; Fryer et al. 2012; Springer et al. 2010a), or otherwise similar schools (Eberts et al. 2002; Fryer 2011; Glazerman and Seifullah 2010; Goodman and Turner 2010; Goodman and Turner 2011; Marsh et al. 2011; McCollum 2001; Schacter and Thum 2005; Solmon et al. 2007; Springer, Ballou and Peng 2008; Springer et al. 2010b; Springer et al. 2009b). ASPIRE was not a randomized trial because it was implemented in virtually all HISD schools and among virtually all HISD employees.

We capitalize on the strengths of our data by comparing changes in outcomes over time between core teachers who did and did not receive awards. Only a few other evaluations have focused on award receipt rather than award eligibility [see (Jinnai 2012; Springer et al. 2009a)]. Although not an ideal test of the effectiveness of ASPIRE, this approach does enable a unique focus on the mechanisms by which ASPIRE, and educator award programs more generally, create change. The multivariate analyses in this evaluation primarily focus on core teachers because these teachers' outcomes are of primary interest and because core teachers were the only employees eligible for awards not determined on the basis of campus level achievement. Core teachers were also well represented in the data files with potential outcomes of interest. Again capitalizing on the strengths of our data, we focused on three outcomes: 1) core teacher retention, 2) core teacher attendance, and 3) math and reading/ELA[12] teachers' mean student test score gains on the math and reading TAKS and Stanford tests. Figure 2 shows the temporal order of and theorized connections between

---

[10] In addition to ASPIRE awards, some HISD educators were simultaneously eligible for award programs funded by the Texas and Governor's Educator Excellence Grants (TEEG and GEEG) and managed by each school (Springer et al. 2009a; 2009b; 2010b). (Per HISD staff, the ASPIRE award program was under the umbrella of the District for Teacher Excellence (D.AT.E.) program.) Per HISD staff, students and teachers at 4 schools, and teachers at 1 school, received financial awards through a Dell Grant for increasing performance on AP tests; performance on AP tests is also a criteria for Strand III awards through ASPIRE.

[11] For instance, Ladd (1999) attributed Dallas' higher levels of student achievement growth than that of other Texas cities to its award program, but also found that Dallas had higher levels of growth than other cities before the program was implemented, and loosely attributed that to positive publicity about the program.

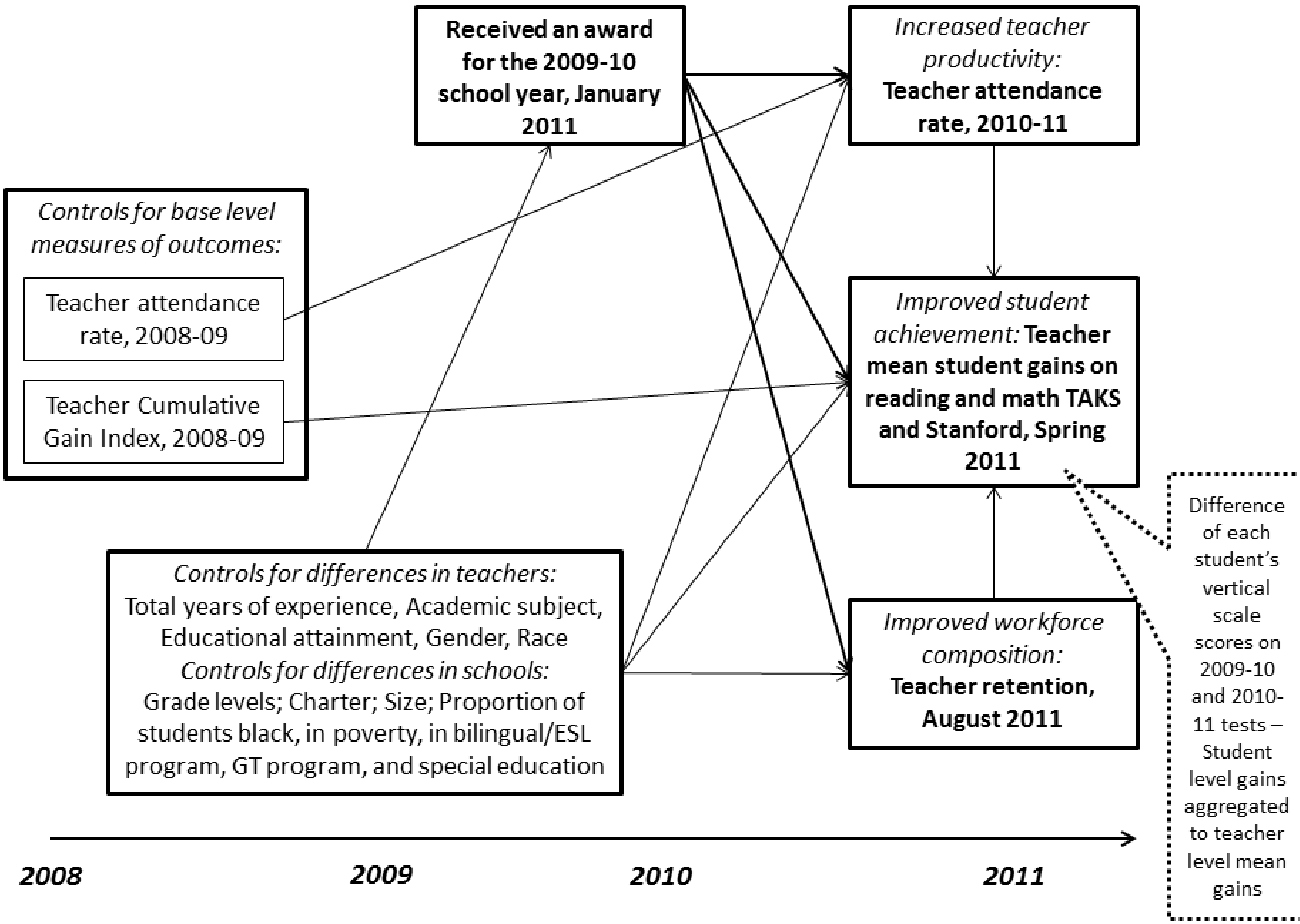
[12] ELA refers to 'English Language Arts.'

the predictors and outcomes of interest in this evaluation. If ASPIRE motivates high quality teachers to continue teaching in the district and low quality teachers to quit teaching or seek employment elsewhere, teachers who receive awards should be more likely to be retained the following year. [It is important to note that considerable debate remains as to the accuracy of value-added measures in identifying high quality teachers (Chetty, Friedman and Rockoff 2011; Kane and Staiger 2008).] If teachers are rewarded for their good attendance by retaining eligibility for the award program and even receiving an attendance bonus, teachers who receive awards should be more motivated to maintain good attendance rates the following year than teachers who do not.[13] Maintaining good attendance rates may also contribute to students' achievement gains and teachers' odds of receiving an award the following year. Similarly, teachers who receive an award for their students' achievement gains may be more motivated to encourage the achievement of their students the following year in order to receive another award than teachers who do not receive an award. We are unable to include the test score gains of students who took versions of the test not on the same vertical scale (e.g., Spanish, special education). See the Dependent Variables section of this evaluation for more details on the construction of these outcome measures.

Comparing pre- and post-program outcomes was another possibility. However, HISD does not have consistent data preceding ASPIRE, and an inability to accurately link students to teachers in the earliest year (2006-07) warrants a focus on more recent years. Moreover, the effect of ASPIRE in more recent years is of urgent policy interest to HISD. Although we received data on ASPIRE awards for the 2010-11 school year, we focused on 2009-10 ASPIRE awards to enable longitudinal analyses that establish temporal order (a first criteria for establishing causality). In other words, we constructed a longitudinal dataset, or a dataset that links measures describing the same people at different points in time, and designed our evaluation to focus on outcomes that occurred after the receipt of the award. Additionally, we focused on 2010-11 outcomes because the state of Texas transitioned from the TAKS to the State of Texas Assessments of Academic Readiness (STAAR) in 2012, making the estimation of gain scores impossible. It's important to note that teachers received official notification of their 2009-10 award in January 2011 (per HISD staff, many were informally aware of their award receipt status as early as October of 2010). Because of the timing of award notification, we focus on the characteristics of teachers' 2010-11 schools when considering the effect of award receipt on teacher outcomes.

[13] Although the attendance bonus was discontinued for the 2010-11 ASPIRE award program, HISD staff report that many teachers were not aware of this change.

Figure 2: Conceptual Model of this Evaluation



In addition to establishing that the cause occurred before the effect (i.e., the award was received before retention, attendance, and achievement gains were measured), causal claims depend upon establishing that a relationship exists between the cause and the effect, and that this relationship is not actually attributable to some other factor related to both the cause and the effect (Neuman 2009). We can use descriptive analyses (see Table 5) to demonstrate whether a relationship appears to exist between receiving an award and the outcomes of interest. Analyses like these, though, are not sufficient evidence that these differences are attributable to receipt of an award. As suggested by the third criterion for causality, the very characteristics that led these teachers to receive an award may be the characteristics that result in their better outcomes. In other words, teachers who receive awards may have had better outcomes than other teachers regardless of whether an award program was in place. [Analyses that focus on overall ASPIRE award receipt are also problematic because the vast majority of core teachers received some award, and so the reference group is small and probably not comparable to the group of teachers who did receive an award. For this reason, this evaluation differentiates by the Strand of award receipt (teachers received Strand I and II awards at much lower rates than Strand III awards).]

Table 5: Differences in the 2010-11 Outcomes of Core Teachers Who Received a 2009-10 ASPIRE Award

	Mean retention rate, August 2011	Mean 2010-11 attendance rate	Mean gain in students' 2010-11 scores on:			
			TAKS math test	TAKS reading test	Stanford math test	Stanford reading test
Received any ASPIRE award in 2009-10:						
No	0.35	0.94	-0.14	0.03	-0.10	-0.11
Yes	0.80	0.96	0.02	-0.02	0.02	-0.01

Note: These analyses exclude core teachers who were ineligible for the 2009-10 award program because of their job, campus, or opting out, as well as core teachers not included in each respective outcome data file.

When randomized experiments are not possible, quasi-experimental methods are the next best option. This evaluation used regression analyses to compare the outcomes of teachers who received and did not receive awards but are otherwise similar (logistic regression models to predict teacher retention, poisson regression models to predict teacher attendance rates, and linear regression models to predict mean student test score gains).[14] In regression analyses, we only include teachers who were not missing on the outcome and who were included in the 2009-10 award data file. In doing so, teachers who were hired in 2010-11 are excluded, which is appropriate since these teachers are not comparable to other non-award-recipients, having never had the chance to be eligible for an award. Poisson regression models are used to model count data, or data that sums the number of events experienced by each case. Our measure of teacher attendance is a count of the number of hours each teacher was present in 2010-11. We included the number of hours teachers were scheduled to work in 2010-11 as an exposure indicator (a measure that indicates the number

[14] Although complicated by the fact that individual teachers experience a multitude of cutoff points for receiving awards through the ASPIRE program, we plan to further test the robustness of these results with regression discontinuity methods in later versions of this evaluation. Preliminary RD analyses with a small subset of teachers suggested results consistent to those presented in this report.

of times the event of interest, being present in this case, could possibly happen). Including the exposure indicator means that we are effectively modeling an attendance rate (the number of hours present conditioned on the number of hours scheduled). Results were similar whether we used linear regression or Poisson regression models, but the data more closely aligned with the assumptions of Poisson models.

We attempted to compare teachers who are as similar as possible, with the exception of the characteristic(s) of interest, by including controls describing each teacher's sex, race, educational attainment, total years of experience, receipt of new hire recruitment stipends, and the characteristics of their school (see Figure 2). To further isolate the effect of receiving an award, we included a measure of teachers' 2008-09 attendance rates to create lagged models when predicting 2010-11 attendance rates. By focusing on gains from 2009-10 to 2010-11 in our student achievement analyses, we are essentially controlling on baseline achievement levels and all of the sociodemographic factors that produce those baseline achievement levels, but we further isolate the effect of receiving an award on student achievement gains by including a measure of each teacher's 2008-09 Cumulative Gain Index (CGI)[15]. We essentially compare year three outcomes (2010-11) of teachers who did and did not receive awards in year two (2009-10), among teachers who were similar on that outcome in year one (2008-09).

Nonetheless, the possibility remains that differences in teachers' outcomes that persist net of controls are attributable to unmeasured factors (e.g., the characteristics of their schools, the composition of their classrooms, their personal attributes). At the most extreme, our approach could be described as a confirmation that the awards were distributed to the appropriate teachers (Cooper and Cohn 1997). We estimated two-level random effects models to account for the clustering of teachers in schools and the possibility that unmeasured distinctions across schools contribute to differences in outcomes (Bollen and Brand 2010). (Three-level models were not an option for gains in student achievement because students are nested in multiple teachers.) We used random rather than fixed effects models because school level characteristics are of interest in this evaluation (Rabe-Hesketh and Skrondal 2012). Unlike fixed effects models, random effects models rely on the assumption that the unmeasured characteristics of the clusters are uncorrelated with the characteristics of the individual cases included in the model (Clarke et al. 2010). Obviously, it is unlikely that this assumption is met in the case of schools and teachers. Clarke et al. (2010) argue that policy relevant education research relies on the expanded use of random effects models. They posit that with some knowledge of the school selection mechanisms, i.e., the characteristics of schools that differentiate the types of teachers that work in them, including controls for these factors can enable the use of random effects models. See Grodsky and Riegle-Crumb (2010) and Riegle-Crumb and Grodsky (2010) for recent examples of education research using random effects models. Results from the random effects models are similar to results obtained in exploratory analyses with cluster-robust standard errors.

Receipt of Strand I, II, and III Awards

Our first set of models compares the outcomes of teachers who received Strand I, II, and/or III awards to those of teachers who did not. These models are intended to illuminate whether there is an independent effect of receiving an award, net of eligibility for the award program, and whether teachers are more responsive to competitive (Strand II) rather than collaborative (Strands I and III) awards. Strand II is of particular interest in this evaluation because it is the only Strand that awards some teachers on the basis of more select groups of students and because HISD pays out the most

[15] Value-added score calculated by EVAAS to compare achievement gains of each teacher's students to those of all HISD students.

for Strand II awards [about $20 million versus $10 million each for Strand I and III awards (Mosier and Stevens 2011)]. Table 7 shows that Strand II awards are determined by value-added scores based on the achievement gains of: A) each teacher's own students (for core teachers of grades 3-8), B) students on their campus in their subject(s) (for core teachers of grades 9-12, and special education teachers of grades 3-12),[16] or C) all third graders on the campus or at the feeder campus in reading and math (for teachers of pre-kindergarten through grade 2). In contrast to the individualistic theories of competition that motivate teacher award programs, teachers often espouse a preference for programs that prioritize cooperation and collaboration over competition with colleagues (Goodman and Turner 2011; McCollum 2001). Many programs incorporate an emphasis on collaboration in an attempt to increase teacher buy-in (Glewwe, Ilias and Kremer 2003), as potentially evidenced by ASPIRE's Strands I and III. But from the perspective that awards not based in competition defeat the original purpose of these programs (Buck and Greene 2011), teachers will be more motivated by receipt of a Strand II award than by receipt of a Strand I or III award, because Strand II awards are more competitive than Strand I or III awards.

All models exclude core teachers who were not eligible for the award program because of their job duties or campus, as well as teachers who opted out of the award program. This ensures the comparison of teachers who received an award to those who did not but were eligible to receive one. We don't exclude teachers who were ineligible for the award program because of their attendance or lack of retention because these teachers ostensibly were unaware through some part of the school year that they would eventually become ineligible, just as other teachers are unsure whether student achievement is high enough to warrant their receipt of an award.

[16] High school core teachers' Strand II awards are really department level awards, with 1/3 based on 9th graders', 1/3 based on 10th graders', and 1/3 based on 11th graders' achievement gains in the department's subject. Because special education teachers are self-contained rather than departmentalized, their Strand II awards are usually based on campus level achievement in all core subjects.

Table 6: Core Teachers' Eligibility for and Receipt of 2009-10 ASPIRE Awards

	Percentage of 2009-10 core teachers[a]	Eligibility for 2009-10 ASPIRE awards[a]		
		Ineligible because:		Eligible
		Attendance	Opted out	
All core teachers	100.0%	5.5%	0.09%	94.4%
Lower need schools	64.2%	5.1%	0.06%	94.9%
Higher need schools	35.8%	6.1%	0.14%	93.7%
Secondary math	7.8%	7.9%	0.13%	92.0%
Secondary science	6.3%	6.9%	0.00%	93.1%
Bilingual/ESL	31.8%	3.7%	0.13%	96.2%
Special education	13.5%	7.0%	0.08%	92.9%
Not hard-to-staff	40.7%	5.6%	0.08%	94.3%

	Among eligible core teachers, percent received:				
	Any ASPIRE award	Strand I	Strand II	Strand III	Attendance bonus
All core teachers	93.6%	62.2%	69.0%	98.1%	11.6%
Lower need schools	94.3%	63.6%	70.6%	98.9%	12.7%
Higher need schools	92.3%	59.6%	66.2%	96.5%	9.7%
Secondary math	91.7%	75.9%	68.8%	98.4%	13.3%
Secondary science	92.3%	75.4%	68.0%	97.5%	8.9%
Bilingual/ESL	95.7%	57.9%	69.6%	98.8%	14.4%
Special education	91.4%	61.6%	74.6%	96.4%	10.0%
Not hard-to-staff	93.2%	61.2%	67.1%	98.0%	10.1%

a - These analyses use all core teachers in the 2009-10 ASPIRE award data file.

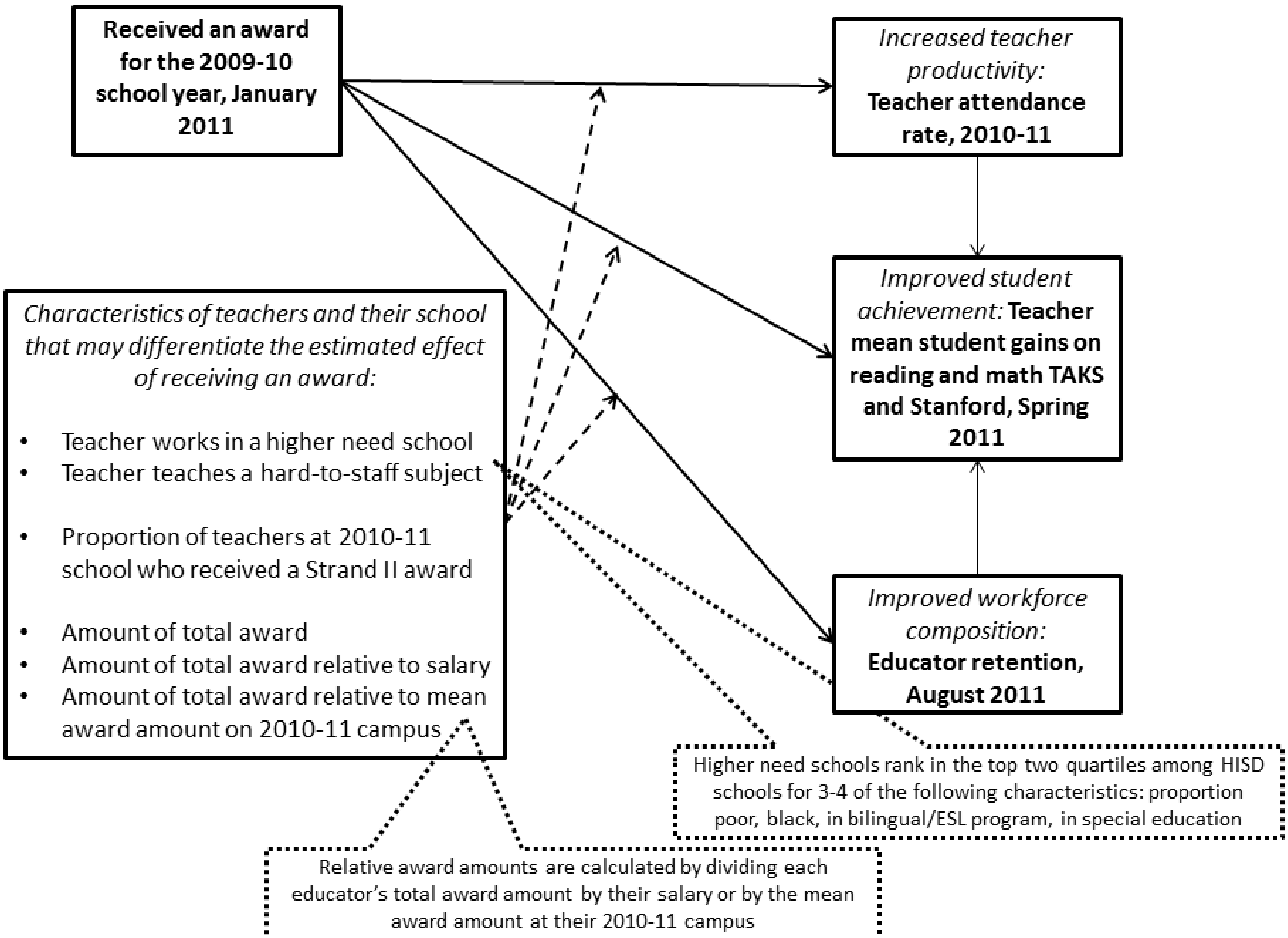
Table 6 shows that 94.4% of core teachers were eligible for the 2009-10 ASPIRE award program, and 93.6% received an ASPIRE award. Core teachers were much less likely to receive Strand I or Strand II awards (62.2% and 69.0%, respectively) than they were Strand III awards (98.1%). Employees who had perfect attendance during 2009-10 received an attendance bonus in the amount of 10% of their total ASPIRE award; employees who missed fewer than two days received an attendance bonus in the amount of 5% of their total award. 11.6% of core teachers received one of these attendance bonuses. It is important to not estimate models focused on particular Strands without controls for the other Strands, because, for example, a seeming effect of a Strand I award could actually be an effect of the Strand III award if teachers who receive Strand I awards are also more likely to receive Strand III awards. We include interactions in this first set of models to locate variation in the estimated effect of receiving a Strand I, II, and/or III award for teachers who work in a higher need school or teach a hard-to-staff subject, and depending on the prevalence of Strand II award receipt among eligible core teachers on the teacher's campus. The following sections provide more theoretical and analytical description of these groups of core teachers.

HIGH NEED SCHOOLS

High need schools are a particular focus for TIF funders, as well as HISD. According to the theories that support award programs, students in high need schools typically have lower average achievement levels because of the lower quality and less motivated teachers in such schools. Award programs are meant to motivate these teachers to work harder, and to encourage the best teachers to stay and the worst to leave (Kingdon and Teal 2007). The theoretical focus on teachers' undifferentiated pay structures makes it less clear how award programs target high need schools, as teachers in both low and high need schools typically experience the same pay structure. The focus on high need schools is also muddied in the design of some award programs, with awards available to teachers at both low- and high need schools (Solmon 2005), and concern that even the most sophisticated analytic methods do not capture the unique challenges teachers at high need schools face (Scherer 2001). Guarino, Brown and Wyse (2011) found that award programs focused on campus level achievement actually accelerated the drain of high quality teachers from high need schools. Similarly, Ramirez (2001) notes that increased teacher motivation may not be a sufficient ingredient to change high need schools. For these reasons, we explore whether teachers in HISD's highest need schools experience a comparable estimated effect of receiving an ASPIRE award as teachers in HISD's lowest need schools. Figure 3 provides a conceptual model of how we expect the estimated effect of receiving an award might vary.

Exploratory analyses showed that students' test scores, teacher attendance rates, and teacher retention rates were the lowest in HISD schools with higher proportions of students in poverty, black students, students in ESL, and students in special education. To align with the focus of this evaluation, we define high need schools using these school characteristics. For each school year, we classified schools whose proportions of poor students placed them in the top two quartiles within the district, to be high need insofar as this characteristic. Similarly, we classified schools whose proportion of black students placed them in the top two quartiles within the district, to be high need in terms of that characteristic. We did the same for the other two characteristics, and finally, summed the number of characteristics that qualified each school to be a high need school: a school ranked at 0 was not in the top two quartiles for any of these four characteristics, while a school ranked at 4 was in the top two quartiles for all four of these characteristics. To facilitate presentation of results, we compare core teachers in schools with 0-2 characteristics ('lower need'), to educators in schools with 3-4 characteristics ('higher need'). Table 6 shows that 35.8% of HISD's core teachers work in higher need schools. Core teachers in higher need schools were more likely than core teachers in lower need school to be ineligible for a 2009-10 award because of poor attendance, and were less likely to receive awards through all three Strands or to receive an attendance bonus.

Figure 3: Conceptual Model of Potential Variation in Effect of Receiving an ASPIRE Award



Hard-to-Staff Subjects

Hard-to-staff subjects are also a particular focus for TIF funders and for HISD. We identified secondary level math and science, any level of bilingual/ESL, and any level of special education as hard-to-staff subjects within HISD. While the undifferentiated pay structures of public school systems are criticized more broadly for creating undesirable recruitment and retention patterns, they are thought to particularly complicate the recruitment and retention of teachers with higher-paying options in the private sector (such as math and science teachers) (Hess 2011). Human Capital Talent Acquisition Selectors within HISD pursue highly qualified educators, with the goal of attracting them to HISD through recruitment stipends and the prospect of receiving an ASPIRE award for their efforts. Although the recruitment and retention of these teachers are targeted more directly by stipends than by ASPIRE awards, these stipends are a corollary of the award program. We explored whether teachers of hard-to-staff subjects experience effects of ASPIRE comparable to those experienced by teachers in not hard-to-staff positions, even net of the influence of recruitment stipends. Teachers in not hard-to-staff positions include both self-contained teachers (e.g., elementary teachers that teach multiple subjects), and teachers who teach subjects like English, history, and social studies. Table 6 shows that 7.8% of HISD core teachers taught secondary level math, 6.3% taught secondary level science, 31.8% were in a bilingual or ESL position, and 13.5% were in a special education position. Secondary level math and special education teachers were most likely to be ineligible for the 2009-10 ASPIRE award program because of poor attendance. Teachers of hard-to-staff subjects were more likely to receive a Strand II award than teachers in not hard-to-staff positions, with special education teachers particularly likely to receive a Strand II award. Secondary level science teachers were the least likely to receive an attendance bonus.

Prevalence of Strand II Award Receipt

We also attempt to understand whether teachers are more responsive to competitive or collaborative awards by exploring whether the estimated effect of receiving a Strand II award varies depending on the proportion of teachers on the campus who received Strand II awards. All eligible employees receive Strand I and III awards on campuses with achievement levels meeting these Strands' criteria. In contrast, some Strand II awards are only distributed to individual teachers on a campus or to all teachers in a department, potentially inspiring competitive comparisons among teachers on the same campus. Although HISD notifies teachers individually of their award status, a Houston newspaper lists teachers' names and award status in a publicly accessible online forum. HISD staff also stated that they have observed discussion and comparison of awards among teachers. These facts suggest that differences in award receipt across teachers may be salient to the public and teachers. If teachers are most responsive to competitive awards, we would expect that teachers who receive Strand II awards in schools in which receipt of Strand II awards is less prevalent will benefit more than teachers who receive Strand II awards in schools in which receipt of Strand II awards is more prevalent.

Method of Strand II Award Determination

The estimated effect of receiving a Strand II award may also vary depending on how the award is determined. If teachers are most responsive to competitive awards, teachers rewarded on the basis of the achievement gains of students they actually teach should respond more positively to Strand II awards than teachers rewarded on the basis of students outside of their own classrooms. Differences in the estimated effect of receiving a Strand II award may also reflect teachers' preferences for being rewarded on the basis of factors they can control (the students in their own classroom versus all students on the campus, for instance) (Goodman and Turner 2010). In a working paper, Imberman and Lovenheim (2012) found that teachers' effort increases as the proportion of students who they teach increases, among those students whose achievement is used to determine awards. And so it is also possible that teachers may respond more positively to Strand II awards than Strand I or III awards, and particularly respond positively if they received their Strand II award on the basis of the achievement of the students they actually teach, because they perceive the award as a result of their own efforts rather than a more abstract collective effort. To continue our exploration of whether teachers respond more positively to competitive rather than collaborative awards, or to awards based on students over whom they feel they have control, we estimated a second set of models with interactions that locate variation in the estimated effect of receiving a Strand II award depending on the group of students whose achievement was used to determine the award. We used each teacher's ASPIRE category to capture these differences; because these measures are correlated to some extent with our constructed measures of hard-to-staff subjects, we excluded the latter in this second set of models.

AMOUNT OF AWARD

Lastly, the extent to which an award acts as a financial incentive may depend on its amount (Storey 2000). We explored variation in the benefits of receiving an award, depending on the total award's raw amount, its amount relative to the teacher's annual salary, and its amount relative to the mean award amount at the teacher's school. Table 8 shows that the raw amount of core teachers' total 2009-10 awards ranged from $100 to $11,300, with the median amount at $3,200. These amounts represented 0.001 to 0.31 of their annual salaries (we divided each teacher's award amount by his or her salary – these estimates are means of those proportions), with the median at 0.06. Finally, these amounts represented 0.03 to 12.42 of the mean award amount on their 2010-11 campus, with the median at 0.89. Teachers with a relative amount of 12.42 received Strand II awards on campuses in which receipt of Strand II awards was uncommon. In these models, we excluded all teachers who did not receive an award, so recipients of a larger award are compared to recipients of a smaller award. We used each teacher's ASPIRE category to control for systematic differences in award amounts among core teachers; because these measures are correlated to some extent with our constructed measures of hard-to-staff subjects, we excluded the latter in models 3 through 5.

