

# Timing Is Everything: Temporal Variation and Measures of School Quality

**WCER Working Paper No. 2015-4**  
**August 2015**

---

**Peter T. Goff, Jihye Kam, and Jacek Kraszewski**  
Department of Educational Leadership and Policy Analysis  
School of Education  
University of Wisconsin–Madison  
[pgoff@wisc.edu](mailto:pgoff@wisc.edu)



**Wisconsin Center for Education Research**

**School of Education • University of Wisconsin–Madison • <http://www.wcer.wisc.edu/>**

Goff, P. T., Kam, J., & Kraszewski, J. (2015). *Timing is everything: Temporal variation and measures of school quality* (WCER Working Paper No. 2015-4). Retrieved from University of Wisconsin–Madison, Wisconsin Center for Education Research website: <http://www.wcer.wisc.edu/publications/workingPapers/papers.php>

# **Timing Is Everything:**

## **Temporal Variation and Measures of School Quality**

**Peter T. Goff, Jihye Kam, and Jacek Kraszewski**

Surveys are a well-established tool used to expand our insight into the mechanics of organizations and the behaviors of people (Ary, Jacobs, Razavieh, & Sorensen, 2010). The analysis of the large-scale, nationally-representative datasets, such as those sponsored by the National Center for Education Statistics (NCES) and the National Science Foundation (NSF), assist in supporting and developing federal, state, and local legislations, regulations, and policies in education. Researchers collect, analyze, evaluate, and convey statistical data related to performance evaluations of students, parents, teachers, staffs, and principals. These national datasets, such as the Early Childhood Longitudinal Study (ECLS), the Schools and Staffing Survey (SASS), the Baccalaureate and Beyond (B&B), and the National Survey of College Graduates (NSCG), span early childhood education to postgraduate labor market outcomes. Practitioners increasingly use surveys to better understand organizational climate (e.g., Halverson, Kelley, & Shaw, 2014), leadership behaviors (e.g., Goldring, Porter, Murphy, Elliott, & Cravens, 2009), and instructional practices (e.g., Balch, 2012).

### **Perspectives and Perceptions**

In the current era of accountability, educators, practitioners, and parents rely, in part, on survey measures to inform summative evaluation, formative development, and school selection decisions. Therefore, the validity and reliability of survey responses regarding school performances are critical to the success of market-based educational reforms (Downey, Hippel, & Hughes, 2008). When the survey data are applied in educational practice, the temporal differences in the survey timing may generate problems. For example, schools using surveys to measure developmental progress typically examine how a score in the current year ( $t$ ) differs from a score in the prior year ( $t-1$ ). If the survey timing matters, the growth is over or underestimated and its true magnitude remains unknown. If fluctuation in timing is idiosyncratic, then this over or underestimation is simply measurement error. However, if the survey timing is correlated with substantive factors of interest, then temporal variation can be a source of bias, undermining decisions and inferences drawn from the data collected. Given the highly cyclic nature of schools and schooling, we might expect that temporal effects—to the extent that they exist at all—might be more pronounced in an educational setting.

In acknowledgement of the complex and multifaceted nature of education, some school districts have moved away from a strictly test-based vision of accountability, expanding performance measures to incorporate the views and perspectives of individuals closely involved. Polk County Public School District in Florida, for example, has developed a leadership evaluation system that integrates a measure of learning-centered leadership, as measured through the Vanderbilt Assessment of Leadership in Education (VALEd) (Porter et al., 2008). A similar system is found in Baltimore, Maryland. The performance measures regarding student and parent

## **Timing Is Everything**

perspectives on school climates, as well as teacher instructional effectiveness captured by the Measures of Effective Teaching Study (MET), are used in the Baltimore school leader effectiveness evaluation (Kane & Staiger, 2012). Given such prevalence of survey measures in education, these measures have to depict an accurate portrayal of differences across schools and over time.

### **Validity and Reliability**

The psychometric considerations of survey research are well developed and widely discussed in the literatures of psychology and are primarily clustered around the concepts of validity and reliability; however, there is little attention paid to the temporal effects of the survey timing on participants' responses. Validity refers to "the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of tests," while reliability refers to "the degree to which an assessment tool produces stable and consistent results" (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1999). In particular, the "internal consistency" or "item homogeneity," which denotes "the reliability of a scale based on the degree of within-scale item inter-correlation," has been long used in psychometric testing (Crobanch, 1951). One factor typically omitted from the psychometric testing pertains to the exogenous temporal variation, even though some psychometric papers directly address this feature (Buhyoff & Wellman, 1979; Kitamura, Shima, Sugawara, Toda, 1999; Cole & Stewart, 2002; Young, Blodgett, & Reardon, 2003; Murray, 2003).

Our study aims to contribute to the literature by examining how the survey timing in terms of day of week, season, and proximity to high-stakes exam is systematically related to teachers' responses regarding their perceptions of students' poverty and behavior, their control in the classroom, student behavior, and their principal's leadership support. The study is organized as follows: Section 1 presents the literature review, while Section 2 and Section 3 provide identification strategies and research methodology, respectively. Section 4 highlights our main findings and Section 5 discusses the findings and possible directions for further research.