Log odds and odds ratios are the default output from logistic and Poisson regression models. However, researchers increasingly emphasize that comparing log odds or odds ratios across models can result in invalid comparisons because of issues of scaling unique to these models (Allison 1999; Hoetker 2007). The interpretation of odds ratios is additionally complicated by their unnatural and asymmetrical scale: negative effects range from 0 to 1, while positive effects start at 1 but go well beyond 2 (Long 1997). Presenting results from these models with predicted probabilities and counts, or marginal effects (which are differences in predicted counts and probabilities), addresses these issues of scaling (Hoetker 2007). Predicted probabilities and counts also facilitate a more intuitive understanding of the results [for example, attendance results are presented in terms of hours or days rather than log odds of hours or factors of comparison (as is the case with odds ratios)]. Furthermore, we estimated the average marginal effect rather than the marginal effect at the mean, because the latter reflects only one of many possible sets of values (often one that doesn't represent the experiences of any real person) (Williams 2011).

Table 7: Variation in Determination of and Receipt of 2009-10 Strand II Awards

	Percent of core teachers in 2009-10	Receive Strand II award based on achievement gains of:	Percent received 2009-10 Strand II award
A: Self-contained core foundation teachers, grades 3-6	11.7%	Own students	81.8%
B: Departmentalized core foundation teachers, grades 3-8	29.6%	Own students	57.3%
C: Core foundation teachers, grades 9-12	16.5%	Students on campus in applicable subject(s), grades 9-11[a]	79.1%
D: Core foundation teachers, grades PK-2	34.8%	All 3rd graders on campus or feeder campus, reading and math	57.6%
E: Special education core foundation teachers, grades 3-12	7.5%	Students on campus in applicable subject(s)	76.4%

Note: These analyses exclude employees who were ineligible for the 2009-10 award program because of their job, campus, or opting out.

a - High school core teachers' Strand II awards are department level awards, with 1/3 based on 9th graders', 1/3 based on 10th graders', and 1/3 based on 11th graders' achievement gains in the department's subject.

b - Because special education teachers are self-contained rather than departmentalized, their Strand II awards are usually based on campus level achievement in all core subjects.

Table 8: Raw and Relative Amounts of Core Teachers' 2009-10 ASPIRE Awards

	Total Award Amounts[a]		
	Raw amount	Total amount as proportion of:	
		Teacher's salary[b]	Mean campus amount[c]
Minimum	$100	0.001	0.03
Median	$3,200	0.06	0.89
Maximum	$11,330	0.31	12.42

a - These analyses exclude core teachers who did not receive any award.

b - This column shows descriptive statistics from proportions estimated by dividing each core teachers' total 2009-10 award amount by their salary.

c - This column shows descriptive statistics on proportions estimated by dividing each core teachers' total 2009-10 award amount by the mean award amount among core teachers on their 2010-11 campus. The large maximum proportion represents core teachers who received Strand II awards on campuses in which receiving Strand II awards was relatively uncommon.

Although often misrepresented, p-values (or statistical significance estimates) indicate the degree of confidence with which effects or differences in effects evident in a sample of cases can be expected to be evident in the entire population. However, we have data on a census rather than a sample of 2009-10 HISD teachers, and we do not intend for these results to be generalized to other years of ASPIRE because the program is modified every year. Nonetheless, we show p-values because of the statistical argument that there is always a theoretical superpopulation and to err on the side of conservative assumptions. We still discuss statistically non-significant findings and emphasize whether the differences in our results are *substantively* significant, i.e., whether any changes in outcomes are large enough to warrant the cost of the ASPIRE program. We also present predicted means and marginal effects from linear regression models because estimations like these are more practically interpretable and have more substantive meaning.

Although limitations remain, the findings in this evaluation, which account for a multitude of teacher and school factors that are likely to be predictors of awards and outcomes, are a marked improvement over descriptive statistics. Moreover, studies relying on experimental data are often limited by attrition (schools and teachers drop out of the study). This was not a problem in this study because virtually everyone in the district participated in the study. We also build on our chances of estimating an unbiased effect of receiving an award by focusing on *changes* in outcomes over time, or gains, rather than simply predicting how receiving an award at time 1 is associated with an outcome at time 2. While the evaluations that employed nationally representative datasets obtained more generalizable results (Belfield and Heywood 2008; Figlio and Kenny 2007; Jones and Hartney 2011), those data limit the exploration of the nuances of award programs that shape their influence. In addition to understanding whether receipt of an award carries an independent effect beyond eligibility, we hope to illuminate whether certain aspects of ASPIRE are more or less effective, and whether ASPIRE benefits certain groups of teachers more than others.

Research Questions

Receiving an Award

- Among core teachers who were eligible for the 2009-10 ASPIRE award program, did the 2010-11 outcomes (retention, attendance, mean student achievement gain) of teachers who received a Strand I, II, and/or III award improve more than those of comparable teachers who did not receive an award?

Higher Need Schools

Award programs may be less effective for teachers at higher need schools because of the possibility that the lower average levels of achievement at higher need schools are not attributable to the qualities of teachers and because of the possibility that the value-added methodology does not sufficiently account for the unique challenges of working in a higher need school.

- Did the outcomes of core teachers in higher need schools who received an award improve to the same extent as those of comparable core teachers in lower need schools?

Hard-to-Staff Subjects

Most districts primarily hope to change the recruitment and retention patterns of teachers of hard-to-staff subjects. If these teachers feel they have better employment options outside of teaching, or that the value-added methodology does not capture the unique challenges of their positions, they may respond differently to the award program than teachers not in hard-to-staff positions.

- Did the outcomes of core teachers of hard-to-staff subjects who received an award improve to the same extent as those of comparable core teachers not in hard-to-staff positions?

Competition or Collaboration

Strand I and III awards are determined entirely on the basis of campus level achievement, while Strand II awards are based on the achievement of more select groups of students. For this reason, we conceptualized Strand II awards as competitive and Strand I and III awards as collaborative. From the perspective that award programs incentivize teachers through competition and reward, the outcomes of teachers should be more improved by receipt of a Strand II award than a Strand I or III award. Teachers may also respond more positively (i.e., experience more improvement in their outcomes) to Strand II awards because they are based on the achievement of students within their control, rather than the achievement of all students on the campus. If teachers are more incentivized by competitive awards, they should also respond more positively to receiving a Strand II award if the receipt of Strand II awards is less prevalent on their campus. Not all Strand II awards are based on the achievement gains of students teachers actually teach, and if teachers are more incentivized by competitive awards, the outcomes of teachers who receive a Strand II award on the basis of students they actually teach should be more improved than those of teachers who receive a Strand II award on the basis of a larger group of students.

- Did the outcomes of core teachers improve more by receiving a Strand II award than by receiving a Strand I or III award?
- Did the outcomes of core teachers who received a Strand II award improve more if they were on a campus in which receipt of Strand II awards was *less* prevalent (i.e., more competitive)?
- Did the outcomes of core teachers who received a Strand II award on the basis of the achievement gains of *their own students* improve more than those of comparable core teachers who received a Strand II award on the basis of the achievement gains of a *larger group of students*?

Award Amount

- Did the outcomes of core teachers who received a larger total ASPIRE award improve more than those of comparable core teachers who received a smaller total award?
- Is the relative amount of the award (relative to their salary and relative to the mean award amount on their campus) more salient to core teachers than the raw amount of the award?

Results

Estimated Effect of Receiving Awards on Core Teachers' Outcomes

This section focuses on the estimated independent effects of receiving 2009-10 Strand I, II, and/or III awards on core teachers' 2010-11 outcomes (retention, attendance, and mean student achievement gains), as well as variation in those effects for core teachers who work in higher need schools or teach hard-to-staff subjects. These results are estimated from random effects models (Model 1 in Appendices 6-11) that include a multitude of controls, with the goal of comparing core teachers who did and did not receive an award, with similar characteristics and who work in similar schools. Importantly, we further isolate the effect of receiving an award by including a control for teachers' 2008-09 attendance rates in models predicting 2010-11 attendance rates, and a control for teachers' 2008-09 CGIs (i.e., EVAAS scores) in models predicting 2010-11 student achievement gains. For teacher retention, estimated effects are represented by differences in the predicted probabilities of retention [(predicted probability for award recipients) – (predicted probability for non-award-recipients)]. For teacher attendance, estimated effects are represented by predicted differences in hours present during 2010-11 [(predicted count for award recipients) – (predicted count for non-award-recipients)]. For student achievement gains, estimated effects are represented by predicted differences in teachers' mean student test score gains from 2009-10 to 2010-11 [(predicted mean for award recipients) – (predicted mean for non-award-recipients)]. We use differences in these differences to contrast teachers in higher need schools to teachers in lower need schools, and to contrast teachers of hard-to-staff subjects to teachers not in hard-to-staff positions.

Core Teacher Retention

Figure 4 shows the estimated effects of receiving 2009-10 Strand I, II, and III awards on core teachers' probability of retention in August 2011. Net of the effect of other awards received, the predicted probability of retention for teachers who received a Strand I award is 3 percentage points higher on average than that of otherwise similar core teachers who did not receive a Strand I award. The differences are 5 percentage points

> Awards through each Strand have a positive effect, but Strand III awards appear to have the largest positive effect on teacher retention.

and 8 percentage points for core teachers who received Strand II and Strand III awards, respectively. Awards through each Strand have a positive effect, but Strand III awards appear to have the largest positive effect on teacher retention. In general, receiving awards improves the retention of core teachers in higher need schools less than comparable core teachers in lower need schools. Similarly, receiving awards improves the retention of teachers of hard-to-staff subjects less than comparable teachers not in hard-to-staff positions. In a marked exception, the predicted probability of retention for secondary level math teachers who received a Strand II award is 5 percentage points higher on average than that of comparable teachers not in hard-to-staff positions who received a Strand II award.

Core Teacher Attendance

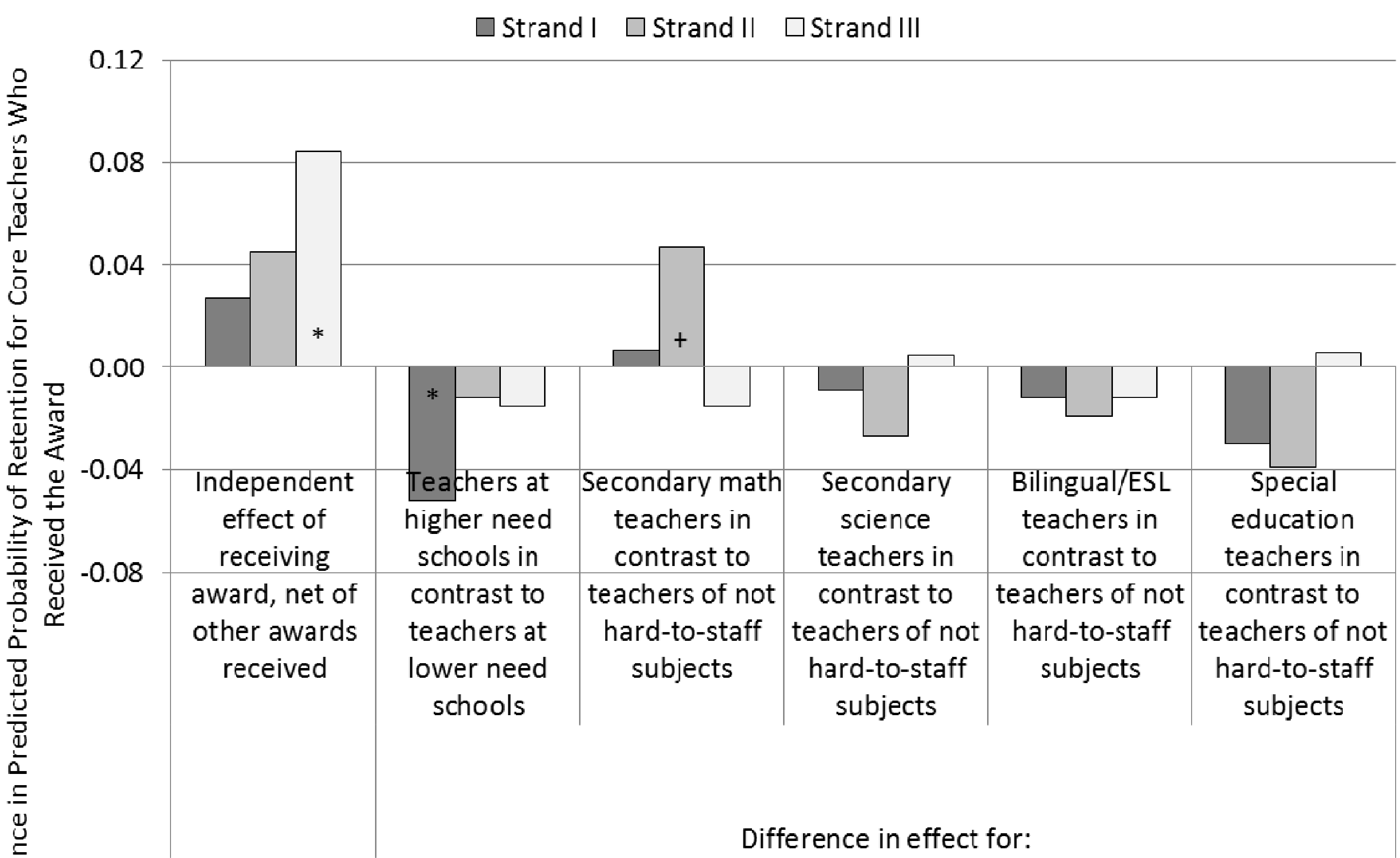
Figure 5 shows the estimated effects of receiving 2009-10 Strand I, II, and III awards on core teachers' 2010-11 attendance rates. Net of the effect of other awards received, core teachers who receive a Strand III award are predicted to be present an additional 15 hours in 2010-11 than otherwise similar teachers who did not receive a Strand III award *and had comparable attendance rates in 2008-09*. To put this finding into perspective, HISD teachers maintain mean attendance rates of 95%, or request to be absent an average of about ten days a year. A difference of 15 hours, or nearly 2 days of work, represents about a 20% improvement in attendance for teachers who receive a Strand III award. Receipt of a Strand II award has a smaller positive estimated effect on teacher attendance than receipt of a Strand III award, and receipt of a Strand I award actually has a negative estimated effect. With the exception of Strand I awards, the attendance of core teachers in higher need schools is improved even more by receipt of an ASPIRE award than that of comparable core teachers in lower need schools. The attendance of secondary level math and special education teachers is improved more by receipt of ASPIRE awards than that of teachers not in hard-to-staff positions, with the exception of Strand II awards for special education teachers. The attendance of secondary level science and bilingual/ESL teachers is improved less by receipt of ASPIRE awards than that of teachers not in hard-to-staff positions.

A difference of 15 hours, or nearly 2 days of work, represents about a 20% improvement in attendance for teachers who receive a Strand III award.

Math and Reading/ELA Teachers' Mean Student Achievement Gains

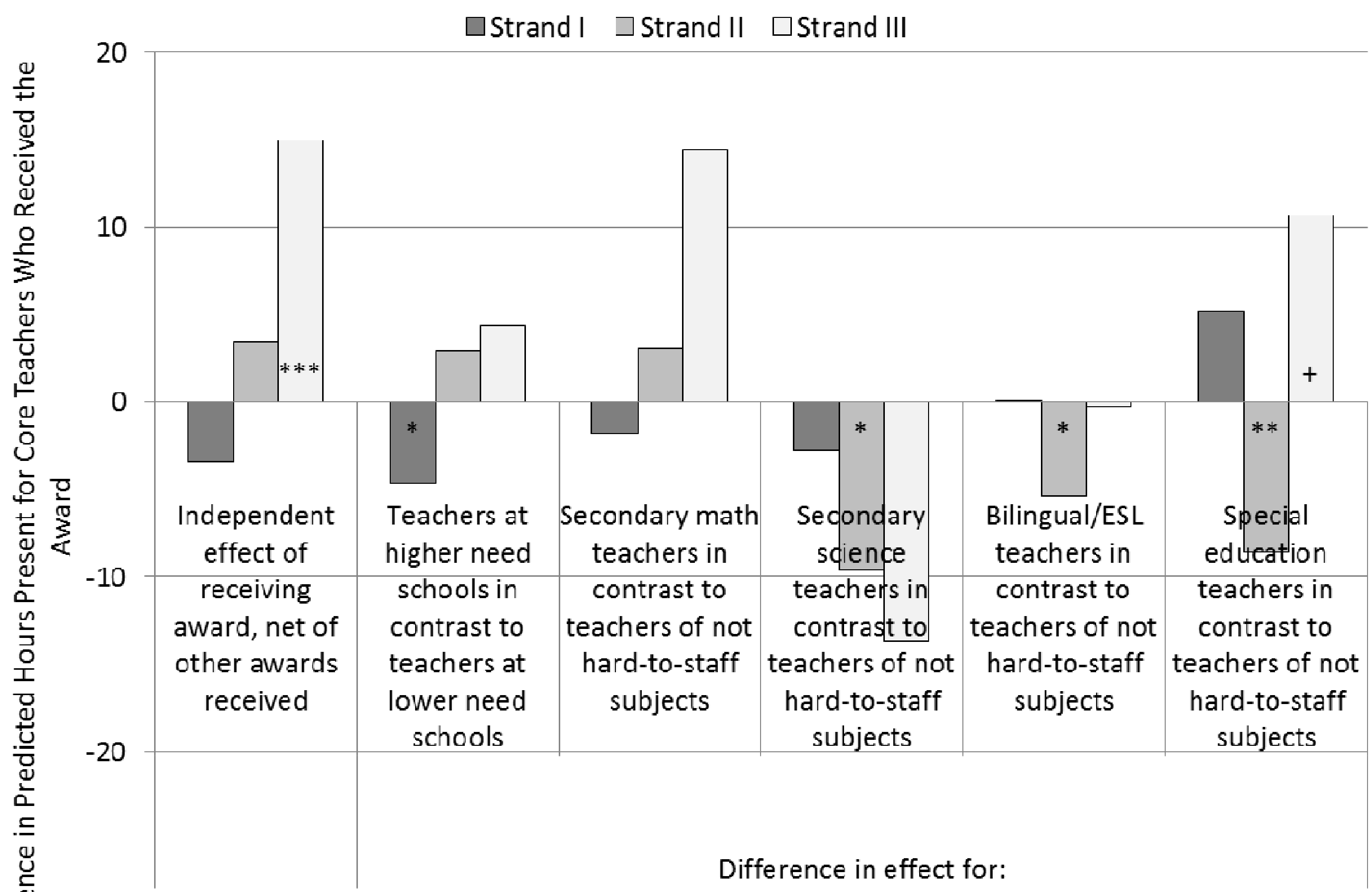
Figure 6 shows the estimated effects of receiving 2009-10 Strand I, II, and III awards on math and reading/ELA teachers' 2010-11 mean student test score gains. We standardized students' test score gains for each test within each grade level, which means a test score gain of zero represents the average gain on that test relative to other students in the same grade level and subject. Students with a positive test score gain experienced more gain than average, relative to others in the same grade level. Students with a negative test score gain experienced less gain than average, relative to others in the same grade level. We then aggregated students' gains to an average for each teacher, with the gains of students linked to multiple teachers contributing to the mean gains of multiple teachers.

Figure 4: Effect of Receiving 2009-10 Awards on Core Teachers' Retention in August 2011, and Differences for Teachers in Higher Need Schools or Hard-to-Staff Subjects



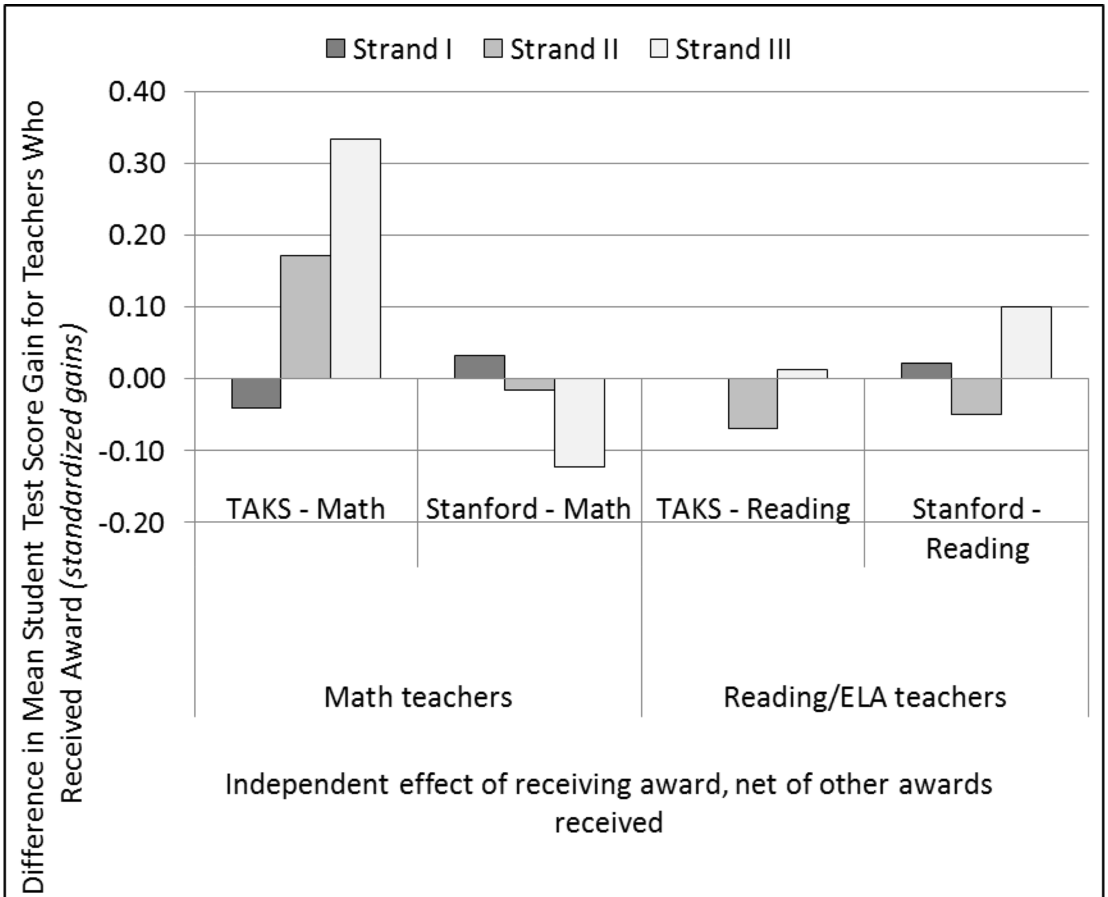
Note: Predicted differences estimated from Model 1 in Appendix 6. Because all controls are included in these models, comparisons are between core teachers similar in terms of their own and their schools' characteristics, with the exception of whether they received the award. This model excluded core teachers ineligible for the award program because of their job, campus, or opting out, as well as charter school teachers (who were, for the most part, not included in the retention data file).
+p<0.10, *p<0.05, **p<0.01, ***p<0.001.

Figure 5: Effect of Receiving 2009-10 Awards on Core Teachers' 2010-11 Attendance, and Differences for Teachers in High Need Schools or Hard-to-Staff Subjects



Note: Predicted differences estimated from Model 1 in Appendix 6. Because all controls are included in these models, comparisons are between core teachers similar in terms of their own and their schools' characteristics, with the exception of whether they received the award. This model excluded core teachers ineligible for the award program because of their job, campus, or opting out, as well as charter school teachers (who were, for the most part, not included in the retention data file). +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

Figure 6: Effect of Receiving 2009-10 Awards on Math and Reading/ELA Teachers' 2010-11 Mean Student Achievement Gains



Note: Predicted differences estimated from Models 1 in Appendices 8-11. Because all controls are included in these models, comparisons are between math or reading/ELA teachers similar in terms of their own and their schools' characteristics, with the exception of whether they received the award. These models excluded math or reading/ELA teachers ineligible for the award program because of their job, campus, or opting out. +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

Net of the effect of other ASPIRE awards received, math teachers who received a Strand III award are predicted to have mean student gains 0.33 standard deviations (SDs) higher on the math TAKS on average than otherwise similar math teachers who had comparable 2008-09 CGIs (i.e., EVAAS value-added scores) but did not receive a 2009-10 Strand III award. It is helpful to

Net of the effect of other ASPIRE awards received, math teachers who received a Strand III award are predicted to have mean student gains 0.33 standard deviations (SDs) higher on the math TAKS on average… This 0.33 SD difference could be perceived as an advantage that moves these teachers' students ahead of the students of more than 25% of their colleagues.

remember when interpreting the size of the effects that because of standardizing, approximately 25% of teachers had mean gain scores between 0 and 0.25 SDs, and 25% had mean gain scores between 0 and -0.25. This 0.33 SD difference could be perceived as an advantage that moves these teachers' students ahead of the students of more than 25% of their colleagues. Math teachers who received a Strand II award are predicted to have mean student gains 0.17 SDs higher on the math TAKS on average than otherwise similar math teachers who did not receive a Strand III award but had comparable 2008-09 CGIs. Mean Stanford math test score gains are actually predicted to be lower among math teachers who receive Strand III awards. The only difference of any note on reading test score gains is a slight advantage in Stanford reading test score gains among reading/ELA teachers who receive Strand III awards. Overall, receiving ASPIRE awards appears to be most beneficial for test score gains on the math TAKS.

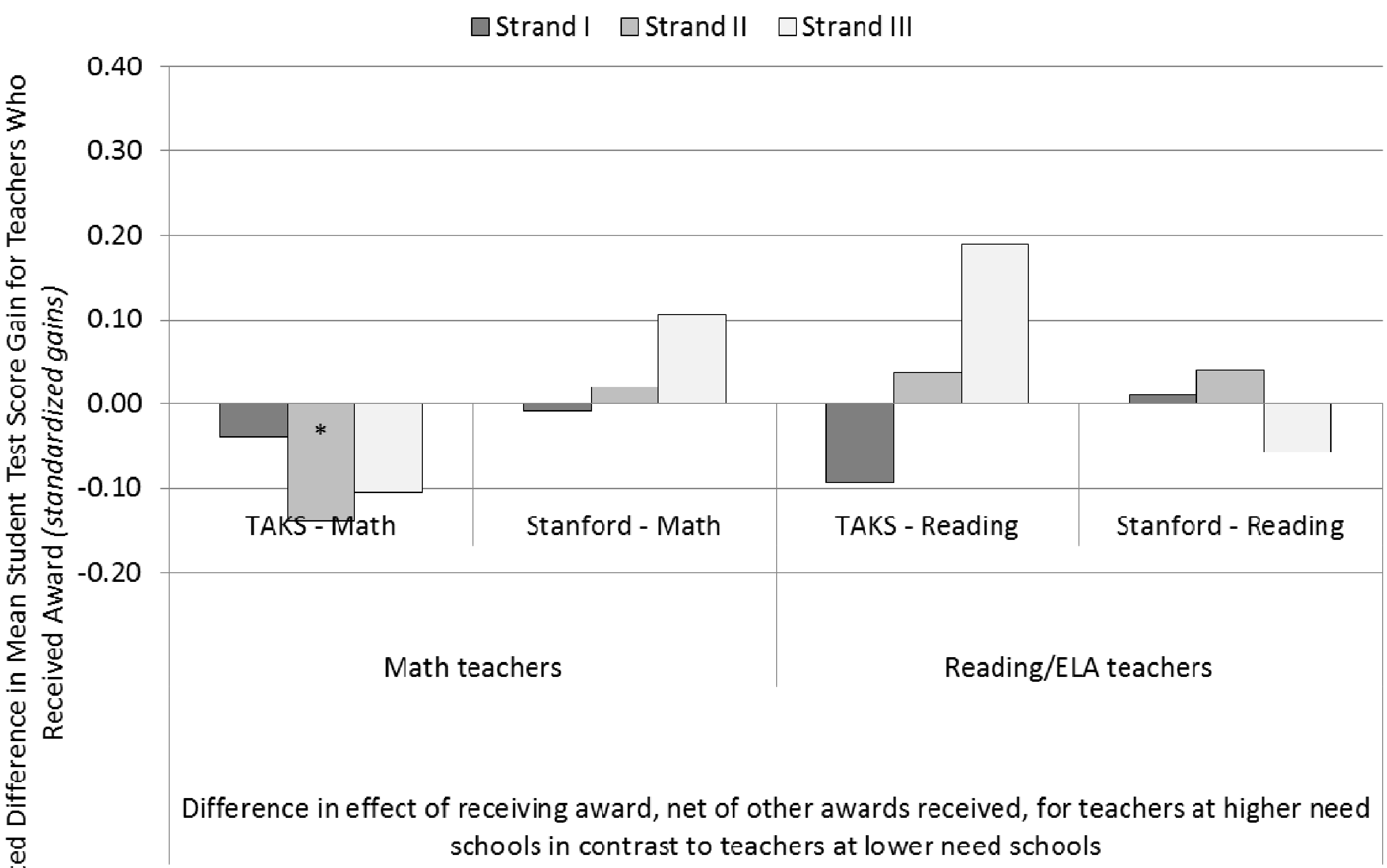
Overall, receiving ASPIRE awards appears to be most beneficial for test score gains on the math TAKS.

Figure 7 shows variation in the estimated effects of receiving 2009-10 Strand I, II, and III awards on math and reading/ELA teachers' 2010-11 mean student test score gains, depending on whether the teacher works in a higher need school. In one example, the mean math TAKS score gains of teachers in higher need schools who received Strand II awards are predicted to be 14 SDs lower than those of otherwise similar math teachers in lower need schools who received Strand II awards. In general, the most marked differences for teachers in higher need schools is that their mean gains on the math TAKS are improved less by receipt of ASPIRE awards, and their mean gains on the reading TAKS are improved slightly more by the receipt of a Strand III award.

Figure 8 shows variation in the estimated effects of receiving 2009-10 Strand I, II, and III awards on 2010-11 mean student test score gains for math teachers who work in a hard-to-staff position. In one example, the mean math TAKS score gains of secondary level math teachers who received a Strand III award are predicted to be 0.48 SDs lower than those of math teachers not in hard-to-staff positions (i.e., elementary level, not special education or bilingual/ESL) who received a Strand III award. Mean score gains on the Stanford math test are also less improved by receipt of a Strand II award for secondary level math teachers than for elementary level math teachers. In other differences of note, special education math teachers' gains on the math TAKS are less improved, and their gains on the Stanford math test are more improved, by receipt of a Strand III award than those of math teachers in not hard-to-staff positions (i.e., elementary level, not special education or bilingual/ESL).

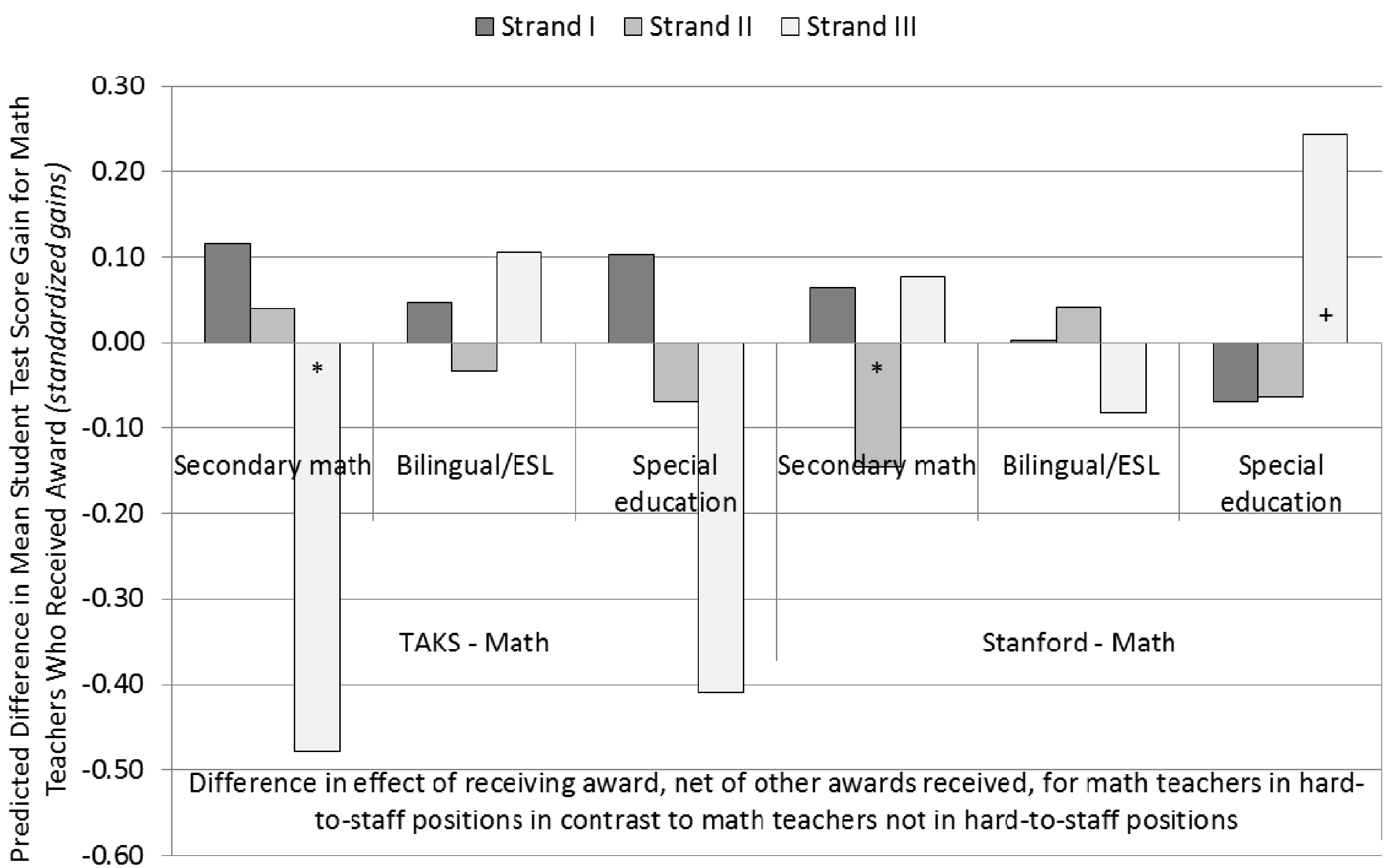
Figure 9 shows variation in the estimated effects of receiving 2009-10 Strand I, II, and III awards on 2010-11 mean student test score gains for reading/ELA teachers who work in hard-to-staff positions. In the only difference of real note, the mean reading TAKS score gains of special education math teachers who received a Strand III award are predicted to be 0.37 SDs lower than those of otherwise similar math teachers not in hard-to-staff positions (i.e., elementary level, not special education or bilingual/ESL) who also received a Strand III award.

Figure 7: Effect of Receiving 2009-10 Awards on 2010-11 Mean Student Achievement Gains for Teachers in Higher Need Schools



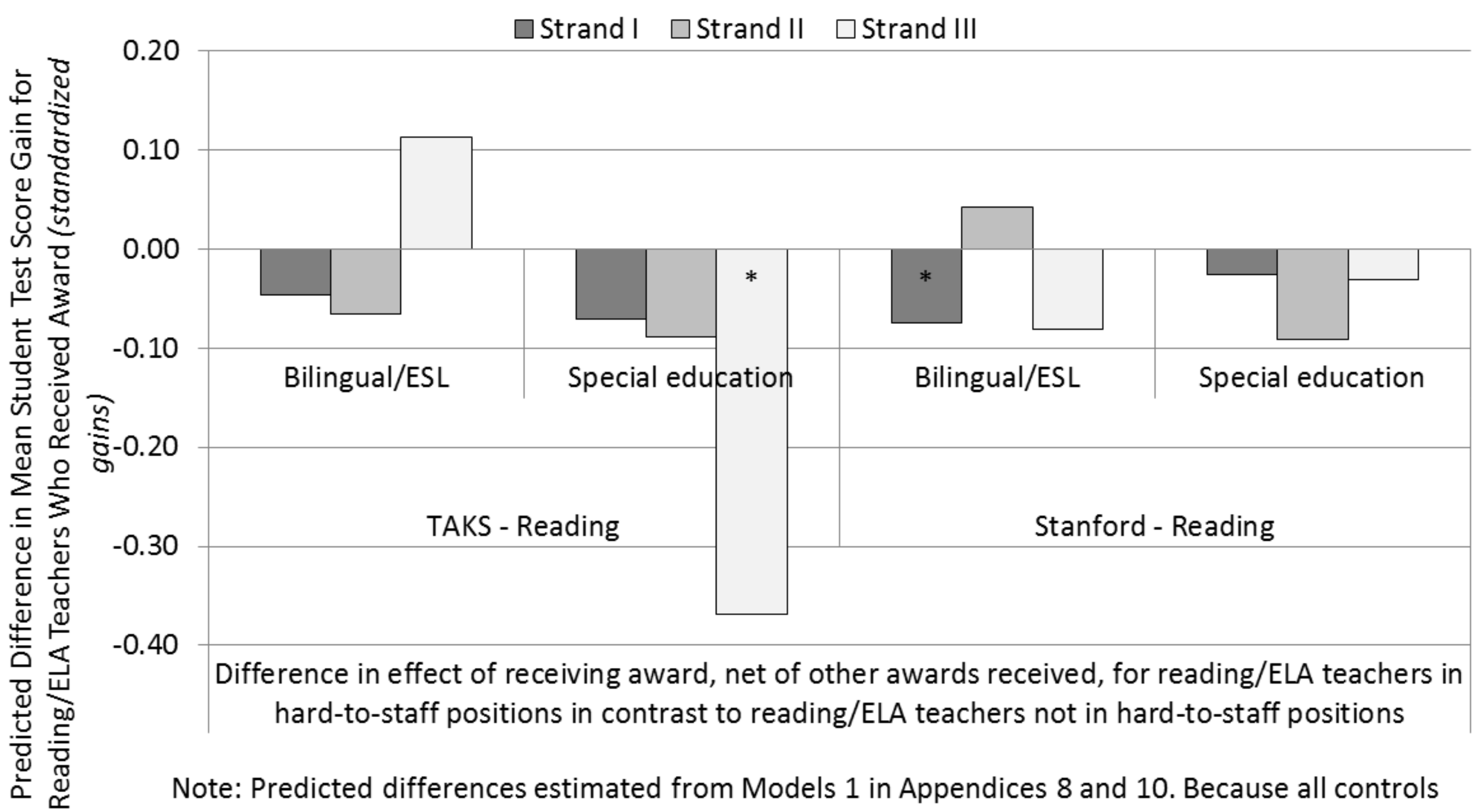
Note: Predicted differences estimated from Models 1 in Appendices 8-11. Because all controls are included in these models, comparisons are between math or reading/ELA teachers similar in terms of their own and their schools' characteristics, with the exception of whether they received the award. These models excluded math or reading/ELA teachers ineligible for the award program because of their job, campus, or opting out. +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

Figure 8: Effect of Receiving 2009-10 Awards on 2010-11 Mean Student Achievement Gains for Math Teachers in Hard-to-Staff Positions



Note: Predicted differences estimated from Models 1 in Appendices 7 and 9. Because all controls are included in these models, comparisons are between math teachers similar in terms of their own and their schools' characteristics, with the exception of whether they received the award. These models excluded math teachers ineligible for the award program because of their job, campus, or opting out. +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

Figure 9: Effect of Receiving 2009-10 Awards on 2010-11 Mean Student Gains for Reading/ELA Teachers in Hard-to-Staff Positions

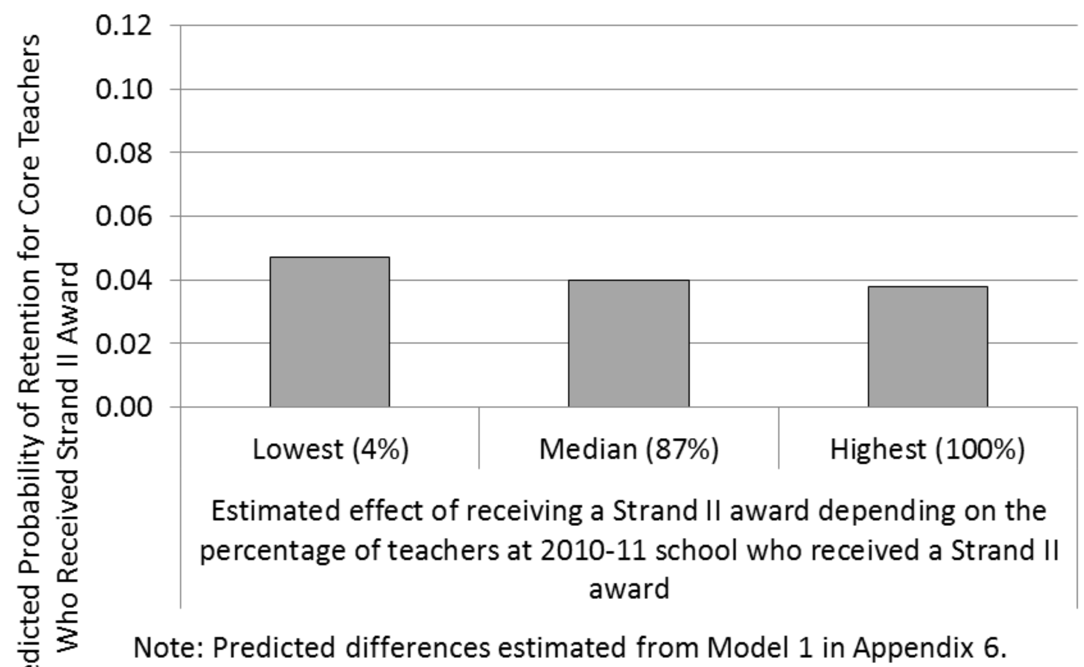


Note: Predicted differences estimated from Models 1 in Appendices 8 and 10. Because all controls are included in these models, comparisons are between reading/ELA teachers similar in terms of their own and their schools' characteristics, with the exception of whether they received the award. These models excluded reading/ELA teachers ineligible for the award program because of their job, campus, or opting out. +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

ESTIMATED EFFECT OF RECEIVING A STRAND II AWARD DEPENDING ON PREVALENCE

This section focuses on variation in the estimated effect of receiving a 2009-10 Strand II award, net of receipt of a Strand I or III award, depending on the prevalence of Strand II award receipt at the teacher's 2010-11 school. These results are estimated from random effects models (Model 1 in Appendices 6-11) that include a multitude of controls, with the goal of comparing core teachers who did and did not receive an award, with similar characteristics and who work in similar schools.

Figure 10: Effect of Receiving a 2009-10 Strand II Award on Core Teacher Retention in August 2011, Depending on Prevalence of Strand II Award Receipt on 2010-11 Campus



Note: Predicted differences estimated from Model 1 in Appendix 6. Because all controls are included in these models, comparisons are between employees similar in terms of their own and their schools' characteristics, with the exception of the characteristics of interest. This model excluded teachers ineligible for the award program because of their job, campus, or opting out, as well as charter school teachers (who were, for the most part, not included in the retention data file). 4% represents the lowest prevalence of Strand II award winning among schools where at least one teacher won a Strand II award. The differences by prevalence of Strand II award winning were not statistically significant.

Figure 10 shows variation in the estimated effect of receiving a 2009-10 Strand II award on core teachers' probability of retention in August 2011, depending on the percentage of teachers at their 2010-11 school who also received Strand II awards. At schools where 4% of eligible core teachers received a Strand II award (the lowest percentage), the predicted probability of retention for a teacher who received a Strand II award is 5 percentage points higher than that of otherwise similar teachers who did not receive a Strand II award. At schools where all eligible core teachers received a Strand II award, the predicted probability of retention for a teacher who received a Strand II award is 4 percentage points higher than that of otherwise similar teachers who did not receive a Strand II award. This suggests teachers respond more positively to receiving Strand II awards in schools in which their receipt is less prevalent, but the difference is small.

Figure 11: Effect of Receiving a 2009-10 Strand II Award on Core Teachers' 2010-11 Attendance, Depending on Prevalence of Strand II Award Receipt on 2010-11 Campus

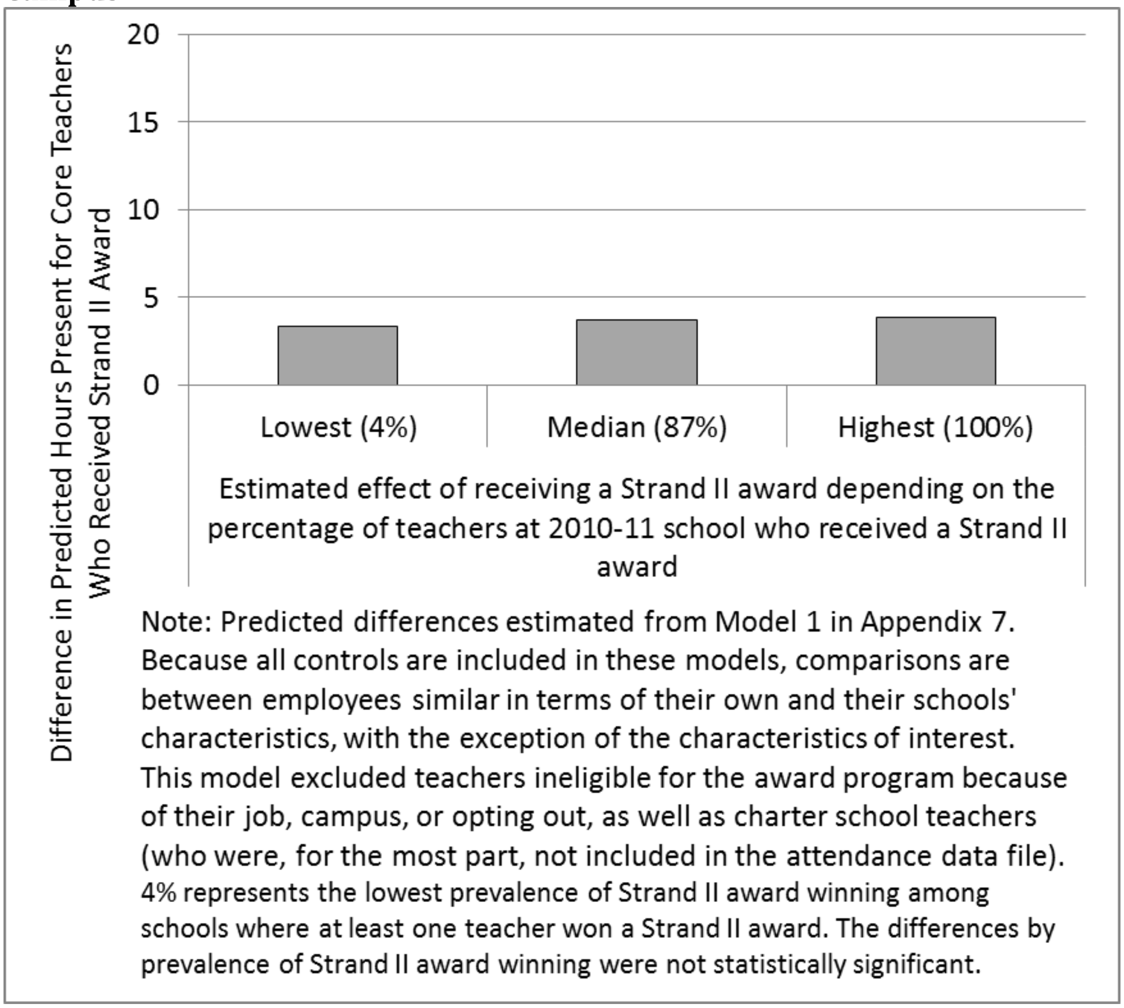
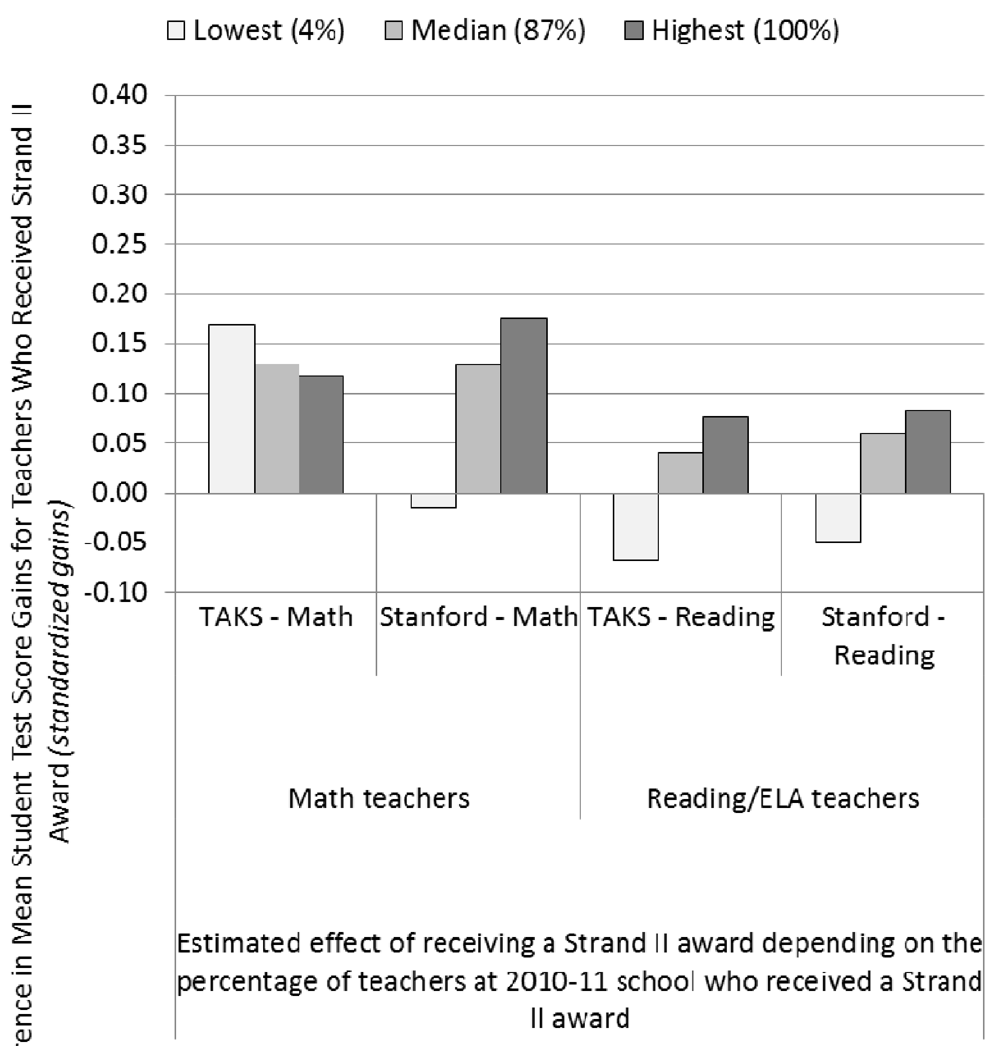


Figure 11 shows variation in the estimated effect of receiving a 2009-10 Strand II award on core teachers' 2010-11 attendance rates, depending on the percentage of teachers at their 2010-11 school who also received Strand II awards. Although the differences are very small, core teachers' attendance is more improved by receiving a Strand II award at a school in which Strand II receipt is more common than at a school in which it is less common.

Figure 12 shows variation in the estimated effect of receiving a 2009-10 Strand II award on math and reading/ELA teachers' 2010-11 mean student achievement gains, depending on the percentage of teachers at their 2010-11 school who also received Strand II awards. We see slightly more differentiation on this outcome than retention or attendance, although not in consistent directions. For the math TAKS, mean student test score gains are most improved by receipt of a Strand II award if the math teacher worked at a school in which receipt of Strand II awards was less prevalent. Mean student test score gains on the other three tests were most improved by receipt of a Strand II award if the teacher worked at a school in which receipt of Strand II awards was more prevalent.

Figure 12: Effect of Receiving a 2009-10 Strand II Award on Math and Reading/ELA Teachers' 2010-11 Mean Student Achievement Gains, Depending on Prevalence of Strand II Award Receipt on 2010-11 Campus



Note: Predicted differences estimated from Models 1 in Appendices 7-10. Because all controls are included in these models, comparisons are between math or reading/ELS teachers similar in terms of their own and their schools' characteristics, with the exception of the characteristics of interest. This model excluded math or reading/ELA teachers ineligible for the award program because of their job, campus, or opting out. 4% represents the lowest prevalence of Strand II award winning among schools where at least one teacher won a Strand II award. The differences by prevalence of Strand II award receipt were statistically significant for Stanford - math ($p<0.05$) and Stanford-reading ($p<0.10$).

Estimated Effect of Receiving a Strand II Award Depending on Award Criteria

This section focuses on variation in the estimated effect of receiving a 2009-10 Strand II award, net of receipt of a Strand I or III award, depending upon the group of students whose achievement gains determined the Strand II award. These results are estimated from random effects models (Model 2 in Appendices 6-11) that include a multitude of controls, with the goal of comparing core teachers who received an award to core teachers who did not, with similar characteristics and who work in similar schools.

Figure 13 shows variation in the estimated effect of receiving a 2009-10 Strand II award on core teachers' retention in August 2011, depending on how their Strand II award was determined (i.e., the teacher's ASPIRE category). Self-contained core teachers' (grades 3-6) predicted probability of retention was improved more by the receipt of a Strand II award than that of any other type of core teacher (nearly 10 percentage points). Although departmentalized core teachers of grades 3-8 also receive Strand II awards on the basis of the students they actually teach, their predicted probability of retention only increased by 4 percentage points on average upon receipt of a Strand II award. The retention of core teachers who receive Strand II awards on the basis of the students in their department (teachers of grades 9-12) experienced the second largest degree of improvement through receipt of a Strand II award (6 percentage points).

Self-contained core teachers' (grades 3-6) predicted probability of retention was improved more by the receipt of a Strand II award than that of any other type of core teacher (nearly 10 percentage points).

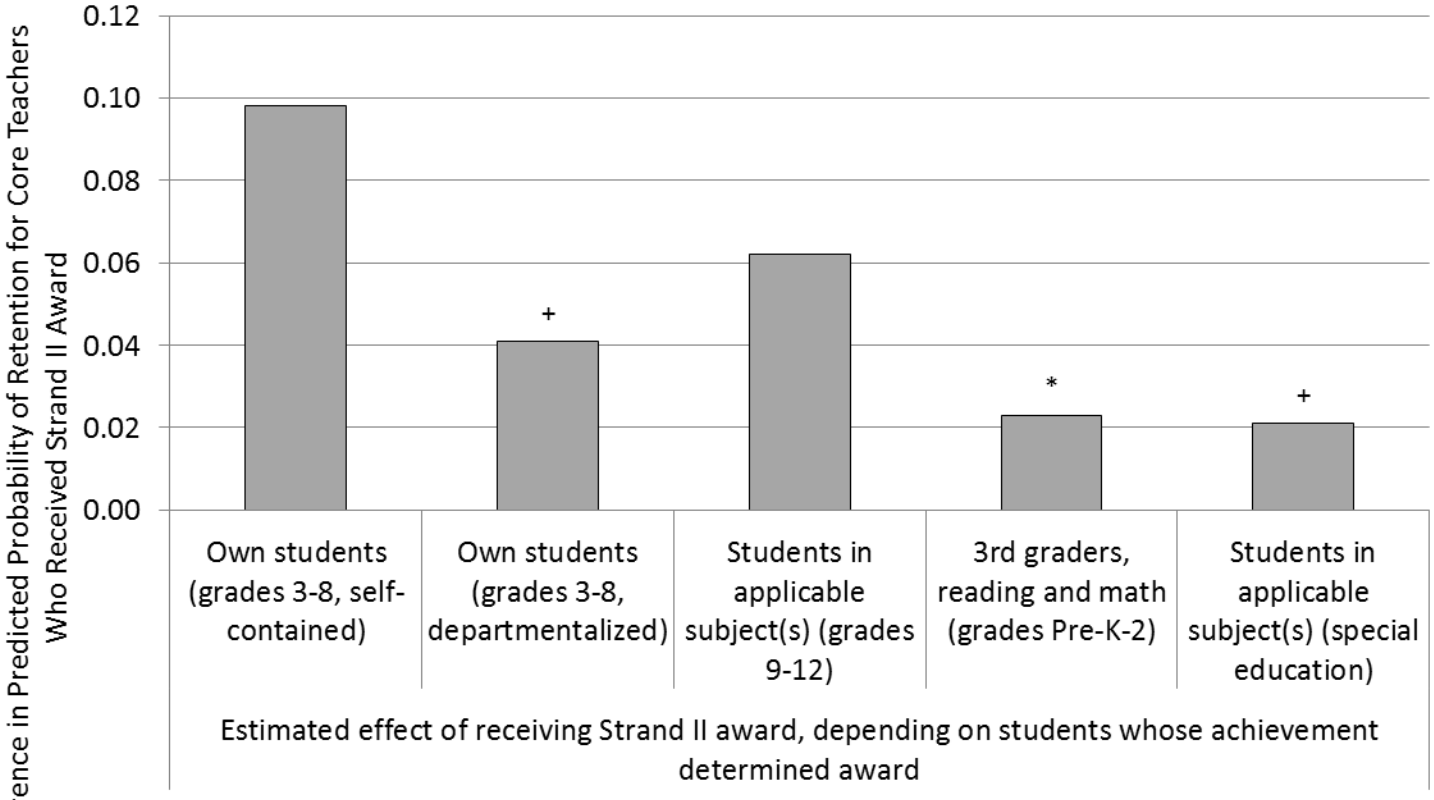
Figure 14 shows variation in the estimated effect of receiving a 2009-10 Strand II award on core teachers' 2010-11 attendance rates, depending on how their Strand II award was determined (i.e., the teacher's ASPIRE category). The attendance of core teachers who receive Strand II awards on the basis of the students in their department (teachers of grades 9-12) was improved more by receipt of a Strand II award than that of any other group of core teachers (an additional 5.3 hours present).

Across all four tests, but the reading TAKS, the mean student test score gains of math and reading/ELA teachers who received Strand II awards on the basis of the students they actually teach were more improved than those of teachers who received Strand II awards on the basis of a different group of students.

Figure 15 shows variation in the estimated effect of receiving a 2009-10 Strand II award on core teachers' 2010-11 mean student achievement gains, depending on how their Strand II award was determined (i.e., the teacher's ASPIRE category). Across all four tests, but the reading TAKS, the mean student test score gains of math and reading/ELA teachers who received Strand II awards on the basis of the students they actually teach were more improved than those of teachers who received Strand II awards on the basis of a different group of students. The mean student achievement gains of special education teachers were notably negatively impacted by receipt of a Strand II award, as were high school teachers' mean student gains on the Stanford math test.