### **1. Literature Review**

Several scholarly streams of thought in the literature focus on the instability of survey responses. Previous research on the instability of survey responses has empirically investigated specific factors such as survey mode and monetary rewards affecting survey responses while attempting to hold all other potential variables constant. We incorporate our hypothesis of the temporal variations in the perceptions of teachers with the methodological issues in the instability of survey responses. For simplicity, the related literatures are categorized under two general classifications: (1) antecedents and consequences of temporal variations, and (2) the seasonal variation in teacher perceptions.

### Antecedents and Consequences of Temporal Variations

Researchers have long recognized the psychological antecedents and consequences of human behaviors with respect to temporal variation. The literatures in psychology have shown distinct patterns in human behavior associated with temperature and other weather variables such as hours of sunlight, duration of daylight, humidity, and air pressure. Suicides, for example, are more likely to happen in spring and summer than in winter (Chew & McCleary, 1995; Ajdacic-Gross et al., 2006). This is in line with a positive effect of seasonal oscillations on crime rates. Warmer temperatures affect the rates of violent behaviors such as robbery (DeFronzo, 1984; Anderson & Anderson, 1984), homicide (Michael & Zumpe, 1983a, 1983b; McDowall & Curtis, 2014; Morken & Linaker, 2000), and sexual assault (Perry & Simpson, 1987; McLean, 2007). In particular, significant variations in violent incidents were observed May through June and October through November (Morken & Linaker, 2000). Violent behavior also peaked on Sundays and Mondays (Sisti, Rocchi, & Preti, 2012; Rastogi et al., 2013).

These temporal variations of crime revealed variations of crime opportunities (Carlsmith & Anderson, 1979) or emotional changes (Hipp, Bauer, Curran, & Bollen, 2004) related to personality disorders and moral responsibilities, which are defined as Seasonal Affective Disorder (SAD), which is a cyclic illness characterized by recurrent episodes of depression in fall and winter months alternating with periods of normal or mild moods in summer and spring months (Rosenthal, 1987, p. 57). There is also compelling evidence of a link between seasonal depression and seasonal variations in stock returns (Kamstra, Kramer, & Levi, 2003) and consumption expenditures (Barrow & McGranahan, 2000). Thus, various aspects of human behavior are significantly influenced by temporal variations.

Organizations also exhibit a predictable ebb and flow of behavior over time. Tax firms have substantial demands on their time and resources during the first half of the year and comparably more flexibility in the third quarter. Hotels, restaurants, and other organizations tied to tourism oscillate between periods of frenetic chaos and relative calm across days, weeks, and years (Baxter & King, 1999). Schools operate on a particularly well-known and established cycle: Annual schooling is regularly punctuated with breaks in the winter and spring, with a longer (typically 3 month) hiatus during the summer. The statewide standardized, high-stakes accountability tests that often inform and dictate school reform efforts are typically offered in early spring. Recent research has shown that effective human-resource management strategies in education are tied tightly to the school calendar (Drake et al., 2014).

The school cycle can dictate how a teacher copes with the demands of the job, and this may be particularly evident among new teachers (Weiss, 1999). During the early weeks of school, teachers establish routines and instructional norms, while also scheduling meetings with parents, administrators, and other teachers. As the school year progresses collaborative interactions may drop, instructional demands accumulate, and evaluations begin. In late winter and spring, teacher burnout increases (Brouwers & Tomic, 2000), in part because teachers cast doubt on their professional abilities, which is related with low self-efficacy (Fernet, Guay, Senécal, & Austin,

## Timing Is Everything

2012). At the end of the school year, teachers focus more on summative assessment and evaluation of student learning (Carson, 2006).

Thus, we see that both organizations and people are subject to systematic temporal fluctuations, and these fluctuations may be of particular note within schools.

### Seasonal Variation in Teacher Responses

A seasonal perspective on learning outcomes has been discussed since Hayns (1978, 1987) found a notably larger achievement gap during the summer months rather than the rest of school year. The author explained this gap by addressing non-school factors developed during the summer break. The subsequent literatures in the seasonality of learning outcomes have since expanded substantially but most have retained a narrow focus on summer learning gains or losses (Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996; Entwistle & Alexander, 1992, 1994; Alexandar, Entwisle, & Olson, 2001). We have found that inquiry regarding the temporal effects of teachers' perceptions of leadership support and student behaviors is comparatively limited. This study contributes to the literature by examining seasonal effects on the perceptions of teachers regarding students' poverty and behavior, their control in the classroom, and the principal leadership across the school year. This paper discusses survey-based research, practice, and policies by identifying and quantifying biases related to temporal variations of the timing of survey administration.