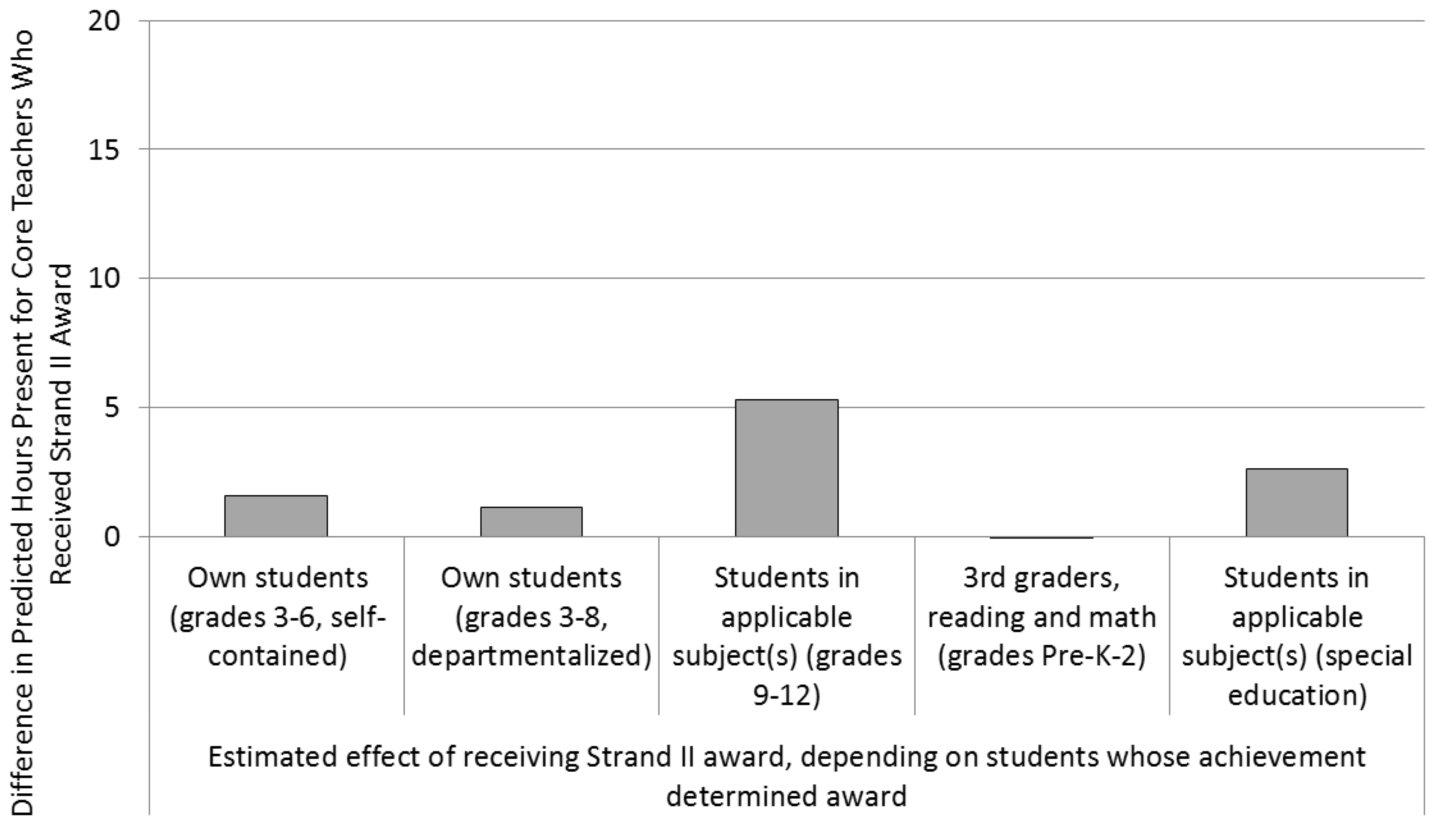
The mean student achievement gains of special education teachers were notably negatively impacted by receipt of a Strand II award, as were high school teachers' mean student gains on the Stanford math test.

Figure 13: Effect of Receiving a 2009-10 Strand II Award on Core Teachers' Retention in August 2011, Depending on Method of Strand II Award Determination



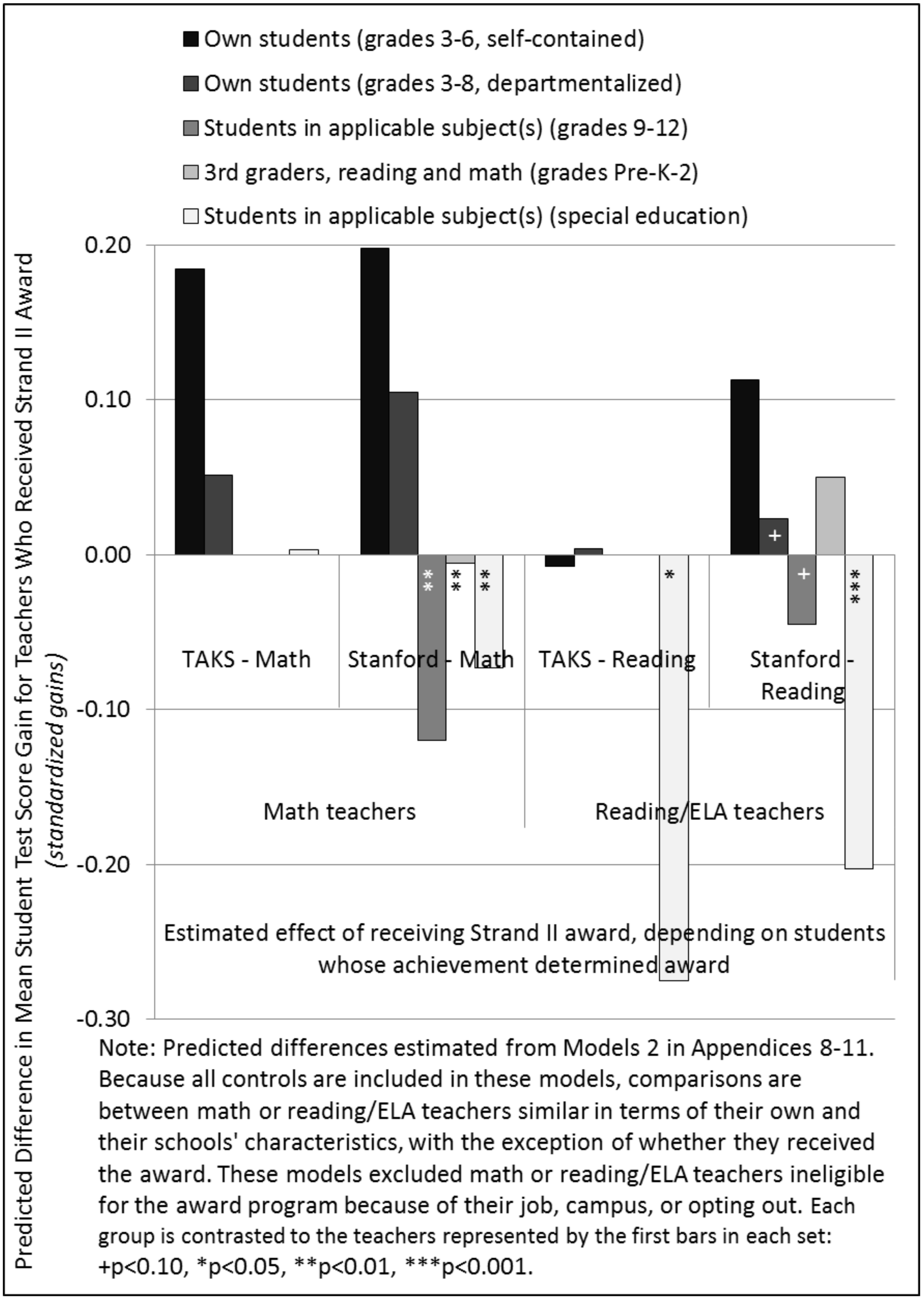
Note: Predicted differences estimated from Model 2 in Appendix 6. Because all controls are included in these models, comparisons are between core teachers similar in terms of their own and their schools' characteristics, with the exception of whether they received the award. This model excluded teachers ineligible for the award program because of their job, campus, or opting out, as well as charter school teachers (who were, for the most part, not included in the retention data file). Each group is contrasted to the teachers represented by the first bar: +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

Figure 14: Effect of Receiving a 2009-10 Strand II Award on Core Teachers' 2010-11 Attendance, Depending on Method of Strand II Award Determination



Note: Predicted differences estimated from Model 2 in Appendix 7. Because all controls are included in these models, comparisons are between core teachers similar in terms of their own and their schools' characteristics, with the exception of whether they received the award. This model excluded teachers ineligible for the award program because of their job, campus, or opting out, as well as charter school teachers (who were, for the most part, not included in the attendance data file). Each group is contrasted to the teachers represented by the first bar: +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

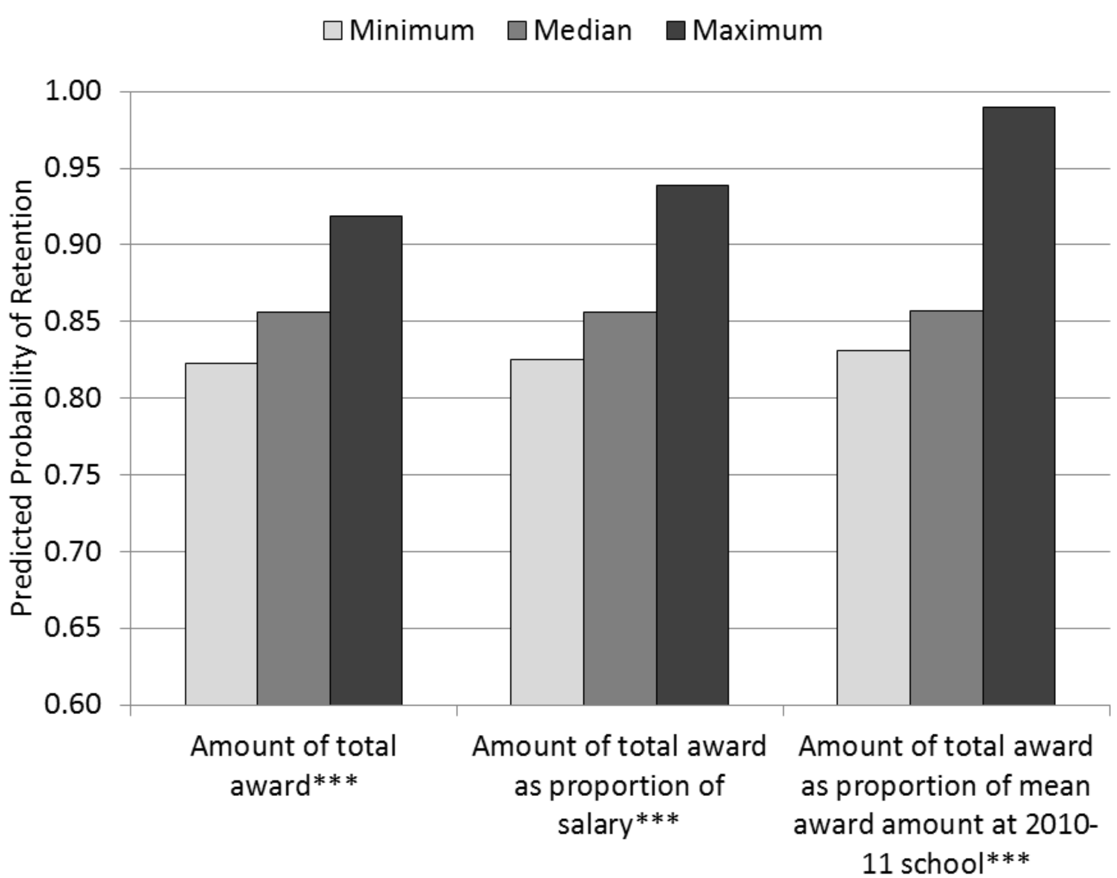
Figure 15: Effect of Receiving a 2009-10 Strand II Award on Math and Reading/ELA Teachers' 2010-11 Mean Student Achievement Gains, Depending on Method of Strand II Award Determination



Award Amounts

This section focuses on variation in the estimated effect of receiving any 2009-10 award, depending on the raw amount of the total award, the amount of the award relative to the core teacher's annual salary, and the amount of the award relative to the mean award amount on the teacher's 2010-11 campus. We focus on the mean award amount in the teacher's 2010-11 school because teachers are notified of their 2009-10 award status and amount in January 2011. While most core teachers did not change schools between 2009-10 and 2010-11, we expect that teachers who did change schools still use the award amounts in their immediate context as a primary base of comparison. These results are estimated from random effects models (Models 3-5 in Appendices 6-11) that include a multitude of controls, with the goal of comparing core teachers who received an award to core teachers who did not, but have similar characteristics and work in similar schools.

Figure 16: Effect of Receiving 2009-10 Award on Core Teachers' Retention in August 2011, Depending on Raw and Relative Award Amount

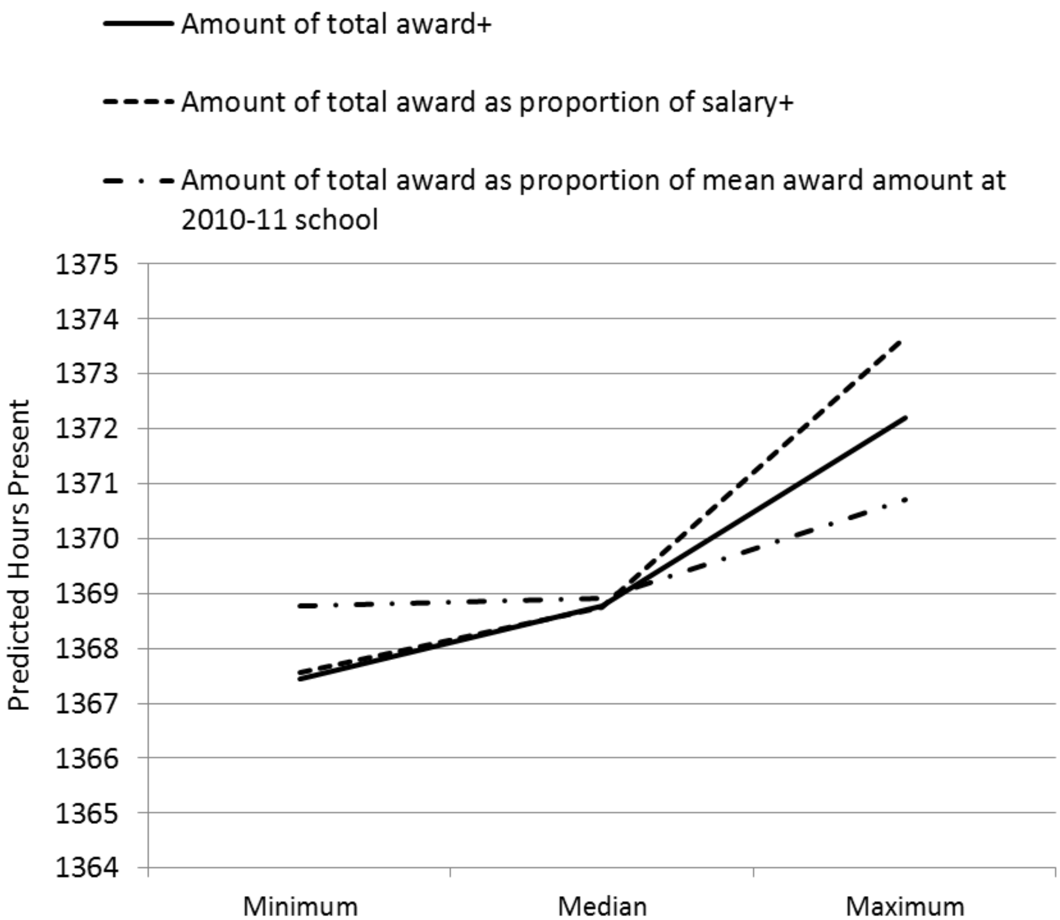


Note: Predicted probabilities estimated from Models 3-5 in Appendix 6. Because all controls are included in these models, comparisons are between teachers who are similar in terms of their own and their schools' characteristics, with the exception of their award amount. These models exclude teachers who did not win an award, as well as charter school teachers (who were, for the most part, not included in the retention data file). +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

Figure 16 shows that the predicted probability of retention was 0.92 for core teachers who received an award of the maximum amount, but only 0.82 for otherwise similar core teachers who received an award of the minimum amount. The difference is even more marked when comparing core teachers whose awards as proportions of their annual salary were at the maximum and minimum (0.94 versus 0.83). We see the most marked differences, though, when comparing core teachers with minimum and maximum awards relative to the mean award amount on their campus (0.99 versus 0.83). This pattern was also evident across mean student test score gains on all four tests (Figures 18, we only show the results for the math TAKS for this reason), with teachers' award amounts relative to the mean amount on their campus being most salient for higher mean test score gains. In Figure 17, core teachers' attendance improved more as their award amount increased relative to their salary than it did as the raw amount increased, but, unlike the other outcomes, teachers' award amount relative to their salary was more salient to improvements in their attendance than their award amount relative to the mean award amount on their campus.

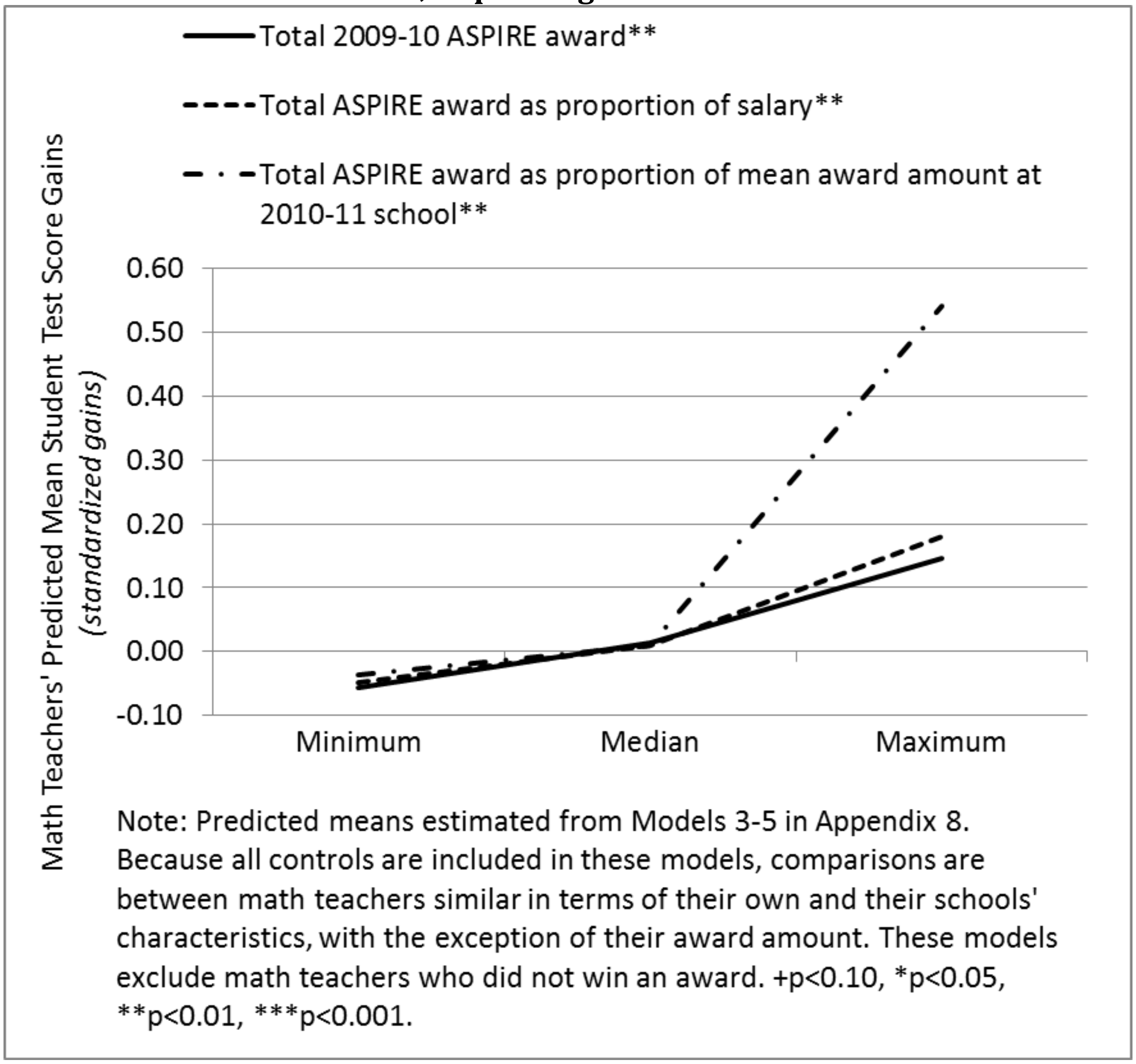
We see the most marked differences, though, when comparing core teachers with minimum and maximum awards relative to the mean award amount on their campus (0.99 versus 0.83).

Figure 17: Effect of Receiving 2009-10 Award on Core Teachers' 2010-11 Attendance, Depending on Raw and Relative Award Amount



Note: Predicted counts estimated from Models 3-5 in Appendix 7. Because all controls are included in these models, comparisons are between employees who are similar in terms of their own and their schools' characteristics, with the exception of their award amount. These models exclude teachers who did not win an award, as well as charter school teachers (who were, for the most part, not included in the attendance data file). +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

Figure 18: Effect of Receiving 2009-10 Award on Math Teachers' 2010-11 Mean Student Gains on Math TAKS, Depending on Raw and Relative Award Amount



SUMMARY OF RESULTS

RECEIVING AN AWARD IMPROVES TEACHER PRODUCTIVITY

Receiving awards does appear to have a positive estimated effect on core teachers' retention, attendance rates, and mean student gains on the math TAKS, net of any effects of award eligibility.

TEACHERS IN HIGHER NEED SCHOOLS BENEFIT LESS FROM RECEIVING AWARDS

In general, the retention and mean student gains on the math TAKS of core teachers in higher need schools were improved less by the receipt of awards than those of core teachers in lower need schools. It may be that financial incentives are less motivating for teachers in higher need schools because of the relatively higher degree of daily challenges in such schools (Ramirez 2001). In contrast, the attendance and mean student gains on the reading TAKS of core teachers in higher need schools were improved more by the receipt of awards (only Strands II and III) than those of core teachers in lower need schools.

Math Teachers Respond More Positively, and Special Education Teachers Less Positively, to Receiving Awards

Most districts primarily hope to change the recruitment and retention patterns of teachers of hard-to-staff subjects. Although teachers of hard-to-staff subjects were more likely to be retained if they received an award, their likelihood of retention was improved less by receiving awards than that of teachers not in hard-to-staff positions. This suggests that while ASPIRE may be contributing positively to district goals for teachers in general, supplementary interventions may be required for teachers of hard-to-staff subjects.

Other distinctions primarily arose for math and special education teachers. Secondary level math teachers responded more positively to receipt of a Strand II award than any other core teachers, as evident in a much improved probability of retention and more improvement in their attendance rates. Unlike reading/ELA teachers, math teachers, as evidenced by their mean student gains on the math TAKS, responded more positively to a Strand II award if they received it on a campus in which Strand II awards were not prevalent. These findings may suggest that math teachers are more responsive to competitive awards than other teachers, or that they are more strategic in their pursuit of awards. It is also possible that the mathematical backgrounds of math teachers may lead them to engage more with the details of the ASPIRE program.

These findings may suggest that math teachers are more responsive to competitive awards than other teachers, or that they are more strategic in their pursuit of awards. It is also possible that the mathematical backgrounds of math teachers may lead them to engage more with the details of the ASPIRE program.

Mean student achievement gains on the math TAKS, though, were predicted to be much lower for secondary level math teachers who received a Strand III award than for elementary level math teachers who received a Strand III award. Elementary schools receive Strand III awards, in part, on the basis of Comparable Improvement rankings, whereas high schools receive Strand III awards, in part, on the basis of increased enrollment in AP/IB coursework. It may be that secondary level math teachers' subsequent lesser gains reflect the increased enrollment of less prepared students in higher level coursework.

The mean student achievement gains of special education teachers who received awards were typically much lower than those of other teachers who received awards. The attendance of special education teachers was more improved by the receipt of a Strand III award, but less by the receipt of a Strand II award, than that of teachers not in hard-to-staff positions. These results possibly reflect the lower control special education teachers have over the achievement of their students. Improving their attendance rates may be one of the few obvious steps they can take to improve their chances of receiving another award.

The mean student achievement gains of special education teachers who received awards were typically much lower than those of other teachers who received awards.

The Most Effective Awards Blend Competition, School-Wide Support, and Accountability Criteria

Some findings suggest that core teachers were most responsive to collaborative awards. The positive effects of receiving a Strand II award on teacher attendance, and student achievement gains on all four tests but the math TAKS, were magnified for teachers at schools with the highest prevalence of Strand II award receipt. Strand III awards, which are based on campus level measures of achievement (implying collaboration), more effectively improved teachers' probability of retention, attendance rates, and student achievement gains than Strand I and II awards. Several other findings, though, suggested that teachers were most responsive to competitive awards. The positive effect of receiving an award on retention was magnified for teachers working at schools with the lowest prevalence of award receipt. Across all outcomes, teachers responded most favorably to awards that were large relative to the mean amount of the awards received on their campus. Across all outcomes, the estimated effect of receiving a Strand II award that was determined on the basis of the achievement of the teachers' own students, or the students in the teacher's department, was typically larger than that of Strand II awards determined on the basis of a larger group of students. Moreover, if Strand III was effective because it was collaboratively based on campus level achievement, it would follow that Strand I would also be more effective than Strand II. However, teachers' attendance rates and mean student achievement gains (on the TAKS) were actually negatively impacted by receipt of a Strand I award.

Across all outcomes, teachers responded most favorably to awards that were large relative to the mean amount of the awards received on their campus.

We interpret the apparent success of collaborative awards as an indicator of the importance of school context and school-wide initiatives. First, the positive effect of receiving an award on student achievement gains was more evident on the TAKS than Stanford tests, which suggests teachers focused on goals emphasized within their schools rather than goals emphasized by the ASPIRE program (TAKS is used for state accountability ratings). In another example of the influence of school context on the effect of ASPIRE, teachers at higher need schools were less likely to receive awards, and their outcomes were less improved by receiving an award than those of teachers at lower need schools. Rather than an indication of teachers' preference for collaborative awards, Strand III, a school level award, may have been most effective because teachers that received this award worked in schools with contexts or school-wide policies that foster teacher success. This interpretation may also explain why teachers sometimes seemed to be more responsive to collaborative awards (i.e., teachers benefitted more from receiving an award in schools in which award receipt was prevalent because these schools' contexts or structural changes support their success).

Across all outcomes, the estimated effect of receiving a Strand II award that was determined on the basis of the achievement of the teachers' own students, or the students in the teacher's department, was typically larger than that of Strand II awards determined on the basis of a larger group of students.

The structure of Strand III may have fostered the development of school-wide initiatives in ways that Strands I and II did not because of its focus on a more targeted set of academic goals (e.g., AP class enrollment, math and reading test scores). Strand I's focus on campus-wide gains in all subjects may be too abstract and broad to feel attainable to individual teachers. The clear set of targets outlined by Strand III may have facilitated structural and procedural change that benefitted teachers and students in a lasting way: teachers at schools that received a 2009-10 Strand III award continued to reap the benefits of those changes in 2010-11, as evidenced by their larger student achievement gains. Campus morale may also generally be higher at schools that receive Strand III awards, making teachers less likely to quit or to disappoint their colleagues with poor attendance. The fact that Strand I awards are determined by test score gains, while Strand III awards are determined by test scores, may also contribute to differences in the influence of awards from each of these Strands. Receipt of Strand III awards is also determined by schools' TEA ratings, which HISD staff cite as a primary focus among school administrators. In other words, Strand III may be most effective because it incorporates the state-level accountability criteria that are a dominant concern at schools. For all of these reasons, award programs that offer some competitive elements, while also being grounded in wider accountability criteria and school wide initiatives, may be the most effective programs. A major limitation of these findings is that very few schools do not receive Strand III awards, making it possible that the apparent effect of receiving a Strand III award is actually attributable to unmeasured distinctions between the schools that do and do not receive these awards. Additional data collection, as discussed below, could provide more (or less) support for our interpretation of these findings.

Larger Awards are More Effective than Smaller Awards

Teachers' retention, attendance, and mean student achievement gains were more improved by a larger award than by a smaller award. The differences were more marked when we compared the outcomes of teachers who received larger awards relative to their annual salary, suggesting that teachers' interpretation of the size of their award is somewhat dependent on the size of their annual salary. Teachers whose awards were large relative to the mean amount of the awards received on their campus experienced the most improved outcomes, suggesting that teachers' interpretation of the size of their awards is most dependent on its size relative to that of their colleagues. This also supports the idea that teachers are most responsive to competitive awards.

…more improved by a larger award than a smaller award… teachers' interpretation of the size of their awards is most dependent on its size relative to that of their colleagues…

Policy Implications

Reduce the Number of Awards and Increase the Amounts

The theoretical origins of award programs, and our findings that support the idea that teachers are more responsive to competitive awards, suggest that the effectiveness of ASPIRE may be improved if fewer teachers receive awards. The distribution of fewer awards would release funds for larger awards, which we found to be more effective than smaller awards. A particularly high proportion of schools receive Strand III awards, which may reduce the meaningfulness and impact of this Strand of awards. The rates of award receipt are also very high among school administrators and some subsets of core teachers. Furthermore, evaluations of ASPIRE would be facilitated by a program design that includes a group of teachers who can serve as a true base of comparison (i.e., a larger group of more comparable non-award recipients).

The theoretical origins of award programs, and our findings that support the idea that teachers are more responsive to competitive awards, suggest that the effectiveness of ASPIRE may be improved if fewer teachers receive awards.

Eliminate Strand I

The apparent negative effect of receiving a Strand I award is troubling. It is possible that schools that received this award were characterized by a high-pressure environment that ultimately resulted in teacher burnout. In contrast, independent positive effects of receiving Strand II and III

The apparent negative effect of receiving a Strand I award is troubling… Narrowing the focus of the award program to Strands II and III might also simplify the design of the program, thereby increasing teacher understanding of and buy-in to the program.

awards were evident across most outcomes for most groups of interest. Narrowing the focus of the award program to Strands II and III might also simplify the design of the program, thereby increasing teacher understanding of and buy-in to the program.

Provide Teachers with the Tools They Need to be Successful

Several findings indicated that ASPIRE might be more effective if skill building were interwoven into the program. The outcomes of teachers in higher need schools and special education teachers were less improved by award receipt, perhaps because these teachers were unsure how to ensure achievement gains among their more disadvantaged students. The relationship between receiving an award and teacher attendance was often distinct from the relationship between award receipt and other outcomes. For instance, Strand II had the largest positive effect on the attendance of high school teachers, whereas elementary level teachers' retention and achievement gains were most responsive to Strand II awards. High school teachers receive Strand II awards on the basis of the achievement of all students in their department (i.e., their subject), and it may be that maintaining good attendance was a key way to signal high effort to colleagues. The attendance of secondary level math and special education teachers was more improved by receipt of a Strand III award than that of all other core teachers, as was the attendance of teachers in higher need schools in contrast to that of teachers in lower need schools; with the heightened degree of challenge and pressure inherent in these positions, maintaining good attendance may be perceived as the most straightforward and relatively simple step to take toward increasing one's likelihood of an award. From the perspective that there is not a tremendous amount of variation in attendance rates, and that teacher attendance is only one small contributor to student success, the possibility arises that teachers in more challenging positions are focusing on symbolic rather than concrete changes.