## 2. Identification Strategies

### Data Description

The SASS is administered to a sample of elementary and secondary schools representative of national and state levels. These surveys, sponsored by the Institute of Education Sciences, were first administrated in the 1987–88 school year and have been re-administered with minor changes six other times over the past 25 years. The survey questionnaires represent a wide range of topics from teacher demand, teacher and principal characteristics, general conditions in schools, principals' and teachers' perceptions of school climate and problems in their schools, teacher compensation, district hiring, and retention practices to basic characteristics of the student population. SASS information is obtained through four questionnaires: The school questionnaire, the teacher questionnaire, the principal questionnaire, and the school district questionnaire. Among them, we focus on four factors: teachers' perceptions of students' poverty, student behavior, teacher control in the classroom, and teachers' perceptions of leadership support using the responses from the teacher questionnaire. The sample of this study consists of 2007–08 SASS, which is the most recent available for analysis. It is composed of 38,240 teachers of 7,931 schools in 4,613 districts across the United States.<sup>1</sup> We restricted the sample to fulltime teachers.

---

<sup>1</sup> For a detailed description of the sampling design used in the SASS, see Tourkin et al. (2010).

## Timing Is Everything

The SASS administrators send out a framework questionnaire to schools once schools have been selected for sampling. The questionnaire asks schools for information on teachers in order to create a sampling frame for the selection of teachers within the school. The framework questionnaire is submitted to the SASS administrators whereas the teacher questionnaires are delivered to a volunteer survey coordinator for each school. The coordinator distributes the questionnaires to selected teachers. The teachers ensure anonymity and confidentiality of their information. The responses are submitted to the SASS administrators, including information regarding the beginning and completion dates of the survey.

### Conceptualizing Temporal Variation in Schools

Schools are organizations that maintain a regular and predictable schedule in accordance with the school district policy and procedures. The common academic year calendars are established for each semester while a detailed curriculum and instruction are designed on a weekly basis. The workload burdens on teachers vary across the days of week based on the intensity of the daily curriculum and instruction. In addition, the tempo and pulse of the school year from late August to early June provides an opportunity for researchers to study how perceptual measures relate to a specific season or time of year. For instance, teachers may be more likely to perceive student poverty as a serious problem in winter rather than in spring because the student poverty rates can be more easily revealed by students' clothing.

When exploring the role of temporal variations in survey measures, we identified three aspects of time that may be most relevant. First, we consider the role that day of the week may play, hypothesizing, for example, that responses on Fridays may be systematically different from responses collected on Tuesdays. Our second perspective of time looks across the entire school year from October<sup>2</sup> to June. Lastly, we consider temporal variations surrounding high-stakes exams, which typically take place during a 3-week window in the spring. This hypothesis demonstrates the challenges and stresses that teachers encounter when preparing their students for high-stakes exams (Kruger, Waddle, & Struzziero, 2007), suggesting that school supports, teacher perspectives, and organizational practices in the weeks leading up to the exams may differ from the weeks following the exam.

As we conceptualize variation in perceptual data over time we identify two sources of variation that may be of interest to researchers and policy makers: “substantive variation” (owing to systematic variation in the substantive construct of interest) and “endogenous variation” (owing to systematic variation in temporal factors related to respondent perception but unrelated to the underlying construct). As an example of the first, when examining student behavior over the course of the school year, we may find that students behave differently during the initial weeks of school as compared to later in the year after classroom norms and routines have been established. As an example of the second, we might envision a scenario where systematic,

---

<sup>2</sup> Although many schools begin in August or September, the survey responses do not begin until October. This gives teachers ample time to acclimate to the school, students, and peers prior to submitting any surveys.

## Timing Is Everything

external pressures (the approaching high-stakes state exams, for example) cause the teacher to perceive student behavior to be more (or less) orderly than the “true” measure of student behavior.

Two plausible hypotheses facilitate an inquiry into these temporal variations in schools. The first suggests that people’s perspectives are sensitive to personal, professional, or environmental influences. Under this hypothesis, we may expect to observe structural variations in survey responses over time, even when the underlying construct we are measuring remains constant. To investigate the prevalence of perceptual change within schools we selected two measures that we expect to remain constant over time: student poverty and teachers’ classroom control. The logic here is that any systematic changes in the perception of these factors are likely *not* related to the underlying construct, but rather are related to external, time-relevant factors.

A second hypothesis suggests that organizational factors within schools change over the course of the school year, and they do so in common, predictable ways. To investigate this second hypothesis we selected two measures that might vary systematically over time: teachers’ perceptions of student behavior and teachers’ perceptions of leadership support. The four measures we selected are discussed in greater detail below.

### Teachers’ Perceptions

**Student poverty.** This item captures the extent to which teachers perceive student poverty to be a problem in the school. The perceptions of teachers are coded as an ordered categorical variable with four responses: “serious problem” (1), “moderate problem” (2), “minor problem” (3), and “not a problem” (4). To easily interpret the results, the poverty scale has been recoded such that student poverty is regarded as a serious problem as the scale increases. Although student poverty varies between and within schools over multiple years, we have found no research to suggest *systematic* variation in student poverty *during* the school year. If student poverty is a time-invariant or a more flexible construct that varies only stochastically relative to the school year, any systematic variation in this measure can reasonably be attributed to perceptual variation. More detailed information on this measure is presented in Table A.3 of Appendix A.

**Control in the classroom.** To measure teachers’ control in the classroom, we use a set of items that solicited teachers’ perspectives of control over six processes: selecting textbooks and other instructional materials; selecting content, topics, and skills to be taught; selecting teaching techniques; evaluating and grading students; disciplining students; and determining the amount of homework to be assigned. The perceptions of teachers regarding their control in the classroom are coded as an ordered categorical variable with four responses: “no control” (1), “minor control” (2), “moderate control” (3), and “a great deal of control” (4). The Cronbach’s alpha is 0.72, which is on par with the accepted reliability standard of 0.70 (Hair, Anderson, Tatham, & Black, 1995; Nunnally, 1978). As with student poverty, we expect teacher control in the classroom to vary between schools (Weiss, 1993) as well as within schools over multiple years; however, we have little reason to believe these measures of classroom control vary within schools over a single school year. Table A.4 of Appendix A presents the related statistics.

## Timing Is Everything

**Leadership support.** We use three items to measure teachers' perceptions of leadership support. The perceptions of teachers on leadership support are coded as an ordered categorical variable with four responses: "strongly agree" (1), "somewhat agree" (2), "somewhat disagree" (3), and "strongly disagree" (4). To easily interpret the results, we recoded the scales in opposite direction. The Cronbach's alpha is 0.728. Teachers' perceptions of leadership support are tied to factors such as teacher efficacy and retention (Boyd et al, 2011). We hypothesize that leadership support is a construct that may evolve over the course of the school year as the demands on the principals' time changes and the principals' relationship with teachers, particularly newly hired teachers, develops. The relevant statistics are viewed in Table A.5 of Appendix A.

**Student behavior.** Our final perceptual measure pertains to teachers' perspectives of student behavior, asking teachers to document the extent to which tardiness, absenteeism, and class-cutting are problematic in their school. The perceptions of teachers regarding student behavior are coded as an ordered categorical variable with four responses: "serious problem" (1), "moderate problem" (2), "minor problem" (3), and "not a problem" (4). The scales are recoded in opposite direction for a consistent interpretation. The Cronbach's alpha is 0.839. Teachers' perceptions of student behavior in winter and summer are also significantly different from fall ( $p = 0.000 < 0.001$ ). In contrast, perceptions of student behavior are not significantly different before and after a high-stakes exam ( $p = 0.194 > 0.1$ ). While we were unable to document any prior work that empirically identifies temporal trends in student behavior, anecdotal evidence suggests that tardiness, absenteeism, and class-cutting are likely to become more prevalent as the school year progresses.

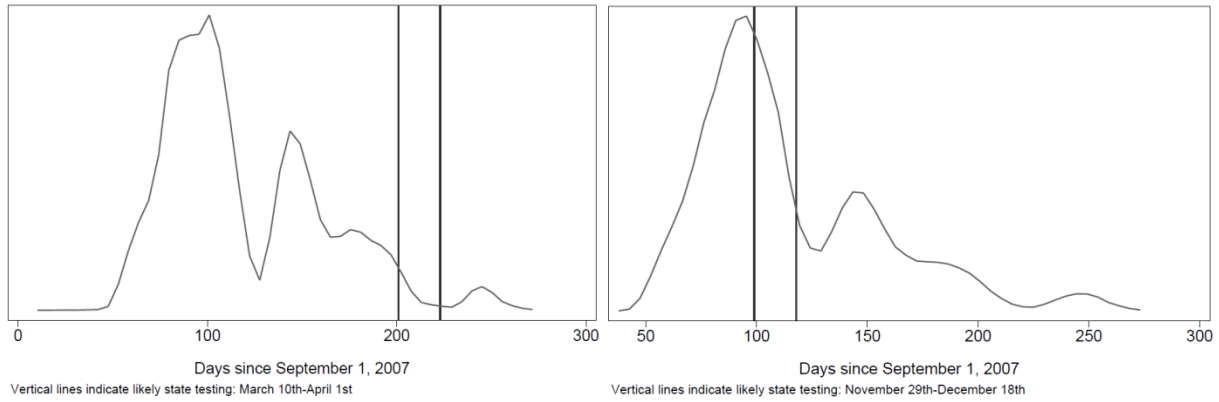
**Data verification.** To examine how survey measures differ before and after high-stakes exams, we searched the state testing dates for each state in the 2007–08 school year by utilizing the Internet Archives Wayback Machine. We restrict our subsample for the temporal variations in terms of the proximity to high-stakes exams to the state comprehensive assessments in Grades 2–8. That means high school teachers are excluded in our analysis. The dates of the testing windows obtained from each state's department of education websites were matched with information provided by the Council of Chief State School Officers (CCSS).