The outcomes of teachers in higher need schools and special education teachers were less improved by award receipt, perhaps because these teachers were unsure how to ensure achievement gains among their more disadvantaged students.

Although TIF grants specify that professional development should be an integral part of award programs, it was not clear how professional development opportunities were interwoven with ASPIRE (outside of online courses describing the structure of the award program). Readily available sources of professional development are an integral part of some award programs, as are rewards for completing professional development courses (Hulleman and Barron 2010; Koppich 2005). Other programs designate successful teachers as lead teachers or mentors (and provide extra compensation) with the goal of building capacity through peer mentorship (Perkins-Gough 2007); this is also a criteria of TIF grants. Integrating initiatives like these into ASPIRE might provide teachers with the tools they need to be more effective and thereby make ASPIRE more effective.

Although TIF grants specify that professional development should be an integral part of award programs, it was not clear how professional development opportunities were interwoven with ASPIRE (outside of online courses describing the structure of the award program).

Improve the Identification of Quality Teachers

Despite not using student achievement levels to construct our measure of a high need school, teachers in these schools were less likely to receive awards, which may be an indication that value-added scores do not capture the differences across student bodies that influence teachers' capacities to enable achievement gains. In addition to calling into question the methods by which teachers are evaluated for an award, facts like these may decrease the degree to which teachers buy into the program. Although ASPIRE awards are determined on the basis of a complex collection of criteria, the vast majority of the criteria involve students' scores on standardized achievement tests. Other award programs incorporate principal or peer evaluations and engagement in professional development, as other indicators of teacher quality (Jacob and Lefgren 2008; Koppich 2005; Podgursky and Springer 2007). More recent experiments are finding that student evaluations of teachers provide valuable and apparently reliable insights into differences across teachers (Ferguson 2012). This issue is certainly larger than HISD. The very theories that motivate programs like these question the validity of educational attainment and years of experience as valid measures of teacher quality, and the debate on how to accurately and fairly evaluate teachers persists. HISD's large and growing set of data files could potentially enable the exploration of the teacher and school qualities most closely associated with student achievement. Better measures of teacher quality would be of major interest to policymakers, practitioners, and researchers.

Increase Non-Merit-Based Financial Incentives for Teachers in Hard-to-Staff Positions

The findings of this evaluation suggest that factors outside of the teachers' own characteristics, such as the degree to which their school is high need, or the subject they teach, differentiate the extent to which the ASPIRE program is effective. If recruiting and retaining teachers in these more challenging schools or subjects remains a district goal, it may be more effective to provide baseline stipends to these teachers, in addition to making these teachers eligible for awards based on performance. To our knowledge, HISD does provide subject-based stipends, but does not provide stipends to teachers working in the highest need schools. Stipends like these might more directly facilitate the recruitment and retention of higher quality teachers for such positions and could even ameliorate teachers' potential frustrations that ASPIRE award determinations for them are based on comparisons to teachers in less challenging positions.

If recruiting and retaining teachers in these more challenging schools or subjects remains a district goal, it may be more effective to provide baseline stipends to these teachers, in addition to making these teachers eligible for awards based on performance.

Obtain a More Informed View of ASPIRE through the Collection of Qualitative Data

Evaluations of ASPIRE, and even ASPIRE itself, might benefit from the more systematic collection of qualitative data. The current survey on teachers' perspectives on ASPIRE is not mandatory and has extremely low response rates. With the possibility that the least or most satisfied teachers are most likely to respond, results from this survey cannot be presented or interpreted with any confidence in their generalizability. The mechanisms that make ASPIRE more or less effective might be illuminated by conducting a mandatory confidential survey of school administrators and teachers on their perceptions of the program, each Strand, and the specific steps they took to increase the likelihood of receiving an award. Information like this could increase understanding of why ASPIRE is less effective for some groups (e.g., teachers in higher need schools) and the specific ways in which award-receiving schools and teachers successfully increased student achievement. Sharing

The mechanisms that make ASPIRE more or less effective might be illuminated by conducting a mandatory confidential survey of school administrators and teachers on their perceptions of the program, each Strand, and the specific steps they took to increase the likelihood of receiving an award.

successful behavioral changes and school wide initiatives (those that resulted in awards) with teachers across the district could potentially increase the capacity of other schools and teachers to enable more gains in student achievement. HISD does have a website with information like this (http://portal.battelleforkids.org/Aspire/Recognize/2008_highachieve.html?sflang=en), but the extent to which HISD teachers make use of it is unclear. Data like this would have implications for school reform in general. Sharing results from the survey, particularly if those results are accompanied with statements on how the district is addressing concerns where possible, would potentially increase teacher investment in ASPIRE. Additionally, entrance and exit surveys might provide valuable information on whether teachers choose to work in HISD because of ASPIRE and the extent to which they choose to leave because of frustrations with ASIPRE.

Clarify Desired Effect of ASPIRE on Teacher Retention

Although HISD does not formally espouse any goals of ASPIRE related to retention, broader theoretical frameworks predict that an effective award program will encourage teachers who don't receive awards to remove themselves from the district or from the teaching profession entirely. This may not be in HISD's best interest if quality teachers are difficult to find, and, particularly, if award receipt becomes more selective. ASPIRE could be presented to teachers as a means of rewarding those who have reached a high level of proficiency and identifying those who need extra mentoring, professional development, etc. (Milanowski 2003), as opposed to encouraging certain teachers to leave. The competitive element of the award program would be retained, but the goals would be in better alignment with an educational context (i.e., recognition and remediation). Unless there are high quality teachers available to replace all lower quality teachers within HISD, taking steps to remediate teachers who did not receive an award would best serve the interests of HISD.

REFERENCES

Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. "Teachers and Student Achievement in the Chicago Public High Schools." *Journal of Labor Economics* 25(1):95-135.

Allison, Paul D. 1999. "Comparing Logit and Probit Coefficients Across Groups." *Sociological Methods and Research* 28(2):186-208.

Andrews, Hans A. 2011. "Supporting Quality Teachers with Recognition." *Australian Journal of Teacher Education* 36(12):59-70.

Aslam, Monazza, and Geeta Kingdon. 2011. "What Can Teachers do to Raise Pupil Achievement?" *Economics of Education Review* 30(3):559-74.

Ballou, Dale. 2001. "Pay for Performance in Public and Private Schools." *Economics of Education Review* 20(1):51-61.

Ballou, Dale, William Sanders, and Paul Wright. 2004. "Controlling for Student Background in Value-Added Assessment of Teachers." *Journal of Educational and Behavioral Statistics* 29(1):37-65.

Bayonas, Holli. 2010. *Guilford County Schools Mission Possible Program: Year 3 (2008–09) External Evaluation Report*. Greensboro, NC: The SERVE Center, University of North Carolina at Greensboro.

Belfield, Clive R., and John S. Heywood. 2008. "Performance Pay for Teachers: Determinants and Consequences." *Economics of Education Review* 27(3):243-52.

Benabou, Roland, and Jean Tirole. 2003. "Intrinsic and Extrinsic Motivation." *Review of Economic Studies* 70(3):489-520.

Bollen, Kenneth A., and Jennie E. Brand. 2010. "A General Panel Model with Random and Fixed Effects: A Structural Equations Approach." *Social Forces* 89(1):1-34.

Brick, Michael. 2012. "When 'Grading' is Degrading." in *The New York Times, The Opinion Pages*. New York, NY.

Buck, Stuart, and Jay P. Greene. 2011. "Blocked, Diluted, and Co-opted." *Education Next* 11(2.

Caillier, James. 2010. "Paying Teachers According to Student Achievement: Questions Regarding Pay-for-Performance Models in Public Education." *Clearing House: A Journal of Educational Strategies, Issues and Ideas* 83(2):58-61.

Carnoy, Martin, Illiana Brodziak, Andres Molina, and Miguel Socias. 2007. "The Limitations of Teacher Pay Incentive Programs Based on Inter-Cohort Comparisons: The Case of Chile's SNED." *Education Finance and Policy* 2(3):189-227.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2011. *The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood (Working Paper 17699)*. Cambridge, MA: National Bureau of Economic Research.

Clabaugh, Gary K. 2009. "Teacher Merit Pay: Is It a Good Idea?" *Educational Horizons* 88(1):16-20.

Clarke, Paul, Claire Crawford, Fiona Steele, and Anna Vignoles. 2010. *The Choice Between Fixed and Random Effects Models: Some Considerations for Educational Research (DoQSS Working Paper No. 10-10)*. London, UK: Department of Quantitative Social Science, Institute of Education, University of London.

Coleman, James S. 1990. *Equality and Achievement in Education*. Boulder, CO: Westview Press.

Cooper, Samuel T., and Elchanan Cohn. 1997. "Estimation of a Frontier Production Function for the South Carolina Educational Process." *Economics of Education Review* 16(3):313-27.

Currall, Steven C., Annette J. Towler, Timothy A. Judge, and Laura Kohn. 2005. "Pay Satisfaction and Organizational Outcomes." *Personnel Psychology* 58(3):613-40.

Darling-Hammond, Linda. 1996. "What Matters Most - A Competent Teacher for Every Child." *Phi Delta Kappan* 78(3):193-200.

Darling-Hammond, Linda, Audrey Amrein-Beardsley, Edward Haertel, and Jesse Rothstein. 2012. "Evaluating Teacher Evaluation." *Phi Delta Kappan* 93(6):8-15.
Dee, Thomas S., and Benjamin J. Keys. 2004. "Does Merit Pay Reward Good Teachers? Evidence from a Randomized Experiment." *Journal of Policy Analysis and Management* 23(3):471-88.
Department of Research and Accountability. 2011. *Houston Independent School District - District and School Profiles 2010-2011*. Houston, TX: Houston Independent School District.
Dillon, Sam. 2010. "Formula to Grade Teachers' Skill Gains Acceptance, and Critics." in *New York Times*.
Eberts, Randall, Kevin Hollenbeck, and Joe Stone. 2002. "Teacher Performance Incentives and Student Outcomes." *Journal of Human Resources* 37(4):913-27.
Farrell, Catherine, and Jonathan Morris. 2009. "Still Searching for the Evidence? Evidence-based Policy, Performance Pay and Teachers." *Journal of Industrial Relations* 51(1):75-94.
Ferguson, Ronald F. 2012. "Can Student Surveys Measure Teaching Quality?" *Phi Delta Kappan* 94(3):24-28.
Figlio, David N., and Lawrence W. Kenny. 2007. "Individual Teacher Incentives and Student Performance." *Journal of Public Economics* 91(5-6):901-14.
Fryer, Jr., Ronald G., Steven D. Levitt, John List, and Sally Sadoff. 2012. *Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment (Working Paper 18237)*. Cambridge, MA: National Bureau of Economic Research.
Fryer, Roland G. 2011. *Teacher Incentives and Student Achievement: Evidence from New York City Public Schools*. Cambridge, MA: National Bureau of Economic Research.
Glazerman, Steven, Hanley Chiang, Alison J. Wellington, Jill M. Constantine, and Daniel Player. 2011. *Impacts of Performance Pay Under the Teacher Incentive Fund: Study Design Report*. Washington, DC: Mathematica Policy Research.
Glazerman, Steven, and Allison Seifullah. 2010. *An Evaluation of the Teacher Advancement Program (TAP) in Chicago: Year Two Impact Report*. Washington, DC: Mathematica Policy Research, Inc.
Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2003. *Teacher Incentives*. Cambridge, MA: National Bureau of Economic Research.
Goldhaber, Dan, Michael DeArmond, and Scott DeBurgomaster. 2008. "Teacher Attitudes About Compensation Reform: Implications for Reform Implementation." *Industrial & Labor Relations Review* 64(3):441-63.
Goodman, Sarena F., and Lesley J. Turner. 2010. *Teacher Incentive Pay and Educational Outcomes: Evidence from the New York City Bonus Program*. New York, NY: Columbia University.
Goodman, Sarena, and Lesley Turner. 2011. "Does Whole-School Performance Pay Improve Student Learning? Evidence from the New York City Schools." *Education Next* 11(2):67-71.
Gratz, Donald B. 2009. "The Problem with Performance Pay." *Educational Leadership* 67(3):76-79.
Grodsky, Eric, and Catherine Riegle-Crumb. 2010. "Those Who Choose and Those Who Don't: Social Background and College Orientation." *The ANNALS of the American Academy of Political and Social Science* 627(1):14-35.
Guarino, Cassandra M., Abigail B. Brown, and Adam E. Wyse. 2011. "Can Districts Keep Good Teachers in the Schools that Need Them Most?" *Economics of Education Review* 30(5):962-79.
Hanshaw, Larry G. 2004. "Value-Related Issues in a Departmental Merit Pay Plan." *Professional Educator* 26(2):57-68.
Hanushek, Eric A. 2011. "The Economic Value of Higher Teacher Quality." *Economics of Education Review* 30(3):466-79.
Hanushek, Eric A., and Steven G. Rivkin. 2007. "Pay, Working Conditions, and Teacher Quality." *The Future of Children* 17(1):69-86.

Hanushek, Eric A., and Ludger Woessmann. 2011. "Overview of the Symposium on Performance Pay for Teachers." *Economics of Education Review* 30(3):391-93.
Harkness, Peter, and Mark Schier. 2011. "Performance Related Pay in Australian Universities: The Case of Swinburne University." *Australian Universities' Review* 53(2):50-58.
Hess, Frederick M. 2011. "Spend Money Like It Matters." *Educational Leadership* 68(4):51-54.
Hoetker, Glenn. 2007. "The Use of Logit and Probit Models in Strategic Management Research: Critical Issues." *Strategic Management Journal* 28):331-43.
Holt, Maurice. 2001. "Performance Pay for Teachers: The Standards Movement's Last Stand?" *Phi Delta Kappan* 83(4):312-17.
Hourigan, Ryan. 2011. "Race to the Top: Implications for Professional Development in Arts Education." *Arts Education Policy Review* 112(2):60-64.
Hulleman, Chris S., and Kenneth E. Barron. 2010. "Performance Pay and Teacher Motivation: Separating Myth from Reality." *Phi Delta Kappan* 91(8):27-31.
Imberman, Scott A., and Michael F. Lovenheim. 2012. *Incentive Strength and Teacher Productivity: Evidence from a Group-Based Teahcer Incentive Pay System (NBER Working Paper No. 18439)*. Cambridge, MA: National Bureau of Economic Research.
Ingersoll, Richard M. 2007. "Short on Power, Long on Responsibility." *Educational Leadership* 65(1):20-25.
Ingvarson, Lawrence, and Ken Rowe. 2008. "Conceptualising and Evaluating Teacher Quality: Substantive and Methodological Issues." *Australian Journal of Education* 52(1):5-35.
Jacob, Brian A. 2005. "Accountability, Incentives, and Behavior: The Impact of High Stakes Testing in the Chicago Public Schools." *Journal of Public Economics* 89):761-96.
Jacob, Brian A., and Lars Lefgren. 2008. "Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education." *Journal of Labor Economics* 26(1):101-36.
Jinnai, Yusuke. 2012. "The Impact of Teacher Performance Pay on Student Achievement: A Regression Discontinuity Approach." *Working Paper*.
Jones, Michael, and Michael Hartney. 2011. "Show Who The Money? Does Performance Pay Attract Higher Quality Teachers?" *Working Paper*.
Kane, Thomas J., and Douglas O. Staiger. 2002a. "The Promise and Pitfalls of Using Imprecise School Accountability Measures." *Journal of Economic Perspectives* 16(4):91-114.
—. 2002b. "Volatility in School Test Scores: Implications for Test-Based Accountability Systems." *Brookings Papers on Education Policy*):235-83.
—. 2008. *Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation (Working Paper 14607)*. Cambridge, MA: National Bureau of Economic Research.
Kelley, Carolyn. 1999. "The Motivational Impact of School-Based Performance Awards." *Journal of Personnel Evaluation in Education* 12(4):309-26.
Kingdon, Geeta Gandhi, and Francis Teal. 2007. "Does Performance Related Pay for Teachers Improve Student Performance? Some Evidence from India." *Economics of Education Review* 26(4):473-86.
Koedel, Cory, and Julian R. Betts. 2009. *Does Student Sorting Invalidate Value-Added Models of Teacher Effectiveness? An Extended Analysis of the Rothstein Critique*. Cambridge, MA: National Bureau of Economic Research.
Koppich, Julia E. 2005. "All Teachers Are Not the Same: A Multiple Approach to Teacher Compensation." *Education Next* 5(1):13-15.
Ladd, Helen F. 1999. "The Dallas School Accountability and Incentive Program: An Evaluation of Its Impacts on Student Outcomes." *Economics of Education Review* 18):1-16.
Lavy, Victor. 2007. "Using Performance-Based Pay to Improve the Quality of Teachers." *The Future of Children* 17(1):87-109.
Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables (Advanced Quantitative Techniques in the Social Sciences)*. Thousand Oaks, CA: Sage Publications, Inc.

Mahony, Pat, Ian Menter, and Ian Hextall. 2004. "The Emotional Impact of Performance-Related Pay on Teachers in England." *British Educational Research Journal* 30(3):435-56.
Marsh, Julie A., Matthew G. Springer, Daniel F. McCaffrey, Kun Yuan, Scott Epstein, Julia Koppich, Nidhi Kalra, Catherine DiMartino, and Art (Xiao) Peng. 2011. *A Big Apple for Educators: New York City's Experiment with Schoolwide Performance Bonuses - Final Evaluation Report*. Santa Monica, CA: The RAND Corporation.
McCollum, Sandra. 2001. "How Merit Pay Improves Education." *Educational Leadership* 58(5):21-24.
Milanowski, Anthony. 2003. "The Varieties of Knowledge and Skill-Based Pay Design: A Comparison of Seven New Pay Systems for K-12 Teachers." *Education Policy Analysis Archives* 11(4.
Morice, Linda C., and James E. Murray. 2003. "Compensation and Teacher Retention: A Success Story." *Educational Leadership* 60(8):40-43.
Mosier, Victoria, and Carla Stevens. 2011. *2009-2010 Final Inquiry Report and ASPIRE Award Payout Report: 2006-2007 through 2009-2010*. Houston, TX: Department of Research and Accountability, Houston Independent School District.
Murnane, Richard J., and David K. Cohen. 1986. "Merit Pay and the Evaluation Problem: Why Most Merit Pay Plans Fail and a Few Survive." *Harvard Educational Review* 56(1):1-17.
Neuman, W. Lawrence. 2009. *Social Research Methods: Qualitative and Quantitative Approaches, 7th Edition*. New York, NY: Pearson.
Noguera, Pedro. 2003. *City Schools and the American Dream: Reclaiming the Promise of Public Education*. New York, NY: Teachers College Press.
Papay, John P. 2011. "Different Tests, Different Answers: The Stability of Teacher Value-Added Estimates across Outcome Measures." *American Educational Research Journal* 48(1):163-93.
Perkins-Gough, Deborah. 2007. "Teachers Weigh In on Performance Pay." *Educational Leadership* 65(1):83-85.
Podgursky, Michael J., and Matthew G. Springer. 2007. "Teacher Performance Pay: A Review." *Journal of Policy Analysis and Management* 26(4):909-49.
Podgursky, Michael, and Matthew G. Springer. 2011. "Teacher Compensation Systems in the United States K-12 Public School System." *National Tax Journal* 64(1):165-92.
Propper, Carol, and Deborah Wilson. 2003. *The Use and Usefulness of Performance Measures in the Public Sector*. Bristol, UK: The University of Bristol.
Rabe-Hesketh, Sophia, and Anders Skrondal. 2012. *Multilevel and Longitudinal Modeling Using Stata, Third Edition*. College Station, TX: The Stata Press.
Ramirez, Al. 2001. "How Merit Pay Undermines Education." *Educational Leadership* 58(5):16-20.
—. 2011. "Merit Pay Misfires." *Educational Leadership* 68(4):55-58.
Ravitch, Diane. 2012. "My View: Rhee is Wrong and Misinformed." CNN: Schools of Thought. Retrieved December 30, 2012 (http://schoolsofthought.blogs.cnn.com/2012/08/09/my-view-rhee-is-wrong-and-misinformed/).
Riegle-Crumb, Catherine, and Eric Grodsky. 2010. "Racial-Ethnic Differences at the Intersection of Math Course-taking and Achievement." *Sociology of Education* 83(3):248-70.
Ritter, Gary W., and Nathan C. Jensen. 2010. "The Delicate Task of Developing an Attractive Merit Pay Plan for Teachers." *Phi Delta Kappan* 91(8):32-37.
Ritter, Gary W., Robert Maranto, and Stuart Buck. 2009. "Harnessing Private Incentives in Public Education." *Review of Public Personnel Administration* 29(3):249-69.
Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica* 73(2):417-58.
Rothstein, Jesse. 2012. *Teacher Quality Policy When Supply Matters (NBER Working Paper No. 18419)*. Cambridge, MA: National Bureau of Economic Research.
Rothstein, Richard. 2004. *Class and Schools: Using Social, Economic, and Educational Reform to Close the Black-White Achievement Gap*. New York, NY: Teachers College Press.

Ryssdal, Kai. 2011. "Freakonomics: Tackling the Problem of Cheating Teachers." Marketplace Podcasts: Freakonomics Radio, October 19, 2011. Retrieved November 16, 2012 (http://www.freakonomics.com/2011/10/19/those-cheating-teachers-a-new-freakonomics-marketplace-podcast/).

Sable, J., C. Plotts, and L. Mitchell. 2010. "Characteristics of the 100 Largest Public Elementary and Secondary School Districts in the United States: 2008-09 (NCES 2011-301)." Retrieved

Sandel, Michael J. 2012. "What Isn't For Sale?" *The Atlantic* April 2012):62-64.

Schacter, John, and Yeow Meng Thum. 2005. "TAPping into High Quality Teachers: Preliminary Results from the Teacher Advancement Program Comprehensive School Reform." *School Effectiveness and School Improvement* 16(3):327-53.

Scherer, Marge. 2001. "Improving the Quality of the Teaching Force: A Conversation with David C. Berliner." *Educational Leadership* 58(8):6-10.

Scott, Timothy. 2011. "A Nation at Risk to Win the Future: The State of Public Education in the U.S." *Journal for Critical Education Policy Studies* 9(1):267-316.

Smith, Stephen Samuel, and Roslyn Arlin Mickelson. 2000. "All That Glitters is Not Gold: School Reform in Charlotte-Mecklenburg." *Educational Evaluation and Policy Analysis* 22(2):101-27.

Solmon, Lewis C. 2005. "Recognizing Differences." *Education Next* 5(1):16-20.

Solmon, Lewis C., J. Todd White, Donna Cohen, and Debbie Woo. 2007. *The Effectiveness of the Teacher Advancement Program*. Santa Monica, CA: National Institute for Excellence in Teaching.

Springer, Matthew G., Dale Ballou, Laura Hamilton, Vi-Nhuan Le, J.R. Lockwood, Daniel F. McCaffrey, Matthew Pepper, and Brian M. Stecher. 2010a. *Teacher Pay for Performance: Experimental Evidence from the Project on Incentives In Teaching*. Nashville, TN: National Center on Performance Incentives at Vanderbilt University.

Springer, Matthew G., Dale Ballou, and Art Xiao Peng. 2008. *Impact of the Teacher Advancement Program on Student Test Score Gains: Findings from an Independent Appraisal*. Nashville, TN: National Center on Performance Incentives.

Springer, Matthew G., Jessica L. Lewis, Mark W. Ehlert, Michael J. Podgursky, Gary D. Crader, Lori L. Taylor, Timothy J. Gronberg, Dennis W. Jansen, Omar S. Lopez, and David A. Stuit. 2010b. *District Awards for Teacher Excellence (D.A.T.E.) Program: Final Evaluation Report*. Nashville, TN: National Center for Performance Incentives.

Springer, Matthew G., Jessica L. Lewis, Michael J. Podgursky, Mark W. Ehlert, Timothy J. Gronberg, Laura S. Hamilton, Dennis W. Jansen, Brian M. Stecher, Lori L. Taylor, Omar S. Lopez, and Art (Xiao) Peng. 2009a. *Texas Educator Excellence Grant (TEEG) Program: Year Three Evaluation Report*. Nashville, TN: National Center on Performance Incentives.

Springer, Matthew G., Jessica L. Lewis, Michael J. Podgursky, Mark W. Ehlert, Lori L. Taylor, Omar S. Lopez, and Art (Xiao) Peng. 2009b. *Governor's Educator Excellence Grant (GEEG) Program: Year Three Evaluation Report*. Nashville, TN: National Center on Performance Incentives.

Storey, Anne. 2000. "A Leap of Faith? Performance Pay for Teachers." *Journal of Education Policy* 15(5):509-23.

—. 2002. "Performance Management in Schools: Could the Balanced Scorecard Help?" *School Leadership & Management* 22(3):321-38.

Strunk, Katharine O., and Dara Zeehandelaar. 2011. "Differentiated Compensation: How California School Districts Use Economic Incentives to Target Teachers." *Journal of Education Finance* 36(3):268-93.

Texas Education Agency. 2011a. *2010-11 Academic Excellence Indicator System: Houston ISD*. Austin, TX: Texas Education Agency.

—. 2011b. "Texas Assessment of Knowledge and Skills - 2011 Test Administrations." Data File Formats. Retrieved January 5, 2013 (http://www.tea.state.tx.us/index3.aspx?id=3306&menu_id=793).

West, Kristine Lamm, and Elton Mykerezi. 2011. "Teachers' Unions and Compensation: The Impact of Collective Bargaining on Salary Schedules and Performance Pay Schemes." *Economics of Education Review* 30(1):99-108.

Williams, Richard. 2011. "Using the Margins Command to Estimate and Interpret Adjusted Predictions and Marginal Effects." in *Stata Conference*. Chicago, IL.

Winters, Marcus A. 2012. *Transforming Tenure: Using Value-Added Modeling to Identify Ineffective Teachers*. New York, NY: Manhattan Institute.

Wright, Robert E. 2003. "Difficulties in Marketing the Concept of Merit Pay for Primary and Secondary Teachers." *Research for Educational Reform* 8(3):38-45.

Zimmerman, Laurie S., Jennifer O'Brien, Victoria Mosier, Anna Olajay-Abarquez, Carla Stevens, and Don Hilber. 2011. *Research Educational Program Report: 2008-2009 ASPIRE Award Program Evaluation*. Houston, TX: Department of Research and Accountability, Houston Independent School District.

APPENDICES

APPENDIX 1: 2009-10 ASPIRE AWARDS PROGRAM AND ELIGIBILITY REQUIREMENTS

Following are the revised program and eligibility requirements for the 2009–2010 ASPIRE Awards.

GENERAL ELIGIBILITY REQUIREMENTS

To be eligible to participate in the 2009–2010 ASPIRE Awards, HISD Employees must meet all of the following general eligibility requirements.

1. Employees must be supervised and evaluated by the principal of the campus where they are serving students. (This does not apply to Category J: Principals)
2. Employees must be employed in a campus-assigned position as of the fall snapshot date.
3. Employees must be continuously employed in an eligible position through the last day of school.
4. Employees must complete the instructional-linkage and assignment-verification process, or have this completed by their principal, through the ASPIRE portal by the submission deadline as published annually. It is recommended that Employees review instructional-linkage and assignment-verification information on the ASPIRE portal for accuracy.
5. Employees may "opt out" of the ASPIRE Award Program during the linkage and verification process. If an educator does not make a selection, the educator will be included for consideration for an ASPIRE Award.
6. Employees eligible under other incentive plans are not eligible for ASPIRE Awards (e.g., Food Services Employees).
7. Hourly Employees in any capacity, including substitute/associate teachers, are not eligible to participate in the ASPIRE Awards. Employees holding an hourly or substitute position must be converted to a non-hourly position by the fall snapshot date in order to be eligible.
8. Employees who take leave of absence during the eligibility period (e.g., temporary disability, but not family medical leave) are not eligible to participate in the ASPIRE Awards.
9. Employees must be in attendance 90 percent of the 175 instructional days identified as the "instructional school year." This means that Employees cannot be absent for more than 10 percent of their scheduled hours to work during the instructional year; and first-year Employees must have been hired by September 17, 2009. The following types of leave will be held harmless (not count as days absent): funeral leave, military leave, family medical leave (must be authorized through Human Resources), assault leave, jury duty, religious holidays, compensatory time, and off-campus duty.

POSITION ELIGIBILITY REQUIREMENTS AND CATEGORIZATION

Different positions within HISD qualify for various aspects of the ASPIRE Award Program. Following are definitions for position categories and eligibility requirements that will be used to categorize Employees for award purposes.

Instructional Position Categories

Employees who qualify as instructional must be certified teaching staff and will fall into either core foundation or elective/ancillary instructional positions as defined below.

Core Foundation Teaching Positions

For Employees to qualify as core foundation instructional staff, Employees must be assigned to a campus, plan lessons, provide direct instruction to students, and be responsible for providing content grades, not conduct or participation grades.

ASPIRE Core Foundation Courses

The ASPIRE Core Foundation Courses include those courses identified by the Texas Education Agency under the Core Foundation areas of English Language Arts/Reading, Mathematics, Science, and Social Studies at the elementary and middle school levels and those Core Foundation courses required for graduation credit in the 4 x 4 Recommended or Distinguished High School Diploma programs and/or those courses that contribute directly to data collected and interpreted as part of the growth measure. Fifty percent of the teaching assignment must be in ASPIRE Core Foundation courses to be considered as a core foundation teacher for the purposes of award.