A large portion of responses were submitted between late October and December before the winter break in the entire sample, whereas the average testing window for spring exams is between March 10th and April 1st, 2008. We included only states offering a high-stakes exam during fall semester when the majority of SASS participants submitted their responses and excluded Oregon because of its wide range of testing window (Table A.2 of Appendix A). In the state test sample, the average testing window for fall exams is between November 29th and December 18th, 2007 (Figure 1).



## Timing Is Everything

**Figure 1. Distribution of Response Dates: Full Sample (Left) and Fall Testing (Right) Sample**



Note: A kernel density plot shows the distribution of teacher survey dates across the school year.

There are two important considerations to be aware of here. First, we were unable to find the state test dates in North Carolina and thus North Carolina was dropped from the model. Second, the state test dates in the 2007–08 school year for Indiana, Iowa, North Dakota, New Hampshire, Rhode Island, Utah, Virginia, and Wisconsin could not be ascertained, so we assigned the testing dates based on the test windows of these states in years before and after.

### 3. Methodology

This study addresses the effects of the survey timing on the teachers' perceptions of their students (behavior and poverty), their principal, and their classroom. Regression analysis is used to explore the relationship between temporal variation in responses with respect to day of week, season, and proximity to high-stakes exams and the perceptions of teachers regarding students' poverty and behavior, their classroom control, and the principal leadership. This section outlines the methods we applied to discuss three phenomena of temporal variations: variations pertaining to day of week, variations across the entire school year, and variations before and after a statewide exam. When examining each of these three aspects of temporal variations, we use the four perceptual measures previously outlined: student poverty, classroom control, leadership support, and student behavior. To facilitate interpretation of our findings, we standardize the scores for each teachers' perception to have within-sample mean 0 and variance 1 after taking mean values over the respective sets of scales (Heckman, Stixrud, & Urzua, 2006).

We use several aspects of the survey administration process to strengthen our identification strategy. To facilitate the data collection process, the SASS begins the survey administration in September and staggers the deployment of its remaining surveys across the school year. Although this staggered timing is not random, it is not purposefully related to factors such as geography, urbanicity, or school quality. Further variation in timing manifests within schools, as the schools are given surveys for all selected teachers within a school at the same time, yet some teachers take longer than others to complete the survey. We take advantage of both within and

## Timing Is Everything

between-school variation in the survey timing to estimate the impact of timing on survey responses.

When exploring the role of temporal variations across the school year in terms of day of week, season, and proximity to high-stakes exams, we use several approaches to mitigate biases that may result from endogenous relationships between response time and teacher characteristics. We first control for observable characteristics at the school and teacher levels, because teachers who promptly complete their survey may have systematically different views from those of their peers who wait to complete the survey. A second strategy for limiting endogenous variation focuses on an examination of temporal variation using only the teachers from each school who responded first. The logic of this approach is that “first responders” might be systematically different from their colleagues who respond later, and an examination of response patterns among first responders will yield a cleaner estimate of temporal variation. In a third approach, we use the date the first responder submitted the survey as an instrument for completion date for all other teachers within the school. This approach is based on the assumption that the completion date of the first responders will be related to the completion date of other teachers in the school, but unrelated to unobserved factors that may be correlated with later responses; first responders are not included in this analysis. To conduct instrumental variable (IV) estimations, the instruments should be uncorrelated with errors but partially correlated with the endogenous treatment variable once the other explanatory variables are controlled. The R-squared and the adjusted R-squared values of the first-stage regression are around 0.882 for estimations in the seasonal variations, ruling out potential concerns regarding weak instrument bias (Stock & Yogo, 2005). The F-statistics of for the seasonal variation IV analyses exceed 600, considerably larger than the minimum rule of thumb value of 10.

We use these various approaches to examine the four survey scales identified above, including two that are expected to exhibit substantive variation—leadership support and student behavior—and two that exhibit perceptual variation: perceptions of poverty and classroom control.

### **Analytical Approach**

As indicated, we examine three sources of temporal variation: day of week, season, and before and after high-stakes state exams. For each scenario, the dependent variables are four measures of teachers’ perceptions: their instructional challenges related to poverty, classroom control, leadership support, and student behavior (see Table A.1 in Appendix A). For many of our models we present results with and without school and teacher control variables. Models without control variables may be of interest to practitioners who use raw, unadjusted survey means, while researchers who frequently integrate control variables may have greater interest in the covariate adjusted regression estimates. School control variables include urbanicity dummies (city, suburb, town, and rural), total enrollment in school, program type dummies (regular, special program emphasis, special education, vocational education, and alternative program), the percentage of students with an individual education program (IEP) and limited English proficiency (LEP), charter school dummy, regional dummies (Northeast, Midwest, South, and

### Timing Is Everything

West), and school level dummies (primary, middle, high, and combined). Teacher control variables include Hispanic ethnic origin dummy, racial dummies (White, Black, Asian, Pacific Islander, and American Indian), total years of teaching experience and its squared, gender dummy, educational dummies for highest degree earned (associate's degree or no college degree, bachelor's degree, master's degree, educational specialist or certificate of advanced graduate studies, and doctoral degree), union dummy, and dummies for grades of students. A descriptive overview of the full sample and the restricted test sample are presented in Table A.7 and Table A.8 of Appendix A.

To analyze the impact of daily variation on teachers' perceptions, we estimate four equations of the form shown below (one for each of the teacher perception variables). We denote  $Y_{i,j}$  as the perceptions of teacher  $i$  in school  $j$ . The model for Scenario 1, day of the week, is specified below:

$$Y_{i,j} = \alpha_{i,j} + \beta_{i,j}Day\ of\ Week_{i,j} + \gamma_jSchool_j + \delta_iTeacher_i + \epsilon_{i,j} \quad (1)$$

where  $Day\ of\ Week_{i,j}$  are six dummy variables for day of week, with reference to Friday.  $School_j$  and  $Teacher_i$  are the school and teacher control variables, respectively, as outlined in the previous paragraph.

Similarly, the IV models for Scenario 2, investigating variation across the school year are below:

$$Z_{i,j} = \alpha_{i,j} + \beta_{i,j}Season_{i,j} + \gamma_jSchool_j + \delta_iTeacher_i + \rho_jFirst\ Response_j + \epsilon_{i,j} \quad (2)$$

$$Y_{i,j} = \alpha_{i,j} + \beta_{i,j}Season_{i,j} + \gamma_jSchool_j + \delta_iTeacher_i + \lambda_j\hat{Z}_j + \epsilon_{i,j}$$

where  $Season_{i,j}$  are dummies for spring and winter seasons, in reference to fall. These variables do not pertain to seasons directly but instead represent periods of time that may be most pertinent to schools: a fall period running from September through December, a winter period from January to March, and a spring period from March to June.

$Z_{i,j}$  is operationalized as the number of days from September 1, 2007; in the same way  $First\ Response_j$ , our excluded instrument, is the number of days since September first for the first respondent in each school. Thus, we are using the response time from the first responders in each school to create an exogenous estimate of the response time for all other individuals in the school.

The models for Scenario 3, investigating how proximity to state exams may influence survey response patterns, are shown below:

## Timing Is Everything

$$Y_{i,j} = \alpha_{i,j} + \beta_{i,j}Prior_{i,j} + \gamma_jSchool_j + \delta_iTeacher_i + \lambda_iTimePrior_i + \epsilon_{i,j} \quad (3)$$

where  $Prior_{i,j}$  is 1 if a teacher  $i$  responds to the survey before the state exam and 0 after the state exam. The variable  $TimePrior$  indicates the number of days before or after the testing window that a survey was completed. As with Scenario 1, we present results for a base model and a covariate-adjusted model.