A. Core Foundation Teachers, Grades 3–6, Self-Contained

To be considered in this category, Employees must qualify as core foundation teachers and teach the majority of the same students in grades 3–6 in at least four out of the five core foundation subject areas. For third grade only, Employees must teach reading, math, and language arts to the majority of the same students to be considered "self-contained." A teacher-level value-added report should be produced for these Employees. For small class sizes, a special analysis may be performed (see Award Model Diagram for further details and definitions).

B. Core Foundation Teachers, Grades 3–8, Departmentalized

To be considered in this category, Employees must qualify as core foundation teachers and teach one to three core foundation subjects to different classes of students in grades 3–8. A teacher-level value-added report should be produced for these Employees. For small class sizes, a special analysis may be performed (see Award Model Diagram for further details and definitions).

C. Core Foundation Teachers, Grades 9–12

To be considered in this category, Employees must qualify as core foundation teachers and teach grades 9–12 core foundation courses the majority of the school day. For a complete list of these courses, please review the 2009 Master Course List with ASPIRE core foundation subjects.

D. Core Foundation Teachers, Pre-Kindergarten through Grade 2

To be considered in this category, Employees must qualify as core foundation instructional staff and teach core foundation subjects to students in Pre-Kindergarten through grade 2 the majority of the school day.

E. Special Education Core Foundation Teachers–No Value-Added Report

To be considered in this category, Employees must qualify as core foundation instructional staff and teach core foundation subjects to Special Education students in grades 3–8 where a value-added report cannot be generated, or teach fewer than seven TAKS or TAKS-accommodated Special Education students in grades 9–12. All other Special Education teachers will be considered under their respective core foundation teacher category (above).

Elective/Ancillary Instructional Positions

F. Elective/Ancillary Teachers

To be considered elective/ancillary teachers, Employees must teach elective or ancillary courses for fifty percent or more of their teaching assignment and do not meet the definition of core foundation teachers (above) in grades PK–12. Courses defined by TEA as "core enrichment" (art, music, dance, theater, PE) are considered an elective course for ASPIRE.

Other Position Categories

In addition to recognizing instructional staff, the ASPIRE Awards also acknowledge the contributions of Employees who contribute to student growth in other ways throughout the school year. Following are the categorizations to recognize these Employees.

G. Instructional Support Staff

Instructional support-staff members are degreed, certified, or licensed professionals assigned to a campus and provide direct support to the instruction of students. If the instructional support-staff member is assigned to multiple campuses, the percentage of assignment to a single campus cannot be less than 40 percent.

For example: counselor, librarian, nurse, speech therapist, speech therapist assistant, evaluation specialist, instructional coordinator, content-area specialist, school-improvement facilitator, API, social worker, literacy coach, Magnet coordinator, or Title I coordinator.

H. Teaching Assistants

Teaching assistants are staff members who have a job classification of teaching assistant and provide direct classroom instructional support to instructional staff.

I. Operational Support Staff

Operational support-staff members are campus-based Employees who do not meet the requirements for instructional staff, instructional support staff, or teaching assistants.

For example: school secretary, data entry clerk, teacher aide, clerk, attendance specialist, business manager, SIMS clerk, computer network specialist, registrars, and CET.

Campus Leadership Categories

The ASPIRE Award Program recognizes school administratorship for their contributions to student progress and achievement based on campus and departmental performance. Certification for these positions is required to be considered for these categories. The following describe the award category eligibility for leadership positions.

J. Principals

To be considered in this category, Employees must meet all eligibility requirements and be the “principal of record” according to HR and PeopleSoft.

K. Assistant Principals/Deans of Instruction/Deans of Students

To be considered in this category, Employees must meet all eligibility requirements, and be coded as an assistant principal, dean of instruction, or dean of students according to HR and PeopleSoft.

Additional Position Eligibility Requirements

1. For an educator who voluntarily transfers from one ASPIRE Award-eligible position to another ASPIRE Award-eligible position during the eligibility period, the award will be determined on the basis of the ASPIRE Award-eligible position the educator held the greatest percentage of the school year (based on the 187-day duty schedule). For example: On September 5 (prior to the fall snapshot), an educator teaches third-grade math (Category B: a departmentalized, core foundation teacher). On February 5, the educator transfers to a content specialist position on the same campus (Category G: an instructional support position). Both assignments are ASPIRE

Award-eligible. However, the award model and eligibility requirements differ. In this case, the greater percentage of the "school year" was spent as a third-grade, departmentalized, core foundation teacher. Therefore, the award amount would be determined on the basis of the job of a third grade, departmentalized, core foundation teacher.

2. For an educator who transfers from an ASPIRE Award-eligible position to a non-eligible position during the eligibility period, he or she will not be eligible for an award (see General Eligibility Requirements: Rules 2 and 3).
3. The ASPIRE Award for Employees who function in multiple categories (above) will be determined on the basis of the job in which they function the majority of their work day.
4. Employees must have credentials for the position in which they function to be eligible under that category. For example: A teacher teaching ninth-grade math must be certified or on permit to teach ninth-grade math to be eligible as a core foundation 9–12 teacher.
5. For Employees who meet the criteria of a core foundation teacher (including Additional Position Eligibility Requirement 3) and for whom a value-added report is produced, the position categorization will be where direct growth can be measured. For example: If a teacher teaches second- and third-grade reading, and a value-added report is obtained for third grade based on the direct measure of student growth is obtained for third grade, the teacher would be eligible under Category B, as a core foundation 3–8, departmentalized teacher. If an educator teaches music the majority of the day, and one class of reading (for which he or she may receive a value-added report), then the educator will be eligible under Category F: Elective/Ancillary Teacher.
6. The production of a value-added report does not necessarily categorize an educator as a core foundation teacher for the purposes of determining ASPIRE Award-position eligibility. For example: If a value-added report is produced to measure the growth of students by a tutor for diagnostic and instructional improvement, the tutor is not eligible as a core foundation teacher unless all the criteria for a core foundation teacher position are met (see the Position Eligibility Requirements and Categorization section).

ASPIRE AWARD CALCULATION AND PAYOUT RULES

The ASPIRE Awards for teachers will be calculated on the basis of the HISD board-approved model. Certain situations require the adoption of the following award-calculation rules to apply the award model appropriately.

1. Employees who work less than full time must work at least 40 percent of the school time (equivalent to two days per week) at the same campus to be eligible to receive a prorated ASPIRE Award. The prorated ASPIRE Award will be based on the full-time equivalent (FTE) of their eligible position, the portion of time spent in the eligible position, and the ASPIRE Award level. For example: A half-time educator or 0.5 FTE who spends all of his or her time at a single campus will be eligible to receive 50 percent of the award. This same educator who works 50 percent of his or her time at two campuses (0.25 FTE at each campus) will not be eligible.
2. Awards for Employees whose job record/position is assigned to non-campus departments or to a school improvement officer for time reporting, but who are assigned to and work on specific campuses a minimum of 40 percent of the time, and report directly to the principal (the principal is responsible for supervising and evaluating the individual educator) will be calculated and prorated on the basis of the percentage of campus assignments. Examples include evaluation specialists, content specialists, speech therapists, and various Special

Education positions. For example: A department-assigned, campus-based educator works 50 percent of his or her time at campus A, 25 percent at campus B, and 25 percent at campus C. If the educator is eligible for an ASPIRE Award based on campus data, then the educator would receive 50 percent of the eligible payout at campus A and would not receive an award for campus B or C.

3. The ASPIRE Award for Employees assigned to multilevel campuses (e.g., T. H. Rogers) will be determined by an average of both campus-award amounts for Strands I and III.
4. Employees must be in good standing at the time of payment. Therefore, an educator under investigation or reassigned pending investigation is not eligible for an ASPIRE Award payment until he or she is cleared of any allegation. If the investigation is concluded with a confirmation of inappropriate educator behavior, the educator is not eligible to receive an ASPIRE Award payment. Additionally, Employees who retire in lieu of termination or resign in lieu of termination are not eligible to receive an ASPIRE Award payment.
5. If an educator meets all of the eligibility requirements for an award and then resigns or retires from the district prior to the payout of the awards, the educator is still eligible for the award. It is incumbent upon the educator to provide the district with correct forwarding information so that the award payment can be processed.
6. For Principals Only: The campus must also be in good standing. If the campus had an approved waiver to the district-testing procedures, and if any testing improprieties are reported and confirmed or otherwise substantiated at the campus, the principal will be ineligible to receive an ASPIRE Award payment.

APPENDIX 2, PART 1 OF 2: 2009-10 ASPIRE AWARDS MODEL DIAGRAM FOR EDUCATORS WHO ARE NOT SCHOOL ADMINISTRATORS

		Core Foundation Teachers, Grades 3-6, Self Contained	Core Foundation Teachers, Grades 3-8, Departmen-talized	Core Foundation Teachers, Grades 9-12	Core Foundation Teachers, Pre-K-Grade 2	Special Education Core Foundation Teachers	Elective/ Ancillary Teachers	Instructional Support Staff	Teaching Assistants	Operational Support Staff
Strand I	Campus Progress Award	Campus overall value-added score across all grades and subjects calculated by EVAAS. Campuses are then split into elementary, middle, and high school levels and compared separately. Only campuses demonstrating positive gain and in Q1/Q2 of their comparison group are awarded. Campuses are rank ordered into quartiles.								
		Max $1500	Max $1500	Max $1500	Max $1500	Max $1500	Max $1500	Max $1000	Max $1000	Max $750
Strand II	Teacher Progress Award	Value-added Cumulative Gain Index by teacher by subject	Value-added Cumulative Gain Index by teacher by subject[a]	Value-added campus Cumulative Gain Index by subject and grade level	Value-added Campus Gain Index in 3rd-grade reading and math[b]	Value-added Campus Gain Index in each applicable subject	Not eligible for Strand II			
		Rank in quartiles comparing self-contained teachers by grade and subject	Rank in quartiles comparing departmen-talized teachers by academic level and subject[a]	Rank high schools in quartiles comparing grade levels and subjects	Rank elementary campuses into quartiles comparing 3rd-grade reading and math	Rank campus scores into quartiles comparing subjects				
		Max $7000	Max $7000	Max $7000	Max $3500	Max $3500				
		a-Teachers of only one grade level in grades 6, 7, or 8 are rank ordered with other teachers of the same grade level and subject.								
		b-Early childhood centers are paired with their feeder campus.								

Appendix 2, Part 2 of 2: 2009-10 ASPIRE Awards Model Diagram for Educators Who are Not School Administrators

<table>
<tr><th></th><th></th><th>Core Foundation Teachers, Grades 3-6, Self Contained</th><th>Core Foundation Teachers, Grades 3-8, Departmen-talized</th><th>Core Foundation Teachers, Grades 9-12</th><th>Core Foundation Teachers, Pre-K-Grade 2</th><th>Special Education Core Foundation Teachers</th><th>Elective/ Ancillary Teachers</th><th>Instructional Support Staff</th><th>Teaching Assistants</th><th>Operational Support Staff</th></tr>
<tr><td rowspan="11">Strand III</td><td rowspan="2">Part A: Campus Elementary and Middle School Improvement Award</td><td colspan="9">Eligible elementary and middle school staff are awarded where their campus earned a: TEA Comparable Improvement ranking in Quartile 1 or 2 of the 40 TEA-designated comparison schools for reading and/or math. TEA rank order of Campus Progress on TAKS Scale Score</td></tr>
<tr><td>Max $1000</td><td>Max $1000</td><td>N/A</td><td>Max $1000</td><td>Max $1000</td><td>Max $1000</td><td>Max $500</td><td>N/A</td><td>N/A</td></tr>
<tr><td rowspan="2">Part A: High School Campus College Credit Participation/ Performance Award</td><td colspan="9">Eligible high school staff are awarded where: 40% or more of students in grades 10–12 enrolled in at least one AP, IB, or dual credit course or improvement in the percentage of students enrolled in at least one AP, IB, or dual credit course campuswide ranked in Quartile 1 or 2. Only positive change will be rewarded. Rank order campus progress if 40% standard is not met.</td></tr>
<tr><td>N/A</td><td>N/A</td><td>Max $500</td><td>N/A</td><td>Max $500</td><td>Max $500</td><td>Max $250</td><td>N/A</td><td>N/A</td></tr>
<tr><td rowspan="3">Part B: Campus Achievement Award</td><td colspan="9">Eligible staff are awarded where their campus earned a: TEA rating of Exemplary or Recognized.</td></tr>
<tr><td>Exemplary $400</td><td>Exemplary $400</td><td>Exemplary $400</td><td>Exemplary $400</td><td>Exemplary $400</td><td>Exemplary $400</td><td>Exemplary $200</td><td>Exemplary $100</td><td>N/A</td></tr>
<tr><td>Recognized $200</td><td>Recognized $200</td><td>Recognized $200</td><td>Recognized $200</td><td>Recognized $200</td><td>Recognized $200</td><td>Recognized $100</td><td>Recognized $50</td><td>N/A</td></tr>
<tr><td rowspan="3">Part C: Campus Writing Achievement Award</td><td colspan="9">Eligible staff are awarded where: 70% or more of students campuswide met the College Readiness Standard on TAKS Writing/ELA, or improvement in the percentage of students meeting the College Readiness Standard on TAKS Writing campuswide ranked in Quartile 1 or 2. Only positive change will be awarded. College Readiness Standard is TAKS Writing/ELA scale score of 2200 or greater AND writing composition score of 3 or better on either the TAKS Writing grade 4 or 7 or the 11th grade TAKS ELA. Rank Order Campus Progress if 70% Standard Is Not Met.</td></tr>
<tr><td>4th/7th Writing $400</td><td>4th/7th Writing $400</td><td>High-School ELA $400</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Instrctnl Staff $200</td><td>Instrctnl Staff $200</td><td>Instrctnl Staff $200</td><td>Instrctnl Staff $200</td><td>Instrctnl Staff $200</td><td>Instrctnl Staff $200</td><td>N/A</td><td>N/A</td><td>N/A</td></tr>
<tr><td colspan="2">Maximum Award</td><td>$10,300</td><td>$10,300</td><td>$10,300</td><td>$6,600</td><td>$6,600</td><td>$3,100</td><td>$1,700</td><td>$1,100</td><td>$750</td></tr>
</table>

APPENDIX 3: 2009-2010 ASPIRE AWARDS FOR EDUCATORS WHO ARE NOT SCHOOL ADMINISTRATORS

Category		Strand I: Campus Progress Award	Strand II: Teacher Progress Award	Strand III: Campus Improvement and Achievement Award				Maximum ASPIRE Award		Attendance Bonus
				A - ES/MS Campus Improvement Award - HS College Credit Part./Perf. Award	B - Campus Achieve-ment Award	C - Campus Writing Achievement Award				
						4th & 7th Grade Writing & HS ELA	Other Instrctnl Staff	4th & 7th Grade Writing & HS ELA	Other Instrctnl Staff	
Core Foundation Teachers	Gr. 3-6, Self Contained	$1,500	$7,000	$1,000	$400	$400	$200	$10,300	$10,100	Perfect attendance: Add 10% of total ASPIRE Award Bonus. Missed less than two days: Add 5% of total ASPIRE Award Bonus.
	Gr. 3-8, Dprtmntlzd	$1,500	$7,000	$1,000	$400	$400	$200	$10,300	$10,100	
	Gr. 9-12	$1,500	$7,000	$1,000	$400	$400	$200	$10,300	$10,100	
	Pre-K-Gr. 2	$1,500	$3,500	$1,000	$400	N/A	$200	$6,600		
	Special Education[a]	$1,500	$3,500	$1,000	$400	N/A	$200	$6,600		
Elctve/Ancllry Teachers		$1,500	N/A	$1,000	$400	N/A	$200	$3,100		
Instructional Support Staff		$1,000	N/A	$500	$200	N/A	N/A	$1,700		
Teaching Assistants		$1,000	N/A	N/A	$100	N/A	N/A	$1,100		
Operational Support Staff		$750	N/A	N/A	N/A	N/A	N/A	$750		

Note: Award amounts shown are maximum amounts for each strand and category. Based on 2009–2010 school year data.

a-For Special Education teachers in grades 3–8 where a value-added report cannot be generated or grades 9–12 with fewer than 7 TAKS or TAKS-accommodated students. All other Special Education teachers will be considered under the respective core instructional category according to the position eligibility rules.

Appendix 4: Acronyms

Acronym	*Meaning*	*Explanation*
AP	Advanced Placement	Rigorous high school courses and assessments developed by the College Board
ASPIRE	Accelerating Student Progress: Increasing Results and Expectations	HISD's educator award program from 2006-2007 through current school year
BA	Bachelor's Degree	Academic degree typically awarded after four years of postsecondary schooling at the undergraduate level
CGI	Cumulative gain index	Score calculated by EVAAS that compares achievement gains of teachers' students to that of all HISD students
ESL	English as a Second Language	Educational program focused on teaching and assimilating students with limited English proficiency
EVAAS	Education Value-Added Assessment System	Value-added metrics from this company are used for ASPIRE award determinations
GT	Gifted and Talented	Educational program focused on enriching the educational experiences of students with high ability
HERC	Houston Education Research Consortium	A research team housed within Rice University's Kinder Institute for Urban Research, and directed by Ruth López Turley, an Associate Professor in Rice University's Department of Sociology
HISD	Houston Independent School District	The district in which ASPIRE is implemented
IB	International Baccalaureate	Rigorous high school courses and assessments developed by the international organization IB
LEP	Limited English Proficient	Official school designation that qualifies student for extra services from their school
PEIMS	Public Education Information Management Systems	TEA requires all Texas school districts to maintain these data files with information on students' sociodemographic and school designation characteristics
TAKS	Texas Assessment of Knowledge and Skills	Standardized test administered to grades 9-11 from 2003 through current school year, and administered to grades 3-8 from 2003 through 2011
TEA	Texas Education Agency	State agency that collects sociodemographic and achievement data on Texas students to construct accountability ratings of schools in compliance with NCLB

APPENDIX 5: DATA

This report uses data collected and managed by HISD's Department of Research and Accountability, and Information Technology Human Capital Management Department (also referred to as 'PeopleSoft'). HERC first received data files with TAKS and Stanford test scores, and sociodemographic information on all of the students in HISD [Public Education Information Management Systems (PEIMS)]. This is data that TEA requires all Texas school districts to maintain (http://www.tea.state.tx.us/peims/). In subsequent stages, HERC received around 100 separate files of data collected by HISD. These data files were typically separated by school year and topic, and contained approximately 3,000 variables total. Depending on the topic, HERC received data files from as early as the 2006-07 school year to as recent as the 2011-12 school year. Employee level data files provided information on employees' demographic and employment characteristics (HISD refers to these as the PeopleSoft data files – PeopleSoft is a human resources software), type of teacher certification, experiences relative to the ASPIRE award program, retention and termination status, attendance rates, receipt of new hire recruitment stipends, student achievement gain indicators as calculated by EVAAS, responses to a survey administered after the completion of online training on the ASPIRE program, and completion of professional development courses. School level data files provided information on each school's qualities relative to their magnet status and EVAAS's indicators of campus level student achievement gains. We constructed most of the school level measures used in this evaluation by aggregating the data describing students' sociodemographic qualities. Lastly, we received data files that linked students to each of their teachers for grades 1-11 for the 2010-11 school year, and grades 3-8 for the 2006-07 through 2008-09 school years. HISD has no formal codebooks, so in addition to cleaning the data in these files and preparing them for merging, we relied on previous HISD reports, information from HISD staff, and internet searches to interpret the meaning of variable names and variable value codes (which were typically inconsistent across data files and school years). Most of these data files had duplicate entries for some subset of employees (e.g., retention status at various time points in the school year, employees who worked at multiple schools). We aggregated these within each data file so that the data was accurately represented by a single entry for each employee.

We do not use the data on teacher certification in this evaluation, because HISD staff advised us that uncertified teachers number too few in HISD to be of policy interest, and because the data file does not capture other differences of potential interest [i.e., teachers who initially receive alternative certification are reclassified (and unidentifiable) once they complete their probationary period]. We do not use the termination data file because the codes describing the reason for termination were not unique (HISD provided a document in which the termination code "CON," in just one example, is defined as "Misconduct," "FML/WC Concurrent," and "Consultant"). The professional development data file excluded certain sources of professional development, and so was not representative. We do not use data from employee surveys related to ASPIRE, because they were not mandatory and had very low response rates (for this reason, we did not even request data from HISD's survey seeking employee feedback on ASPIRE). HISD staff does not have confidence that the data in the 2006-07 teacher-student linkage file are accurate, because of glitches with the software and linking processes that were addressed after this first year.

INDEPENDENT VARIABLES

We chose the variables we would use through an iterative process of theoretical considerations and extensive exploratory analyses. Some basic variables (e.g., employee race, sex, job title, years of experience) were available in multiple data files. In addition to triangulating data across data files, we addressed missing values on independent variables through multiple imputation by the MICE system of chained equations.

ASPIRE Awards, 2009-10

ASPIRE Category. HISD uses 11 ASPIRE categories to determine the Strands through which employees are eligible to receive awards, and the method by which award determinations are made within each Strand. These categories include: a) Self-Contained Core Foundation Employees, Grades 3-6, b) Departmentalized Core Foundation Employees, Grades 3-8, c) Core Foundation Employees, Grades 9-12, d) Core Foundation Employees, Grades PK-2, e) Special Education Core Foundation Employees, Grades 3-12, f) Elective/Ancillary Employees, g) Instructional Support Staff, h) Teaching Assistant, i) Operational Support Staff, j) Principals, and k) Assistant Principals (more details available in Appendix 1). We use these categories as controls to ensure the comparison of like employees, but also to detect differentiation in the estimated effect of receiving a Strand II award depending on how the award determination was made.

Eligibility for Awards. The eligibility indicator in the award data file was dichotomous and grouped all ineligible employees together, regardless of the reason for ineligibility. We used the data file's indicators of eligibility, having opted out, and ASPIRE category, as well as missing values on award amounts and conversations with HISD staff to create more fine-toothed distinctions in eligibility. We distinguish employees based on whether they were 1) ineligible for the award program because of their job category or campus of employment (i.e., their ASPIRE category was 'X,' 'Z,' or missing), 2) ineligible for the award program because of attendance (they had a valid ASPIRE category but were missing on all award amount indicators), 3) eligible for the award program but opted out, 4) eligible for the award program (they had a valid ASPIRE category and values of zero or greater on the total award amount indicator). Among those who were eligible for the award program in general, we determined which Strands each employee was eligible for based on whether the award amount for each Strand was missing (ineligible), or zero or higher (eligible).

Receiving Awards. We constructed dichotomous indicators of whether each employee received any ASPIRE award, a Strand I award, a Strand II award, and/or a Strand III award by assigning a 1 for all cases in which the employee's award amount for that Strand was greater than zero.

Total Award Amount. We used the measure describing the amount each employee was awarded for their total 2009-10 award, but recoded missing values to zero for ineligible employees we wished to include in regression analyses (i.e., ineligible because of attendance or retention). In linear regression analyses, we use a version of this variable divided by 3 in order to increase the number of significant digits in our results. To determine whether employees' interpret the amount of their award by comparing it to other salient amounts of money, we also constructed measures of employees' award amounts as a proportion of their annual salary, and a proportion of the mean amount of the awards received at their 2010-11 school (employees received awards for 2009-10 in January 2011).

Hard-to-Staff Subjects, 2009-10

Although HISD does not have a formal list of hard-to-staff subjects, convention and conversations with HISD staff led us to classify secondary level math, secondary level science, bilingual/ESL, and special education as hard-to-staff subjects. For these analyses, teachers not in hard-to-staff positions form the reference group. Teachers not in hard-to-staff positions include both self-contained teachers (e.g., elementary teachers that teach multiple subjects), and teachers who teach subjects like English, history, and social studies. We constructed these indicators using text from character variables stating each employee's job title in the PeopleSoft data files. For instance, teachers whose job titles included "Science," "Biology," "Chemistry," or "Physics" were coded as science teachers. This method is imperfect because teachers may teach across multiple subjects, but by relying on their main HISD categorization (i.e., their job title), we may gain insight into average differences in the effects of ASPIRE for teachers of hard-to-staff subjects.

High Need Schools, 2010-11

HISD employs several measures of high need schools, depending on the focus of the project, report, or the interests of the audience. We chose to not use average levels of achievement as a measure of high need, because it is difficult to discern whether low levels of achievement are attributable to the social backgrounds of the students at the school, or the qualities of the employees. We also chose to not use a measure of the mean years of experience among each school's employees, because schools with more experienced employees actually had poorer attendance rates. Student achievement, employee attendance rates, and employee retention rates are the lowest in HISD's schools with higher proportions of students in poverty, black students, students in bilingual or ESL programs, and students in special education. To align with the focus of this evaluation, we define high need schools using these school characteristics. For each school year, we classified schools whose proportions of poor students placed them in the top two quartiles within the district, to be high need insofar as this characteristic. Similarly, we classified schools whose proportion of black students placed them in the top two quartiles within the district, to be high need in terms of that characteristic. We did the same for the other two characteristics, and finally, summed the number of characteristics that qualified each school to be a high need school. For instance, schools that are ranked 0 were not in the top 2 quartiles for any of these four characteristics, whereas schools that are ranked 3 were in the top 2 quartiles for three of these characteristics. To facilitate the presentation of results, we compare employees in schools with 0-2 characteristics ('lower need'), to employees in schools with 3-4 characteristics that qualify them to be high need ('higher need'). This indicator does not produce multicollinearity, even when included in regression analyses with the original continuous measures of these school level characteristics.

Employee Characteristics, 2009-10

Sex. We distinguish male employees from female employees using indicators from the PeopleSoft data files.

Race. We constructed a five category measure of race using indicators from the PeopleSoft data files. HISD employees were first asked if they are Hispanic, and were then allowed to choose as many races as they liked from these options: white, black, Asian, Pacific Islander, American Indian. We attempted to combine these responses in a meaningful way, trying to both separate the groups with the largest representations, and to align with Census classifications. We code employees who only selected white and indicated they were not Hispanic as 'White, non-Hispanic', employees who only selected black and indicated they were not Hispanic as 'Black, non-Hispanic', employees who only selected white and indicated they were Hispanic as 'Hispanic, white,' employees who only selected Asian and indicated they were not Hispanic as 'Asian', employees who selected some different combination from the ones already described as 'Other.' Employees who chose Hispanic and white comprised the vast majority of the group of employees who chose Hispanic at all, and so employees who chose Hispanic and black seemed to be a better fit with the multi-racial employees encompassed in the 'Other' category.

Educational Attainment. We constructed three-category ordinal variables describing each employee's educational attainment from character variables in the PeopleSoft data files. We coded employees with values of 'Less than HS,' 'HS Graduate or Equivalent,' 'Some College,' 'Technical School,' 'Associate Degree,' and '2-Year College Degree' as having completed less than a Bachelor's degree (BA). We coded employees with values of 'Bachelor's Level Degree' or 'Some Graduate School' as having completed a BA. We coded employees with values of 'Master's Level Degree,' ‘Doctorate (Academic),’ ‘Doctorate (Professional),’ or ‘Post-Doctorate’ as having completed a Master’s degree (MA) or higher. We set employees with values of ‘Not Indicated’ or ‘NCLB – Employee Assistant’ to missing (i.e., to be imputed). As a first step in addressing missing values, we triangulated data on each employee’s educational attainment from other school years. In analyses focused on teachers, we compare teachers with BAs or lower to teachers with an MA or higher, because so few teachers have less than a BA.

New Hire Recruitment Stipend. We constructed a dichotomous measure for teachers of whether they received a new hire recruitment stipend. The data file included employees who had received one or more of these year-one and year-two recruitment stipends: ‘Recruitment Incentive1 Y1 342A,’ ‘Recruitment Incentive2 Y1 342B,’ ‘Recruitment Incentive3 Y1 342C,’ ‘Recruitment Incentive4 Y2 BIL,’ ‘Recruitment Incentive5 Y2 Crit,’ ‘Recruitment Incentive6 Y2 Oth.’ We primarily use this variable as a control to isolate the effect of receiving an award for employees whose outcomes may also be affected by receiving a new hire stipend. Our analyses related to recruitment stipends are limited, because we only received data on new hire recruitment stipends (first and second years of employment), which potentially excludes stipends non-new-hires receive because they teach a hard-to-staff subject, for example.

Total Years of Experience. We used indicators from the PeopleSoft data files describing employees’ years of experience within and outside of HISD to construct ordinal measures of whether they had a) 3 or less, b) 4-10, c) 11-25, or d) more than 25 total years of experience. We interpret this variable as years of experience in their current position. We chose to use an ordinal measure rather than the continuous measure because exploratory analyses demonstrated that years of experience did not have a linear relationship with our outcomes. In essence, a continuous measure would have eclipsed interesting relationships (e.g., teachers with the most experience have the poorest outcomes in some cases). As a first step in addressing missing values, we triangulated data on each employee’s years of experience using indicators from the retention data files. Although a squared measure of employees’ total years of experience in current position slightly improved the fit of our models in early exploratory analyses [see Dee and Keys (2004) for more support for this decision], we did not use this measure because it was highly correlated with the ordinal version of years of experience. Indicators of employee’s salaries were also available in the PeopleSoft data files, but were highly correlated with years of experience (approximately $\alpha=0.95$). As documented in our literature review, employees’ salaries are not highly differentiated, so we chose to use years of experience as a measure of differences across employees rather than salary. We also constructed dichotomous measures designating teachers with zero years of experience as first-year teachers.