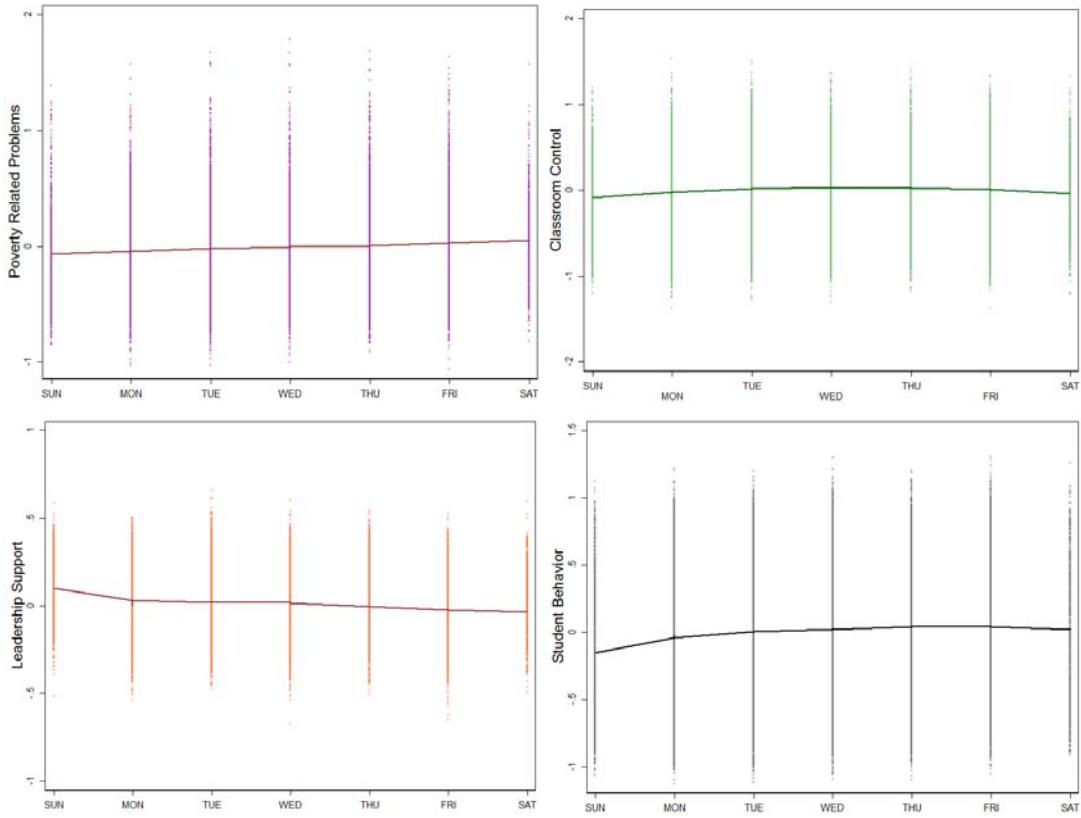
### 4. Findings

The findings from our regression analyses show that temporal variation may introduce unwanted biases into survey responses. We find evidence that teachers' perceptions are sensitive to temporal variation, even when the underlying construct of interest remains constant. We also find evidence that these constructs themselves—commonly used in formative constructs and increasingly used in high-stakes evaluations—show some systematic variation over time. This section outlines the evidence supporting these findings, drawing from the regression models outlined above.

#### Weekly Variation

Scenario 1 investigates whether the responding day of week affects the survey responses. Our findings are presented graphically in Figure 2, which depicts the predicted values from the covariate-adjusted model. When we examine the two measures used to examine perceptual changes (poverty and classroom control), we see little variation across days, although teachers were more likely to report poverty to be an instructional barrier on Mondays rather than on Fridays. When we examine the two measures intended to measure changes in constructs—leadership support and student behavior—we see that teachers are more likely to view principals as supportive on Sundays and Tuesdays (relative to Fridays), and student behavior is reported to be lower on Mondays. Full results are presented in Table B.1 of Appendix B. Similar trends among the coefficients are noted when the sample is reduced to only first responders (see Table B.2 of Appendix B); however, the 75% reduction in sample size erases much of the statistical significance. In sum, these results suggest that surveys completed on Mondays may yield lower ratings of perceptions regarding students, and surveys completed on Fridays may yield lower ratings of perceptions regarding school leadership.

Figure 2. Predicted Values for Day of the Week (Scenario 1)



### Seasonal Variation

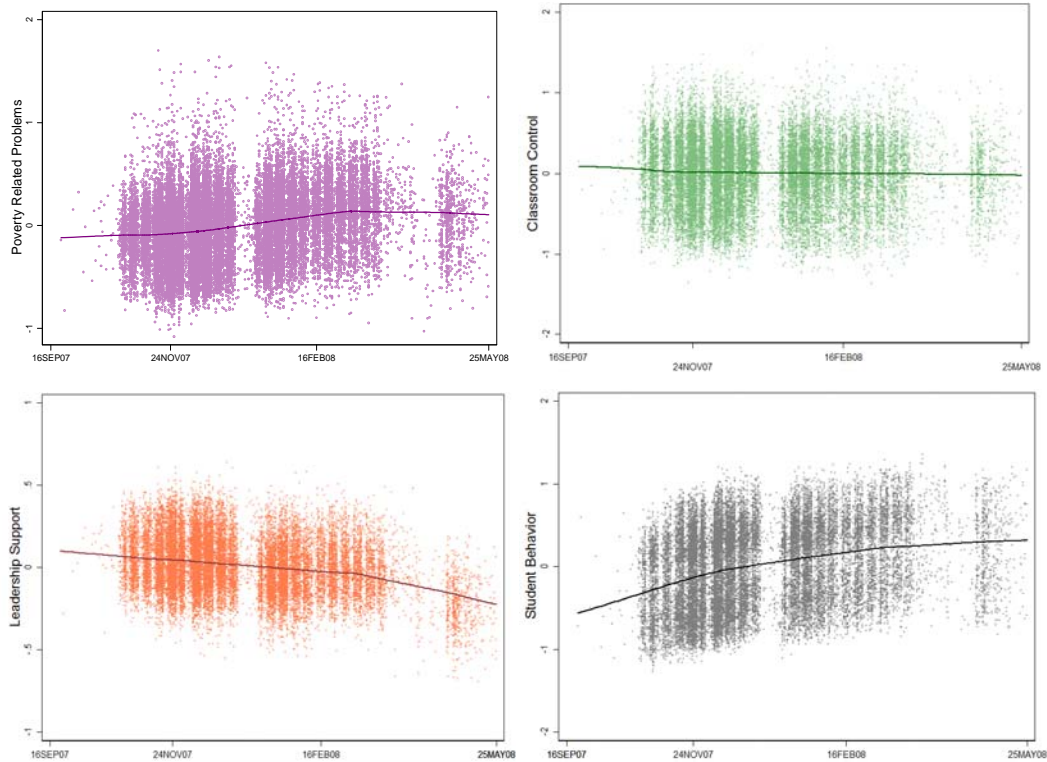
In the second set of our analyses, we examine variation in survey responses across the school year. The findings from the IV estimations of the covariate-adjusted models are presented in Figure 3. The top two panels represent the two constructs we hypothesized to be constant across the school year: teachers’ perceptions of poverty related problems and classroom control. The lower two panels represent the two constructs we predicted may vary across the school year. Teachers’ perceptions of instructional challenges related to poverty, as well as the aspects of classroom control, do not demonstrate significant change across the school year in any of the five models we specified (see Table B.3 in Appendix B). The consistency of predicted values for student poverty-related challenges and classroom control suggests that teachers’ perspectives are not strongly influenced by seasonality. This, in turn, implies that any measurable changes that occur across the school year are likely attributable to changes in the underlying construct of interest.