Cumulative Gain Index (CGI), 2008-09. We use EVAAS’s measure of the value teachers added in 2008-09 as a strong control in models predicting math and reading/ELA teachers’ mean student test score gains. EVAAS converts students’ test score data to a Normal Curve Equivalent (NCE) scale (2006 TAKS state data acts as the anchor for this scale) (see http://www.sas.com/resources/asset/SAS-EVAAS-Statistical-Models.pdf). Student scores are aggregated to the teacher level into an average NCE score for each teacher for each subject in each year. Each teacher’s 2008-09 gain score is calculated by subtracting their NCE score for 2007-08 from their NCE score from 2008-09. The gain index is then calculated as the teacher’s NCE gain minus the district average gain for that grade and subject divided by the teacher’s standard error. Teachers’ CGI are a calculation across grades for a given subject. The CGIs of Core Foundation Teachers, Grades 3-6, Self-Contained and Core Foundation Teachers, Grades 3-8, Departmentalized are based on students they actually teach. The CGIs of Core Foundation Teachers, Grades 9-12 are based on students on their campus in their subject(s). Because we did not receive linkage data files for high school teachers in 2008-09, we did not have specific information on these teachers’ grade levels and subjects. We were able to use our constructed measures of eligibility for various ASPIRE awards (there are more specific measures of the awards that specify grade level and subject) to determine which gain indices from the data file of campus value-added measures to aggregate as a mean for each high school teacher. We use an average of the math and reading cumulative gain indices for Core Foundation Teachers, Pre-K-Grade 2 (who are only included in models predicting mean test scores gains on Stanford tests, because TAKS are not administered prior to grade 3). In cases where the grade level(s) or subject(s) of a teacher were unclear (particularly Special Education Core Foundation Teachers), we assigned teachers the campus composite gain index as their CGI.

School Characteristics, 2010-11

To further ensure we are comparing employees who are as similar as possible with the exception of the characteristic(s) of interest, we include controls describing the grade level of each employee's school, the size of the student body, and the proportions of students at the school that are black, in poverty, in a bilingual or ESL program, in special education, or in the Gifted and Talented (GT) program. We primarily focus on employees' 2010-11 schools in order to capture the setting in which receiving an award would ostensibly affect their outcomes (employees received their award for 2009-10 in January 2011). We constructed a measure describing the grade level of each employee's school from a character variable in the 2010-11 PeopleSoft data file; we typically compare employees in Early Childhood Centers, middle schools, high schools, and 'other' schools (e.g., atypical combinations of grade levels) to employees in elementary schools. As a first step in addressing missing values, we triangulated data on the grade level of each employee's school with an indicator from the retention data files. The other variables describing schools were aggregated from student level data in the 2010-11 PEIMS data file. As the PEIMS data is a census of all of the students in HISD, we were able to construct a measure of school size by summing the number of students linked to each school ID. Similarly, we aggregated student level measures with a mean for the proportional measures of the student body. We chose to focus on the proportion of black rather than Hispanic students, because Hispanic students are the majority in HISD, and because schools with higher proportions of black students exhibited more differences in ASPIRE award distributions and poorer outcomes than did schools with higher proportions of Hispanic students. Exploratory analyses suggested that bilingual programs are more common among younger students, and ESL programs are most common among older students; because our analyses combine students across grade levels, we created a single indicator for bilingual or ESL program participation. The proportion of students who are LEP was highly correlated with the proportion in bilingual or ESL programs; we chose to use the latter measure because it captured a smaller but still sizeable minority within HISD. Because the vast majority of the HISD student population is economically disadvantaged, we chose to focus on the smaller set that is in poverty to capture more distinctive schools within this district.

Dependent Variables

Teacher Retained in HISD, August 2011

We used six measures from the 2010-11 retention data file to describe employee retention: job function, employment status, and school of employment at times 1 and 2. For most employees, times 1 and 2 are Augusts of two subsequent years; for 2009-10, time 2 usually represents August 2011. Following the lead of HISD research staff, we classify employees with a time 2 status of 'Active,' 'Leave,' 'Paid Leave,' or 'Suspended' as retained, and employees with a time 2 status of 'Terminated,' 'Retired,' or 'Death' as not retained. Per HISD staff, employees who quit are coded as 'Terminated.' Because the termination codes were not unique identifiers, we are unable at this point to distinguish employees who quit from those who were terminated. We classified employees who were in the 2009-10 ASPIRE award file but not in the 2010-11 retention data file at all as not retained. At the time of these analyses, the data file correcting employee statuses that were incorrect in the primary retention file for 2010-11 had not yet been released. Our results should not be too biased by this, as updates are provided on around 10 to 150 employees each year, which means that the 2010-11 retention data we have is inaccurate for 1% of HISD employees at most. We exclude charter school teachers from retention analyses, because these teachers were not well represented in the retention data file.

2010-11 Teacher Attendance Rate

HISD's attendance data files offered four key variables: hours scheduled, hours present, mandatory absence hours, requested absence hours. Requested absence hours include funeral leave, personal leave, religious holiday, sick leave, unpaid leave, vacation pay, local personal leave, supplemental sick leave, and state sick leave; whereas mandatory absence hours include compensatory time, jury duty, military leave, worker's compensation, and assault leave (Zimmerman et al. 2011). Across most of the attendance data files, the indicator for hours scheduled did not equal the sum of hours present, mandatory absence hours, and requested absence hours. This same data issue is documented in HISD reports (Zimmerman et al. 2011), and we took their lead to construct our own measure of hours scheduled by summing the other three measures. We also constructed a measure of hours scheduled that excluded mandatory absence hours, to again align with methods used in previous evaluations of the ASPIRE program (Zimmerman et al. 2011). Because these two measures of hours scheduled did not lead to different results, we present the results using the measure of hours scheduled that includes mandatory absence hours.

We focus on hours present rather than hours absent, in order to be consistent with previous HISD reports, and to facilitate more intuitive results (in which a positive association between receiving an award and attendance indicates a desired effect of the ASPIRE program). In order to maintain comparability in our measure of hours present across teachers scheduled for different numbers of hours, we include hours scheduled as an exposure indicator in our Poisson regression analyses. For similar reasons, we identified outliers (more than three standard deviations above or below the mean) using a standardized version of a proportional measure of hours present (hours present divided by hours scheduled). To ensure outliers did not bias our results from regression analyses, we set the cases identified as outliers to missing on our dependent variable, hours present. We exclude charter school teachers from attendance analyses, because these teachers were not well represented in the attendance data file.

Math and Reading Teachers' 2010-11 Mean Student Test Score Gains

To evaluate the effect of receiving an award on student achievement, we constructed measures describing math and reading/ELA teachers' mean student gains from 2009-10 to 2010-11 on the math and reading[17] TAKS and Stanford tests. The Stanford Achievement Test Series, Tenth Edition (referred to as 'Stanford tests' in this evaluation) are used in schools across the nation to assess children's growth from kindergarten through high school. The TAKS were standardized tests administered to Texas students in grades 3-11 from 2003 through 2012. We focus on the math and reading versions of these tests because they are administered in more grade levels than other subjects, are used more extensively for accountability purposes, and are used more often for ASPIRE award determinations. We do not expect features of these tests to disadvantage certain schools within HISD, because the administration of these tests is district-wide, and students in all schools take these tests on a consistent date set by the state (http://www.houstonisd.org/site/Default.aspx?PageID=31515).

---

[17] The TAKS reading test is formally referred to as a 'reading/ELA' test (English Language Arts and Reading).

Students' scores on the English-language Stanford tests were in single data files for each school year. Students' scores on the English- and Spanish-language TAKS were in single data files for each school year, but some grade levels were in separate files. Because grade promotion is determined by students' scores on the TAKS for grades 5 and 8, the TAKS are administered earlier in the school year for these grade levels so that these students have the opportunity to retake the test if they don't pass. In the case of duplicate test scores, we privileged test scores from the grade 5 and 8 data files over test scores from the grades 3-10 data file. In the case of duplicate test scores within the same data file, we privileged the higher test score (which is consistent with HISD processes).

Test scores are vertically scaled so that growth in knowledge is represented by the difference in a student's test scores in the same subject from one grade level to the next. All Stanford test scores are vertically scaled, so we are able to model gains using these test scores among students in grades 1-11. Although there is concern that vertical scales are not valid in the high school years because students take different levels of coursework, a counter argument is that students are grouped by ability and have different opportunities to learn throughout elementary school as well. Mean test scores within grades 9 through 11 also increase with each additional year of schooling as would be expected with vertically scaled scores. TEA started vertically scaling TAKS scores in 2009 for grades 3-8, so we are able to model gains using these test scores among students in grades 4-8. We exclude all test versions that did not have a comparable vertical scale. The 'TAKS-M' (TAKS-Modified) and 'TAKS-Alt' (TAKS-Alternate) versions are different assessments from the regular TAKS which were created for students in special education with more severe disabilities (Texas Education Agency 2011b), and these scores are not vertically scaled at all. Most students flagged as special education in the Stanford test data files had valid test scores.

We were unable to find a conclusive statement of whether the Accommodated and LAT (linguistically accommodated test) versions of the TAKS are on the same vertical scale as the regular TAKS, even after extensive searches through documentation accompanying the data files, online TEA records, and an email and phone call with TEA representatives. These test versions indicate that students received accommodations during the administration of the test (e.g., extended time, different setting), because they are in special education for a less severe disability, or are not proficient in English. The TEA representative did say that they conduct analyses with the A, K, and L versions of the test (which are data file codes that align with the regular, Accommodated, and LAT versions). Our exploratory analyses showed that the ranges, and the 'Met Standard' and 'Commended Performance' cutoff values, were very similar across these three versions of the TAKS. Based on these pieces of evidence, we include students who took the regular, Accommodated, and LAT versions of the TAKS in this evaluation's achievement analyses.

We exclude students who took Spanish versions of the TAKS or Stanford, because these scores were not on the same vertical scale. Students are most likely to take Spanish versions of tests in the early elementary grades. We estimate that our analyses exclude from 10-30% of students in the earlier grades because they took a Spanish version of the test.

The TAKS data files also include an indicator of whether each student's test score is valid; we exclude test scores that are not valid (e.g., student is LEP exempt, student was ill or caught cheating). It was not uncommon for a student's grade level to be inconsistent both within and across test data files from the same school year. Some students may have multiple grade levels because of grade and school mobility, while some grade level inconsistencies appeared to be data errors. We triangulated data from all of the test data files and the linkage files to address missing values and achieve a consistent grade level for each student. We exclude students who were held back or skipped grade(s) between 2009-10 and 2010-11, because their growth on test scores wouldn't be comparable to the other students; fortunately, these students only comprised 2-3% of the students with valid test scores. We subtracted each student's score on each test in 2009-10 from their score on each test in 2010-11 to construct a gain score for each of the four tests. Because we only construct gain scores for students who had a valid test score, on the correct vertical scale, for both school years, students can have gain scores for one, two, three, or all four tests. We standardized students' gains within each test and grade level so that each student's gains were relative to those of other students in their grade, and comparable across grade levels.

The linking data files characterize educators as working in the subjects of Language Arts, Math, Reading, Science, and/or Social Studies (educators who worked in multiple subjects appeared multiple times in the data files). Linking decisions are made at the campus level which results in some inconsistencies in linking decisions across campuses. For instance, some schools link special education teachers in inclusive settings to only special education students, whereas others link them to both special and regular education students in the inclusive class. The grade level of the employee is not always consistent with the grade level of the student(s). The campus-based nature of these linkages may also explain why Elective/Ancillary Teachers, Instructional Support Staff, and Teaching Assistants were linked to students to varying degrees (though to a much lesser degree than Core Teachers). We first restrict our analytic sample for achievement analyses to core teachers, as indicated by their ASPIRE category. Because our analyses focus on math and reading tests, we only link math teachers to students' math test scores, and reading/ELA teachers to students' reading test scores. Regardless of linkages, we only include core teachers of pre-kindergarten through grade 2 (as indicated by their ASPIRE category) in analyses focused on Stanford tests (the TAKS is not administered until grade 3). Because we rely on campus-level decisions of the students that should be linked to each teacher in our achievement analyses, as HISD does for award decisions, the teachers included in our analyses should also have CGIs. After linking students to their teachers, we aggregated students' gain scores to a mean student test score gain for each teacher and each test. Although teachers in charter schools are less likely to have mean student test score gains than teachers in other schools, their representation was sufficient to warrant inclusion; we include a flag indicating charter status in these regression analyses.

APPENDIX 6: REGRESSION MODELS PREDICTING CORE TEACHER RETENTION

Table 9, Part 1 of 3: Log Odds from Logistic Regression Models Predicting Core Teachers' Retention in August 2011

	Model 1			Model 2			Model 3			Model 4			Model 5		
	B	(SE)		B	(SE)		B	(SE)		B	(SE)		B	(SE)	
2009-10 ASPIRE Awards															
Received Strand I award	0.226	(0.15)		-0.017	(0.09)										
Received Strand II award	0.369	(0.24)		0.783	(0.23)	**									
Received Strand III award	0.610	(0.25)	*	0.564	(0.15)	***									
Total award amount							0.084	(0.02)	***						
Total award amount as proportion of salary										4.055	(0.82)	***			
Total award amount as proportion of school's mean award amount													0.251	(0.06)	***
Interactions with 'Received Strand I award'															
School is higher need	-0.446	(0.19)	*												
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Secondary math	0.034	(0.28)													
Secondary science	-0.078	(0.32)													
Bilingual/ESL	-0.105	(0.17)													
Special education	-0.258	(0.23)													
Interactions with 'Received Strand II award'															
School is higher need	-0.050	(0.16)													
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Secondary math	0.424	(0.26)	+												
Secondary science	-0.180	(0.30)													
Bilingual/ESL	-0.025	(0.17)													
Special education	-0.267	(0.23)													
Proportion teachers on campus who received Strand II award	-0.049	(0.32)													

Table 9, Part 2 of 3: Log Odds from Logistic Regression Models Predicting Core Teachers' Retention in August 2011

	Model 1, cont.			Model 2, cont.			Model 3, cont.			Model 4, cont.			Model 5, cont.		
	B	(SE)		B	(SE)		B	(SE)		B	(SE)		B	(SE)	
Strand II award determined on basis of:															
Own students (grades 3-6, self-contained) (ref)				-											
Own students (grades 3-8, departmentalized)				-0.449	(0.26)	+									
Students in applicable subject(s) (grades 9-12)				-0.337	(0.31)										
3rd graders, reading and math (grades Pre-K-2)				-0.583	(0.26)	*									
Students in applicable subject(s) (special ed)				-0.601	(0.34)	+									
Interactions with 'Received Strand III award'															
School is higher need	-0.027	(0.30)													
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Secondary math	-0.183	(0.51)													
Secondary science	0.018	(0.55)													
Bilingual/ESL	-0.087	(0.36)													
Special education	0.192	(0.36)													
Teacher Characteristics															
Male	0.068	(0.08)		0.082	(0.08)		0.075	(0.08)		0.074	(0.08)		0.076	(0.08)	
Race:															
White, non-Hispanic (ref)	-			-			-			-			-		
Black, non-Hispanic	0.629	(0.09)	***	0.622	(0.09)	***	0.621	(0.09)	***	0.620	(0.09)	***	0.623	(0.09)	***
White, Hispanic	0.572	(0.11)	***	0.524	(0.10)	***	0.540	(0.11)	***	0.539	(0.11)	***	0.540	(0.11)	***
Asian	0.737	(0.17)	***	0.737	(0.17)	***	0.800	(0.18)	***	0.802	(0.18)	***	0.806	(0.18)	***
Other race	0.352	(0.16)	*	0.321	(0.16)	*	0.324	(0.16)	*	0.322	(0.16)	+	0.326	(0.16)	*
Educational attainment:															
Less than Master's degree (ref)	-			-			-			-			-		
Master's degree or PhD	0.028	(0.07)		0.036	(0.07)		0.024	(0.07)		0.032	(0.07)		0.026	(0.07)	
Years of experience:															
3 or fewer	-0.707	(0.08)	***	-0.696	(0.08)	***	-0.727	(0.09)	***	-0.745	(0.09)	***	-0.726	(0.09)	***
4-10 (ref)	-			-			-			-			-		
11-25	0.021	(0.09)		0.018	(0.09)		0.068	(0.10)		0.103	(0.10)		0.061	(0.10)	
More than 25	-1.062	(0.10)	***	-1.059	(0.10)	***	-1.029	(0.11)	***	-0.955	(0.11)	***	-1.038	(0.11)	***
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Secondary math	-0.224	(0.46)													
Secondary science	-0.076	(0.51)													
Bilingual/ESL	0.013	(0.34)													
Special education	0.232	(0.32)													

Table 9, Part 3 of 3: Log Odds from Logistic Regression Models Predicting Core Teachers' Retention in August 2011

	Model 1, cont.			Model 2, cont.			Model 3, cont.			Model 4, cont.			Model 5, cont.		
	B	(SE)		B	(SE)		B	(SE)		B	(SE)		B	(SE)	
Strand II award determined on basis of:															
Own students (grades 3-6, self-contained) (ref)				-			-			-			-		
Own students (grades 3-8, departmentalized)				0.263	(0.23)		-0.171	(0.14)		-0.174	(0.14)		-0.164	(0.14)	
Students in applicable subject(s) (grades 9-12)				-0.029	(0.32)		-0.269	(0.23)		-0.266	(0.23)		-0.217	(0.23)	
3rd graders, reading and math (grades Pre-K-2)				0.418	(0.22)	+	-0.023	(0.12)		-0.028	(0.12)		-0.005	(0.12)	
Students in applicable subject(s) (special ed)				0.429	(0.30)		0.047	(0.18)		0.040	(0.18)		0.057	(0.18)	
School Characteristics															
Grade level of school:															
Early Childhood Center	0.738	(0.63)		0.692	(0.62)		0.659	(0.63)		0.664	(0.63)		0.577	(0.62)	
Elementary (ref)	-			-			-			-			-		
Middle school	-0.260	(0.15)	+	-0.265	(0.16)	+	-0.337	(0.16)	*	-0.334	(0.16)	*	-0.249	(0.16)	
High school	-0.473	(0.19)	*	-0.248	(0.24)		-0.380	(0.25)		-0.381	(0.25)		-0.341	(0.25)	
Other	-1.083	(0.27)	***	-0.924	(0.28)	**	-0.972	(0.29)	**	-0.973	(0.29)	**	-0.921	(0.29)	**
Size of student body	0.000	(0.00)		0.000	(0.00)		0.000	(0.00)		0.000	(0.00)		0.000	(0.00)	
Proportion student body:															
Black	0.155	(0.30)		0.126	(0.30)		0.085	(0.30)		0.085	(0.30)		0.107	(0.30)	
In poverty	-1.829	(0.54)	**	-1.662	(0.54)	**	-1.674	(0.56)	**	-1.666	(0.56)	**	-1.644	(0.56)	**
In bilingual/ESL program	0.091	(0.37)		0.143	(0.37)		-0.015	(0.38)		-0.012	(0.38)		0.071	(0.38)	
In GT program	1.223	(0.46)	**	1.328	(0.46)	**	1.091	(0.46)	*	1.097	(0.46)	*	1.135	(0.47)	*
In special education	0.750	(0.71)		0.882	(0.71)		0.851	(0.73)		0.843	(0.73)		0.638	(0.73)	
School is higher need	0.248	(0.28)		-0.041	(0.12)		-0.058	(0.12)		-0.059	(0.12)		-0.084	(0.12)	
Proportion teachers on campus who received Strand II award	0.163	(0.26)													
Constant	0.994	(0.41)	*	0.822	(0.39)	*	1.863	(0.33)	***	1.859	(0.33)	***	1.819	(0.33)	***
Log of panel level var. component	-1.542	(0.22)	***	-1.514	(0.22)	***	-1.489	(0.22)	***	-1.486	(0.22)	***	-1.488	(0.22)	***
Between-subject SD	0.462	(0.05)	***	0.469	(0.05)	***	0.475	(0.05)	***	0.476	(0.05)	***	0.475	(0.05)	***
Interclass correlation	0.061	(0.01)	***	0.063	(0.01)	***	0.064	(0.01)	***	0.064	(0.01)	***	0.064	(0.01)	***
Core teachers (n)	8553			8553			8250			8250			8250		
Schools (n)	276			276			275			275			275		

Note: All of these models exclude core teachers ineligible for the 2009-10 ASPIRE program because of their job, campus, or opting out, as well as charter school teachers (who were generally not included in the retention data file). Models 3-5 additionally exclude core teachers who did not receive any 2009-10 award. *** p<0.001, ** p<0.01, * p<0.05, +p<0.10.

Appendix 7: Regression Models Predicting Core Teacher Attendance

Table 10, Part 1 of 3: Log Odds from Poisson Regression Models Predicting Core Teachers' Hours Present in 2010-11, with Hours Scheduled as an Exposure Indicator

	Model 1			Model 2			Model 3			Model 4			Model 5		
	B	(SE)		B	(SE)		B	(SE)		B	(SE)		B	(SE)	
2009-10 ASPIRE Awards															
Received Strand I award	-0.002	(0.00)		-0.003	(0.00)	*									
Received Strand II award	0.002	(0.00)		0.001	(0.00)										
Received Strand III award	0.011	(0.00)	**	0.014	(0.00)	***									
Total award amount							0.000	(0.00)	+						
Total award amount as proportion of salary										0.015	(0.01)	+			
Total award amount as proportion of school's mean award amount													0.000	(0.00)	
Interactions with 'Received Strand I award'															
School is higher need	-0.003	(0.00)													
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Secondary math	-0.001	(0.00)													
Secondary science	-0.002	(0.00)													
Bilingual/ESL	0.000	(0.00)													
Special education	0.004	(0.00)	+												
Interactions with 'Received Strand II award'															
School is higher need	0.002	(0.00)													
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Secondary math	0.002	(0.00)													
Secondary science	-0.007	(0.00)	*												
Bilingual/ESL	-0.004	(0.00)	*												
Special education	-0.006	(0.00)	**												
Proportion teachers on campus who received Strand II award	0.000	(0.00)													

Table 10, Part 2 of 3: Log Odds from Poisson Regression Models Predicting Core Teachers' Hours Present in 2010-11, with Hours Scheduled as an Exposure Indicator

	Model 1, cont.			Model 2, cont.			Model 3, cont.			Model 4, cont.			Model 5, cont.		
	B	(SE)		B	(SE)		B	(SE)		B	(SE)		B	(SE)	
Strand II award determined on basis of:															
Own students (grades 3-6, self-contained) (ref)				-											
Own students (grades 3-8, departmentalized)				0.000	(0.00)										
Students in applicable subject(s) (grades 9-12)				0.003	(0.00)										
3rd graders, reading and math (grades Pre-K-2)				-0.001	(0.00)										
Students in applicable subject(s) (special ed)				0.001	(0.00)										
Interactions with 'Received Strand III award'															
School is higher need	0.003	(0.00)													
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Secondary math	0.011	(0.01)													
Secondary science	-0.010	(0.01)													
Bilingual/ESL	0.000	(0.00)													
Special education	0.008	(0.00)	+												
Teacher Characteristics															
Male	0.005	(0.00)	***	0.004	(0.00)	***	0.005	(0.00)	***	0.005	(0.00)	***	0.005	(0.00)	***
Race:															
White, non-Hispanic (ref)	-			-			-			-			-		
Black, non-Hispanic	0.001	(0.00)		0.001	(0.00)		0.001	(0.00)		0.001	(0.00)		0.001	(0.00)	
White, Hispanic	0.002	(0.00)	+	0.002	(0.00)	*	0.002	(0.00)	*	0.002	(0.00)	*	0.002	(0.00)	*
Asian	0.006	(0.00)	***	0.006	(0.00)	***	0.007	(0.00)	***	0.007	(0.00)	***	0.007	(0.00)	***
Other race	0.003	(0.00)	+	0.003	(0.00)	*	0.003	(0.00)	+	0.003	(0.00)	+	0.003	(0.00)	+
Educational attainment:															
Less than Master's degree (ref)	-			-			-			-			-		
Master's degree or PhD	0.000	(0.00)		-0.001	(0.00)		0.000	(0.00)		0.000	(0.00)		0.000	(0.00)	
Years of experience:															
3 or fewer	0.001	(0.00)		0.001	(0.00)		0.001	(0.00)		0.001	(0.00)		0.001	(0.00)	
4-10 (ref)	-			-			-			-			-		
11-25	0.002	(0.00)	**	0.002	(0.00)	**	0.003	(0.00)	***	0.003	(0.00)	***	0.003	(0.00)	***
More than 25	-0.002	(0.00)	+	-0.002	(0.00)	+	-0.001	(0.00)		-0.001	(0.00)		-0.001	(0.00)	
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Secondary math	-0.008	(0.01)													
Secondary science	0.017	(0.01)	**												
Bilingual/ESL	0.003	(0.00)													
Special education	-0.006	(0.00)													

Table 10, Part 3 of 3: Log Odds from Poisson Regression Models Predicting Core Teachers' Hours Present in 2010-11, with Hours Scheduled as an Exposure Indicator

	Model 1, cont.			Model 2, cont.			Model 3, cont.			Model 4, cont.			Model 5, cont.		
	B	(SE)		B	(SE)		B	(SE)		B	(SE)		B	(SE)	
Received a new hire recruitment s	0.002	(0.00)	*	0.002	(0.00)	*	0.002	(0.00)	*	0.002	(0.00)	*	0.002	(0.00)	*
Strand II award determined on basis of:															
Own students (grades 3-6, self-contained) (ref)				-			-			-			-		
Own students (grades 3-8, departmentalized)				-0.002	(0.00)		-0.003	(0.00)	*	-0.003	(0.00)	*	-0.003	(0.00)	*
Students in applicable subject(s) (grades 9-12)				-0.003	(0.00)		0.000	(0.00)		0.000	(0.00)		0.000	(0.00)	
3rd graders, reading and math (grades Pre-K-2)				-0.004	(0.00)	+	-0.005	(0.00)	***	-0.005	(0.00)	***	-0.005	(0.00)	***
Students in applicable subject(s) (special ed)				-0.005	(0.00)		-0.003	(0.00)	+	-0.003	(0.00)	+	-0.003	(0.00)	*
2008-09 attendance rate	0.092	(0.01)	***	0.093	(0.01)	***	0.092	(0.01)	***	0.092	(0.01)	***	0.092	(0.01)	***
School Characteristics															
Grade level of school:															
Early Childhood Center	0.005	(0.01)		0.006	(0.01)		0.005	(0.01)		0.005	(0.01)		0.005	(0.01)	
Elementary (ref)	-			-			-			-			-		
Middle school	-0.004	(0.00)	*	-0.005	(0.00)	*	-0.005	(0.00)	*	-0.005	(0.00)	*	-0.005	(0.00)	*
High school	-0.003	(0.00)		-0.003	(0.00)		-0.005	(0.00)		-0.005	(0.00)		-0.005	(0.00)	
Other	-0.009	(0.00)	*	-0.009	(0.00)	*	-0.012	(0.00)	**	-0.012	(0.00)	**	-0.012	(0.00)	**
Size of student body	0.000	(0.00)		0.000	(0.00)		0.000	(0.00)	+	0.000	(0.00)	+	0.000	(0.00)	+
Proportion student body:															
Black	-0.001	(0.00)		-0.001	(0.00)		-0.002	(0.00)		-0.002	(0.00)		-0.002	(0.00)	
In poverty	-0.012	(0.01)	+	-0.012	(0.01)		-0.015	(0.01)	*	-0.015	(0.01)	*	-0.015	(0.01)	*
In bilingual/ESL program	0.003	(0.00)		0.003	(0.00)		0.003	(0.00)		0.003	(0.00)		0.003	(0.00)	
In GT program	0.007	(0.01)		0.007	(0.01)		0.005	(0.01)		0.005	(0.01)		0.005	(0.01)	
In special education	0.005	(0.01)		0.005	(0.01)		0.009	(0.01)		0.008	(0.01)		0.008	(0.01)	
School is higher need	-0.003	(0.00)		-0.001	(0.00)		0.000	(0.00)		0.000	(0.00)		-0.001	(0.00)	
Proportion teachers on campus who received Strand II award	0.004	(0.00)													
Constant	-0.141	(0.01)	***	-0.137	(0.01)	***	-0.121	(0.01)	***	-0.121	(0.01)	***	-0.120	(0.01)	***
Log of panel level var. component	1.000	-	***	1.000	-	***	1.000	-	***	1.000	-	***	1.000	-	***
Between-subject SD	-9.581	(0.13)	***	-9.565	(0.13)	***	-9.635	(0.13)	***	-9.635	(0.13)	***	-9.620	(0.13)	***
Interclass correlation	0.000	(0.00)	***	0.000	(0.00)	***	0.000	(0.00)	***	0.000	(0.00)	***	0.000	(0.00)	***
Core teachers (n)	8313			8313			8040			8040			8040		
Schools (n)	274			274			273			273			273		

Note: All of these models exclude core teachers ineligible for the 2009-10 ASPIRE program because of their job, campus, or opting out, as well as charter school teachers (who were generally not included in the attendance data file). Models 3-5 additionally exclude core teachers who did not receive any 2009-10 award. *** p<0.001, ** p<0.01, * p<0.05, +p<0.10.

APPENDIX 8: REGRESSION MODELS PREDICTING MATH TEACHERS' MEAN STUDENT GAINS ON MATH TAKS

Table 11, Part 1 of 3: Coefficients from Linear Regression Models Predicting Math Teachers' Mean Student Gains on the Math TAKS from 2009-10 to 2010-11

	Model 1			Model 2			Model 3			Model 4			Model 5		
	B	(SE)		B	(SE)		B	(SE)		B	(SE)		B	(SE)	
	Model 1			Model 2			Model 3[a]			Model 4[b]			Model 5[b]		
2009-10 ASPIRE Awards															
Received Strand I award	-0.041	(0.06)		-0.036	(0.04)										
Received Strand II award	0.171	(0.11)		0.184	(0.08)	*									
Received Strand III award	0.334	(0.21)		0.095	(0.10)										
Total award amount							0.019	(0.01)	**						
Total award amount as proportion of salary										0.821	(0.31)	**			
Total award amount as proportion of school's mean award amount													0.047	(0.02)	**
Interactions with 'Received Strand I award'															
School is higher need	-0.039	(0.08)													
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Secondary math	0.115	(0.10)													
Bilingual/ESL	0.047	(0.08)													
Special education	0.102	(0.10)													
Interactions with 'Received Strand II award'															
School is higher need	-0.141	(0.07)	*												
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Secondary math	0.042	(0.08)													
Bilingual/ESL	-0.032	(0.10)													
Special education	-0.067	(0.11)													
Proportion teachers on campus who received Strand II award	-0.053	(0.14)													

Table 11, Part 2 of 3: Coefficients from Linear Regression Models Predicting Math Teachers' Mean Student Gains on the Math TAKS from 2009-10 to 2010-11

	Model 1, cont.			Model 2, cont.			Model 3, cont.			Model 4, cont.			Model 5, cont.		
	B	(SE)		B	(SE)		B	(SE)		B	(SE)		B	(SE)	
Strand II award determined on basis of:															
Own students (grades 3-6, self-contained) (ref)				-											
Own students (grades 3-8, departmentalized)				-0.133	(0.09)										
Students in applicable subject(s) (special ed)				-0.182	(0.14)										
Interactions with 'Received Strand III award'															
School is higher need	-0.104	(0.21)													
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Secondary math	-0.480	(0.24)	*												
Bilingual/ESL	0.105	(0.27)													
Special education	-0.410	(0.40)													
Teacher Characteristics															
Male	0.019	(0.03)		0.015	(0.03)		0.008	(0.03)		0.007	(0.03)		0.010	(0.03)	
Race:															
White, non-Hispanic (ref)	-			-			-			-			-		
Black, non-Hispanic	0.007	(0.04)		0.006	(0.04)		-0.002	(0.04)		-0.002	(0.04)		0.001	(0.04)	
White, Hispanic	0.080	(0.05)		0.082	(0.05)	+	0.087	(0.05)	+	0.086	(0.05)	+	0.086	(0.05)	+
Asian	0.011	(0.07)		0.014	(0.07)		0.007	(0.07)		0.008	(0.07)		0.011	(0.07)	
Other race	0.079	(0.08)		0.073	(0.08)		0.095	(0.08)		0.094	(0.08)		0.098	(0.08)	
Educational attainment:															
Less than Master's degree (ref)	-			-			-			-			-		
Master's degree or PhD	0.018	(0.03)		0.014	(0.03)		0.017	(0.03)		0.018	(0.03)		0.016	(0.03)	
Years of experience:															
3 or fewer	0.027	(0.04)		0.025	(0.04)		0.027	(0.04)		0.022	(0.04)		0.027	(0.04)	
4-10 (ref)	-			-			-			-			-		
11-25	-0.011	(0.04)		-0.020	(0.04)		-0.014	(0.04)		-0.006	(0.04)		-0.019	(0.04)	
More than 25	-0.104	(0.06)	+	-0.102	(0.06)	+	-0.098	(0.06)	+	-0.082	(0.06)		-0.100	(0.06)	+
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Secondary math	0.367	(0.22)													
Bilingual/ESL	-0.171	(0.26)													
Special education	0.309	(0.39)													

Table 11, Part 3 of 3: Coefficients from Linear Regression Models Predicting Math Teachers' Mean Student Gains on the Math TAKS from 2009-10 to 2010-11

	Model 1, cont.			Model 2, cont.			Model 3, cont.			Model 4, cont.			Model 5, cont.		
	B	(SE)		B	(SE)		B	(SE)		B	(SE)		B	(SE)	
Received a new hire recruitment s	-0.048	(0.05)		-0.082	(0.04)	*	-0.093	(0.04)	*	-0.093	(0.04)	*	-0.091	(0.04)	*
Strand II award determined on basis of:															
Own students (grades 3-6, self-contained) (ref)				-			-			-			-		
Own students (grades 3-8, departmentalized)				0.093	(0.09)		-0.039	(0.04)		-0.040	(0.04)		-0.036	(0.04)	
Students in applicable subject(s) (special ed)				0.112	(0.13)		-0.025	(0.06)		-0.027	(0.06)		-0.026	(0.06)	
2008-09 CGI	0.055	(0.02)	*	0.060	(0.02)	*	0.052	(0.02)	*	0.054	(0.02)	*	0.059	(0.02)	*
School Characteristics															
Grade level of school:															
Elementary (ref)	-			-			-			-			-		
Middle school	-0.047	(0.08)		-0.041	(0.07)		-0.053	(0.07)		-0.052	(0.07)		-0.032	(0.07)	
Other	-0.216	(0.14)		-0.211	(0.14)		-0.204	(0.15)		-0.204	(0.15)		-0.203	(0.15)	
Size of student body	0.000	(0.00)		0.000	(0.00)		0.000	(0.00)		0.000	(0.00)		0.000	(0.00)	
Proportion student body:															
Black	0.047	(0.13)		0.046	(0.13)		0.043	(0.14)		0.041	(0.14)		0.054	(0.14)	
In poverty	-0.202	(0.22)		-0.195	(0.23)		-0.236	(0.24)		-0.233	(0.24)		-0.213	(0.23)	
In bilingual/ESL program	0.138	(0.17)		0.118	(0.17)		0.104	(0.17)		0.104	(0.17)		0.125	(0.17)	
In GT program	0.086	(0.21)		0.129	(0.22)		0.070	(0.22)		0.073	(0.22)		0.093	(0.22)	
In special education	0.424	(0.55)		0.655	(0.56)		0.638	(0.59)		0.636	(0.59)		0.536	(0.58)	
School is higher need	0.225	(0.21)		-0.002	(0.05)		0.000	(0.05)		0.000	(0.05)		-0.009	(0.05)	
Proportion teachers on campus who received Strand II award	-0.064	(0.13)													
Constant	-0.407	(0.27)		-0.318	(0.19)	+	-0.091	(0.15)		-0.088	(0.15)		-0.110	(0.15)	
Log of panel level var. component	0.137		***	0.165		***	0.175		***	0.175		***	0.170		***
Between-subject SD	0.440		***	0.440		***	0.440		***	0.441		***	0.442		***
Interclass correlation	0.088	-		0.123	-		0.136	-		0.136	-		0.129	-	
Core teachers (n)	1112			1112			1087			1087			1087		
Schools (n)	217			217			217			217			217		

Note: All of these models exclude core teachers ineligible for the 2009-10 ASPIRE program because of their job, campus, or opting out. Models 3-5 additionally exclude core teachers who did not receive any 2009-10 award. *** p<0.001, ** p<0.01, * p<0.05, +p<0.10.

Appendix 9: Regression Models Predicting Reading/ELA Teachers' Mean Student Gains on Reading TAKS

Table 12, Part 1 of 3: Coefficients from Linear Regression Models Predicting Reading/ELA Teachers' Mean Student Gains on the Reading TAKS from 2009-10 to 2010-11

	Model 1		Model 2			Model 3		Model 4		Model 5	
	B	(SE)	B	(SE)		B	(SE)	B	(SE)	B	(SE)
2009-10 ASPIRE Awards											
Received Strand I award	0.000	(0.06)	-0.062	(0.04)	+						
Received Strand II award	-0.069	(0.09)	-0.010	(0.07)							
Received Strand III award	0.013	(0.14)	0.058	(0.08)							
Total award amount						0.003	(0.01)				
Total award amount as proportion of salary								0.120	(0.28)		
Total award amount as proportion of school's mean award amount										0.022	(0.02)
Interactions with 'Received Strand I award'											
School is higher need	-0.094	(0.08)									
Hard to staff subjects:											
Not hard-to-staff subject (ref)	-										
Bilingual/ESL	-0.045	(0.06)									
Special education	-0.070	(0.08)									
Interactions with 'Received Strand II award'											
School is higher need	0.037	(0.06)									
Hard to staff subjects:											
Not hard-to-staff subject (ref)	-										
Bilingual/ESL	-0.064	(0.06)									
Special education	-0.087	(0.09)									
Proportion teachers on campus who received Strand II award	0.147	(0.13)									

Table 12, Part 2 of 3: Coefficients from Linear Regression Models Predicting Reading/ELA Teachers' Mean Student Gains on the Reading TAKS from 2009-10 to 2010-11

	Model 1, cont.			Model 2, cont.			Model 3, cont.			Model 4, cont.			Model 5, cont.		
	B	(SE)		B	(SE)		B	(SE)		B	(SE)		B	(SE)	
Strand II award determined on basis of:															
Own students (grades 3-6, self-contained) (ref)				-											
Own students (grades 3-8, departmentalized)				0.013	(0.08)										
Students in applicable subject(s) (special ed)				-0.266	(0.13)	*									
Interactions with 'Received Strand III award'															
School is higher need	0.190	(0.16)													
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Bilingual/ESL	0.113	(0.21)													
Special education	-0.368	(0.19)	*												
Teacher Characteristics															
Male	0.030	(0.03)		0.032	(0.03)		0.027	(0.03)		0.027	(0.03)		0.029	(0.03)	
Race:															
White, non-Hispanic (ref)	-			-			-			-			-		
Black, non-Hispanic	0.067	(0.03)	*	0.064	(0.03)	+	0.070	(0.03)	*	0.070	(0.03)	*	0.070	(0.03)	*
White, Hispanic	-0.042	(0.04)		-0.054	(0.04)		-0.044	(0.04)		-0.044	(0.04)		-0.044	(0.04)	
Asian	0.075	(0.06)		0.060	(0.06)		0.083	(0.06)		0.083	(0.06)		0.084	(0.06)	
Other race	0.121	(0.07)	+	0.108	(0.07)		0.133	(0.07)	*	0.133	(0.07)	*	0.134	(0.07)	*
Educational attainment:															
Less than Master's degree (ref)	-			-			-			-			-		
Master's degree or PhD	-0.026	(0.03)		-0.027	(0.03)		-0.022	(0.03)		-0.022	(0.03)		-0.021	(0.03)	
Years of experience:															
3 or fewer	-0.001	(0.03)		0.006	(0.03)		0.011	(0.03)		0.010	(0.03)		0.013	(0.03)	
4-10 (ref)	-			-			-			-			-		
11-25	-0.029	(0.03)		-0.029	(0.03)		-0.033	(0.03)		-0.031	(0.03)		-0.032	(0.03)	
More than 25	0.022	(0.04)		0.027	(0.04)		0.026	(0.04)		0.028	(0.04)		0.026	(0.04)	
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Bilingual/ESL	-0.088	(0.21)													
Special education	0.445	(0.17)	*												

Table 12, Part 3 of 3: Coefficients from Linear Regression Models Predicting Reading/ELA Teachers' Mean Student Gains on the Reading TAKS from 2009-10 to 2010-11

	Model 1, cont.			Model 2, cont.			Model 3, cont.			Model 4, cont.			Model 5, cont.		
	B	(SE)		B	(SE)		B	(SE)		B	(SE)		B	(SE)	
Received a new hire recruitment s	-0.050	(0.04)		-0.078	(0.03)	**	-0.083	(0.03)	**	-0.083	(0.03)	**	-0.083	(0.03)	**
Strand II award determined on basis of:															
Own students (grades 3-6, self-contained) (ref)				-			-			-			-		
Own students (grades 3-8, departmentalized)				-0.059	(0.07)		-0.053	(0.03)		-0.053	(0.03)		-0.051	(0.03)	
Students in applicable subject(s) (special ed)				0.210	(0.12)	+	0.003	(0.05)		0.003	(0.05)		0.011	(0.05)	
2008-09 CGI	0.046	(0.02)	*	0.046	(0.02)	**	0.042	(0.02)	*	0.042	(0.02)	*	0.040	(0.02)	*
School Characteristics															
Grade level of school:															
Elementary (ref)	-			-			-			-			-		
Middle school	-0.061	(0.06)		-0.042	(0.06)		-0.055	(0.06)		-0.055	(0.06)		-0.049	(0.06)	
Other	-0.130	(0.13)		-0.124	(0.13)		-0.123	(0.13)		-0.123	(0.13)		-0.122	(0.13)	
Charter	-0.225	(0.43)		-0.200	(0.43)		-0.172	(0.43)		-0.172	(0.43)		-0.189	(0.43)	
Size of student body	0.000	(0.00)		0.000	(0.00)		0.000	(0.00)		0.000	(0.00)		0.000	(0.00)	
Proportion student body:															
Black	0.324	(0.13)	*	0.322	(0.13)	*	0.281	(0.13)	*	0.281	(0.13)	*	0.281	(0.13)	*
In poverty	-0.414	(0.23)	+	-0.389	(0.22)	+	-0.437	(0.22)	+	-0.436	(0.22)	+	-0.427	(0.22)	+
In bilingual/ESL program	0.003	(0.16)		0.004	(0.16)		-0.042	(0.16)		-0.042	(0.16)		-0.040	(0.16)	
In GT program	0.196	(0.21)		0.238	(0.20)		0.186	(0.21)		0.186	(0.21)		0.181	(0.21)	
In special education	0.941	(0.35)	**	1.014	(0.34)	**	0.894	(0.36)	*	0.894	(0.36)	*	0.883	(0.36)	*
School is higher need	-0.258	(0.16)		-0.100	(0.05)	*	-0.077	(0.05)		-0.077	(0.05)		-0.080	(0.05)	
Proportion teachers on campus who received Strand II award	-0.120	(0.12)													
Constant	-0.013	(0.21)		-0.091	(0.16)		0.000	(0.13)		0.000	(0.13)		-0.026	(0.13)	
Log of panel level var. component	0.190		***	0.180		***	0.192		***	0.192		***	0.194		***
Between-subject SD	0.406		***	0.408		***	0.404		***	0.404		***	0.404		***
Interclass correlation	0.180	-		0.164	-		0.183	-		0.183	-		0.188	-	
Core teachers (n)	1298			1298			1267			1267			1267		
Schools (n)	220			220			219			219			219		

Note: All of these models exclude core teachers ineligible for the 2009-10 ASPIRE program because of their job, campus, or opting out. Models 3-5 additionally exclude core teachers who did not receive any 2009-10 award. *** p<0.001, ** p<0.01, * p<0.05, +p<0.10.

Appendix 10: Regression Models Predicting Math Teachers' Mean Student Gains on Stanford Math Test

Table 13, Part 1 of 3: Coefficients from Linear Regression Models Predicting Math Teachers' Mean Student Gains on the Stanford Math Test from 2009-10 to 2010-11

	Model 1			Model 2			Model 3			Model 4			Model 5		
	B	(SE)		B	(SE)		B	(SE)		B	(SE)		B	(SE)	
2009-10 ASPIRE Awards															
Received Strand I award	0.033	(0.04)		0.017	(0.03)										
Received Strand II award	-0.016	(0.06)		0.199	(0.05)	***									
Received Strand III award	-0.122	(0.10)		0.010	(0.05)										
Total award amount							0.019	(0.00)	***						
Total award amount as proportion of salary										0.903	(0.22)	***			
Total award amount as proportion of school's mean award amount													0.057	(0.01)	***
Interactions with 'Received Strand I award'															
School is higher need	-0.009	(0.05)													
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Secondary math	0.065	(0.07)													
Bilingual/ESL	0.002	(0.05)													
Special education	-0.068	(0.06)													
Interactions with 'Received Strand II award'															
School is higher need	0.020	(0.04)													
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Secondary math	-0.145	(0.06)	*												
Bilingual/ESL	0.042	(0.05)													
Special education	-0.061	(0.07)													
Proportion teachers on campus who received Strand II award	0.194	(0.09)	*												

Table 13, Part 2 of 3: Coefficients from Linear Regression Models Predicting Math Teachers' Mean Student Gains on the Stanford Math Test from 2009-10 to 2010-11

	Model 1, cont.			Model 2, cont.			Model 3, cont.			Model 4, cont.			Model 5, cont.		
	B	(SE)		B	(SE)		B	(SE)		B	(SE)		B	(SE)	
Strand II award determined on basis of:															
Own students (grades 3-6, self-contained) (ref)				-											
Own students (grades 3-8, departmentalized)				-0.093	(0.06)										
Students in applicable subject(s) (grades 9-12)				-0.319	(0.11)	**									
3rd graders, reading and math (grades Pre-K-2)				-0.205	(0.06)	**									
Students in applicable subject(s) (special ed)				-0.272	(0.09)	**									
Interactions with 'Received Strand III award'															
School is higher need	0.108	(0.11)													
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Secondary math	0.078	(0.14)													
Bilingual/ESL	-0.083	(0.15)													
Special education	0.246	(0.13)	+												
Teacher Characteristics															
Male	-0.022	(0.02)		-0.015	(0.02)		-0.012	(0.02)		-0.012	(0.02)		-0.011	(0.02)	
Race:															
White, non-Hispanic (ref)	-			-			-			-			-		
Black, non-Hispanic	0.011	(0.02)		0.016	(0.02)		0.008	(0.02)		0.008	(0.02)		0.012	(0.02)	
White, Hispanic	-0.012	(0.03)		-0.018	(0.03)		-0.010	(0.03)		-0.011	(0.03)		-0.009	(0.03)	
Asian	0.070	(0.04)	+	0.068	(0.04)	+	0.066	(0.04)	+	0.067	(0.04)	+	0.067	(0.04)	+
Other race	-0.069	(0.05)		-0.060	(0.05)		-0.059	(0.05)		-0.060	(0.05)		-0.057	(0.05)	
Educational attainment:															
Less than Master's degree (ref)	-			-			-			-			-		
Master's degree or PhD	0.052	(0.02)	**	0.049	(0.02)	**	0.052	(0.02)	**	0.053	(0.02)	**	0.052	(0.02)	**
Years of experience:															
3 or fewer	0.018	(0.02)		0.016	(0.02)		0.016	(0.02)		0.011	(0.02)		0.015	(0.02)	
4-10 (ref)	-			-			-			-			-		
11-25	-0.019	(0.02)		-0.019	(0.02)		-0.017	(0.02)		-0.009	(0.02)		-0.019	(0.02)	
More than 25	-0.042	(0.03)		-0.045	(0.03)		-0.040	(0.03)		-0.024	(0.03)		-0.042	(0.03)	
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Secondary math	-0.256	(0.13)	+												
Bilingual/ESL	0.046	(0.14)													
Special education	-0.312	(0.13)	*												

Table 13, Part 3 of 3: Coefficients from Linear Regression Models Predicting Math Teachers' Mean Student Gains on the Stanford Math Test from 2009-10 to 2010-11

	Model 1, cont.			Model 2, cont.			Model 3, cont.			Model 4, cont.			Model 5, cont.		
	B	(SE)		B	(SE)		B	(SE)		B	(SE)		B	(SE)	
Received a new hire recruitment s	0.021	(0.03)		0.021	(0.03)		0.008	(0.03)		0.008	(0.03)		0.010	(0.03)	
Strand II award determined on basis of:															
Own students (grades 3-6, self-contained) (ref)				-			-			-			-		
Own students (grades 3-8, departmentalized)				-0.003	(0.06)		-0.097	(0.03)	**	-0.099	(0.03)	**	-0.095	(0.03)	**
Students in applicable subject(s) (grades 9-12)				-0.086	(0.12)		-0.340	(0.08)	***	-0.340	(0.08)	***	-0.325	(0.08)	***
3rd graders, reading and math (grades Pre-K-2)				0.129	(0.06)	*	-0.014	(0.03)		-0.017	(0.03)		-0.009	(0.03)	
Students in applicable subject(s) (special ed)				-0.012	(0.08)		-0.194	(0.04)	***	-0.197	(0.04)	***	-0.191	(0.04)	***
2008-09 CGI	0.043	(0.01)	**	0.041	(0.01)	**	0.038	(0.01)	**	0.039	(0.01)	**	0.043	(0.01)	**
School Characteristics															
Grade level of school:															
Elementary (ref)	-			-			-			-			-		
Middle school	0.159	(0.05)	**	0.056	(0.04)		0.040	(0.05)		0.041	(0.05)		0.064	(0.05)	
High school	0.201	(0.07)	**	0.261	(0.08)	**	0.231	(0.08)	**	0.233	(0.08)	**	0.239	(0.08)	**
Other	0.126	(0.10)		0.127	(0.10)		0.087	(0.10)		0.090	(0.10)		0.105	(0.10)	
Size of student body	0.000	(0.00)		0.000	(0.00)		0.000	(0.00)		0.000	(0.00)		0.000	(0.00)	
Proportion student body:															
Black	-0.141	(0.09)	+	-0.142	(0.08)	+	-0.144	(0.09)		-0.145	(0.09)		-0.138	(0.09)	
In poverty	-0.386	(0.16)	*	-0.384	(0.16)	*	-0.374	(0.17)	*	-0.374	(0.17)	*	-0.356	(0.16)	*
In bilingual/ESL program	0.044	(0.11)		0.036	(0.11)		0.018	(0.11)		0.019	(0.11)		0.036	(0.11)	
In GT program	-0.029	(0.14)		-0.079	(0.13)		-0.053	(0.14)		-0.054	(0.14)		-0.038	(0.14)	
In special education	0.433	(0.22)	+	0.497	(0.22)	*	0.545	(0.22)	*	0.540	(0.22)	*	0.498	(0.22)	*
School is higher need	-0.100	(0.11)		0.014	(0.03)		0.020	(0.04)		0.020	(0.04)		0.010	(0.03)	
Proportion teachers on campus who received Strand II award	-0.200	(0.08)	*												
Constant	0.299	(0.14)	*	0.011	(0.11)		0.114	(0.09)		0.118	(0.09)		0.098	(0.09)	
Log of panel level var. component	0.141		***	0.138		***	0.152		***	0.152		***	0.149		***
Between-subject SD	0.456		***	0.454		***	0.451		***	0.451		***	0.452		***
Interclass correlation	0.088	-		0.085	-		0.102	-		0.102	-		0.098	-	
Core teachers (n)	2906			2906			2825			2825			2825		
Schools (n)	262			262			262			262			262		

Note: All of these models exclude core teachers ineligible for the 2009-10 ASPIRE program because of their job, campus, or opting out. Models 3-5 additionally exclude core teachers who did not receive any 2009-10 award. *** $p<0.001$, ** $p<0.01$, * $p<0.05$, +$p<0.10$.

APPENDIX 11: REGRESSION MODELS PREDICTING READING/ELA TEACHERS' MEAN STUDENT GAINS ON STANFORD READING TEST

Table 14, Part 1 of 3: Coefficients from Linear Regression Models Predicting Reading/ELA Teachers' Mean Student Gains on the Stanford Reading Test from 2009-10 to 2010-11

	Model 1			Model 2			Model 3			Model 4			Model 5		
	B	(SE)		B	(SE)		B	(SE)		B	(SE)		B	(SE)	
2009-10 ASPIRE Awards															
Received Strand I award	0.023	(0.04)		-0.002	(0.03)										
Received Strand II award	-0.050	(0.05)		0.114	(0.05)	*									
Received Strand III award	0.100	(0.08)		0.092	(0.05)	+									
Total award amount							0.006	(0.00)							
Total award amount as proportion of salary										0.284	(0.21)				
Total award amount as proportion of school's mean award amount													0.028	(0.01)	*
Interactions with 'Received Strand I award'															
School is higher need	0.011	(0.06)													
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Bilingual/ESL	-0.076	(0.04)	*												
Special education	-0.027	(0.05)													
Interactions with 'Received Strand II award'															
School is higher need	0.041	(0.04)													
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Bilingual/ESL	0.044	(0.04)													
Special education	-0.090	(0.06)													
Proportion teachers on campus who received Strand II award	0.134	(0.08)	+												

Table 14, Part 2 of 3: Coefficients from Linear Regression Models Predicting Reading/ELA Teachers' Mean Student Gains on the Stanford Reading Test from 2009-10 to 2010-11

	Model 1, cont.			Model 2, cont.			Model 3, cont.			Model 4, cont.			Model 5, cont.		
	B	(SE)		B	(SE)		B	(SE)		B	(SE)		B	(SE)	
Strand II award determined on basis of:															
Own students (grades 3-6, self-contained) (ref)				-											
Own students (grades 3-8, departmentalized)				-0.091	(0.05)	+									
Students in applicable subject(s) (grades 9-12)				-0.160	(0.09)	+									
3rd graders, reading and math (grades Pre-K-2)				-0.063	(0.06)										
Students in applicable subject(s) (special ed)				-0.318	(0.08)	***									
Interactions with 'Received Strand III award'															
School is higher need	-0.058	(0.10)													
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Bilingual/ESL	-0.082	(0.12)													
Special education	-0.032	(0.11)													
Teacher Characteristics															
Male	-0.035	(0.02)		-0.035	(0.02)		-0.043	(0.02)	+	-0.043	(0.02)	+	-0.042	(0.02)	+
Race:															
White, non-Hispanic (ref)	-			-			-			-			-		
Black, non-Hispanic	0.019	(0.02)		0.017	(0.02)		0.013	(0.02)		0.013	(0.02)		0.014	(0.02)	
White, Hispanic	-0.035	(0.03)		-0.045	(0.03)	+	-0.046	(0.03)	+	-0.046	(0.03)	+	-0.046	(0.03)	+
Asian	0.083	(0.04)	*	0.077	(0.04)	+	0.070	(0.04)	+	0.070	(0.04)	+	0.070	(0.04)	+
Other race	-0.085	(0.04)	*	-0.078	(0.04)	+	-0.079	(0.04)	+	-0.079	(0.04)	+	-0.077	(0.04)	+
Educational attainment:															
Less than Master's degree (ref)	-			-			-			-			-		
Master's degree or PhD	0.022	(0.02)		0.019	(0.02)		0.018	(0.02)		0.019	(0.02)		0.019	(0.02)	
Years of experience:															
3 or fewer	0.015	(0.02)		0.020	(0.02)		0.023	(0.02)		0.022	(0.02)		0.024	(0.02)	
4-10 (ref)	-			-			-			-			-		
11-25	-0.026	(0.02)		-0.029	(0.02)		-0.029	(0.02)		-0.027	(0.02)		-0.029	(0.02)	
More than 25	-0.039	(0.03)		-0.038	(0.03)		-0.035	(0.03)		-0.029	(0.03)		-0.035	(0.03)	
Hard to staff subjects:															
Not hard-to-staff subject (ref)	-														
Bilingual/ESL	0.048	(0.11)													
Special education	-0.016	(0.10)													

Table 14, Part 3 of 3: Coefficients from Linear Regression Models Predicting Reading/ELA Teachers' Mean Student Gains on the Stanford Reading Test from 2009-10 to 2010-11

	Model 1, cont.			Model 2, cont.			Model 3, cont.			Model 4, cont.			Model 5, cont.		
	B	(SE)		B	(SE)		B	(SE)		B	(SE)		B	(SE)	
Received a new hire recruitment s	-0.066	(0.02)	**	-0.090	(0.02)	***	-0.094	(0.02)	***	-0.094	(0.02)	***	-0.093	(0.02)	***
Strand II award determined on basis of:															
Own students (grades 3-6, self-contained) (ref)				-			-			-			-		
Own students (grades 3-8, departmentalized)				-0.023	(0.05)		-0.090	(0.03)	**	-0.090	(0.03)	**	-0.088	(0.03)	**
Students in applicable subject(s) (grades 9-12)				0.118	(0.10)		-0.026	(0.07)		-0.026	(0.07)		-0.019	(0.07)	
3rd graders, reading and math (grades Pre-K-2)				0.028	(0.05)		-0.022	(0.02)		-0.023	(0.02)		-0.013	(0.02)	
Students in applicable subject(s) (special ed)				0.086	(0.07)		-0.166	(0.04)	***	-0.166	(0.04)	***	-0.159	(0.04)	***
2008-09 CGI	0.037	(0.01)	**	0.039	(0.01)	**	0.034	(0.01)	*	0.035	(0.01)	**	0.033	(0.01)	*
School Characteristics															
Grade level of school:															
Elementary (ref)	-			-			-			-			-		
Middle school	0.006	(0.04)		0.047	(0.04)		0.031	(0.04)		0.031	(0.04)		0.040	(0.04)	
High school	-0.039	(0.06)		-0.079	(0.08)		-0.075	(0.08)		-0.075	(0.08)		-0.072	(0.08)	
Other	0.093	(0.10)		0.095	(0.10)		0.091	(0.10)		0.092	(0.10)		0.096	(0.10)	
Charter	-0.010	(0.30)		0.024	(0.30)		0.022	(0.30)		0.021	(0.30)		0.015	(0.30)	
Size of student body	0.000	(0.00)	+	0.000	(0.00)		0.000	(0.00)	+	0.000	(0.00)	+	0.000	(0.00)	
Proportion student body:															
Black	0.033	(0.09)		0.034	(0.09)		0.035	(0.09)		0.035	(0.09)		0.038	(0.09)	
In poverty	-0.601	(0.17)	***	-0.617	(0.16)	***	-0.634	(0.16)	***	-0.633	(0.16)	***	-0.628	(0.16)	***
In bilingual/ESL program	-0.080	(0.11)		-0.087	(0.11)		-0.106	(0.11)		-0.105	(0.11)		-0.100	(0.11)	
In GT program	-0.149	(0.14)		-0.143	(0.14)		-0.170	(0.14)		-0.169	(0.14)		-0.168	(0.14)	
In special education	-0.251	(0.22)		-0.251	(0.22)		-0.311	(0.22)		-0.312	(0.22)		-0.329	(0.22)	
School is higher need	0.031	(0.10)		0.009	(0.04)		0.011	(0.04)		0.011	(0.04)		0.007	(0.04)	
Proportion teachers on campus who received Strand II award	-0.146	(0.08)	+												
Constant	0.285	(0.13)	*	0.187	(0.11)	+	0.369	(0.09)	***	0.369	(0.09)	***	0.351	(0.09)	***
Log of panel level var. component	0.184		***	0.179		***	0.176		***	0.176		***	0.175		***
Between-subject SD	0.406		***	0.405		***	0.403		***	0.403		***	0.404		***
Interclass correlation	0.170	-		0.163	-		0.160	-		0.160	-		0.159	-	
Core teachers (n)	3145			3145			3056			3056			3056		
Schools (n)	263			263			262			262			262		

Note: All of these models exclude core teachers ineligible for the 2009-10 ASPIRE program because of their job, campus, or opting out. Models 3-5 additionally exclude core teachers who did not receive any 2009-10 award. *** p<0.001, ** p<0.01, * p<0.05, +p<0.10.