The two constructs in the bottom panel of Figure 3, leadership support and student behavior, show notable variation across the school year. Teachers’ perceptions of leadership support is greatest at the start of the school year, declines through March before dropping sharply at the end of the school year. This result is evident in the base model, the model with school and teacher covariates, and the model with covariates that limited the sample to first-responders. Although

## Timing Is Everything

the same pattern is evident graphically when using predicted values, temporal variation in leadership support is not significant in the IV models. Teachers' perceptions of student behavior problems surrounding truancy, tardiness, and absenteeism increase at a modest, though consistent, rate throughout the course of the school year. In particular, teachers are more likely to negatively perceive student behavior in spring season (Table B.6 of Appendix B). These findings are robust across all model specifications.

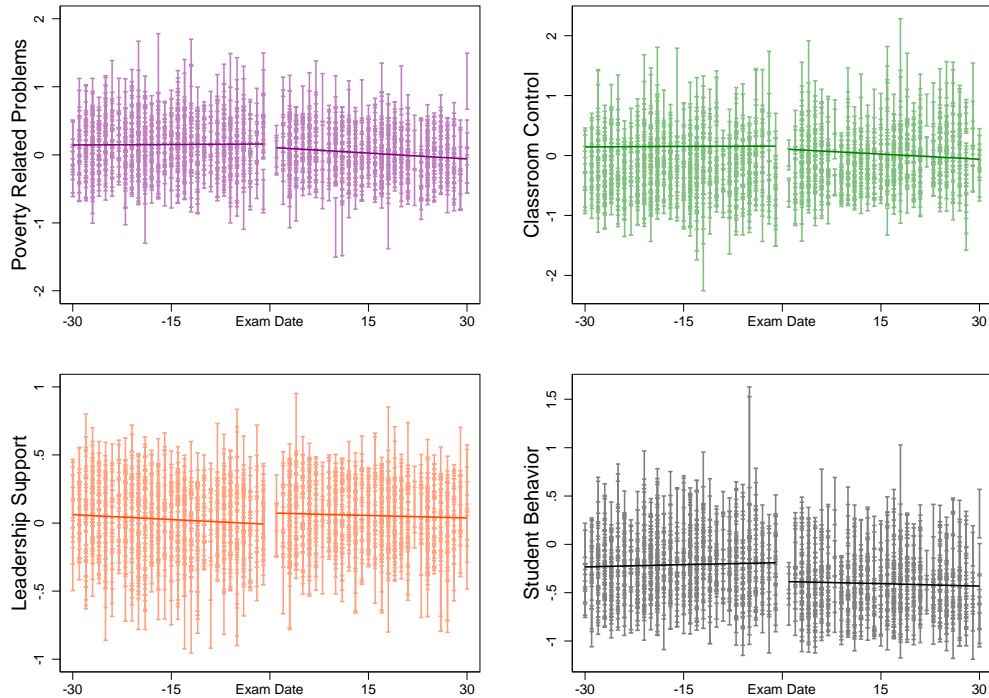
**Figure 3. Predicted Values across the School Year (Scenario 2)**



## Variation Before and After State Exams

Our final inquiry investigated the extent to which the pressures of high-stakes testing may substantially change perspectives and behaviors within schools. Figure 4 presents the min-max range of predicted values for each day generated from our covariate controlled models across each of our four survey measures. The x-axis shows the number of days before and after the exam period. For the three survey measures of student poverty-related challenges, leadership support, and student behavior, the covariate-controlled models yielded null results, suggesting that measures collected before or after testing are not significantly different. In contrast, teachers who responded to the survey prior to the high-stakes test are somewhat more likely to report greater classroom control (Table B.7 of Appendix B). These results are substantively unchanged when we limit the sample to first responders.

Figure 4. Predicted Values Pre- and Post-testing (Scenario 3)



## 5. Discussion and Further Research

The findings from this research are threefold. First, teachers’ report stronger leadership support and more challenging student behaviors early in the week. Second, substantive constructs appear to change predictably over the course of the school year, particularly declines in student behavior and, to a lesser extent, in leadership support. Third, despite the publicity given to the pressures arising around high-stakes state exams, we find little evidence that responses before the exam are systematically different than those before. This last point comes with the caveat that responses to low-stakes surveys, such as the SASS may differ from responses to high-stakes surveys.

For state and district-level practitioners these findings imply that reasonable latitude may be given to schools to choose survey dates that best complement local schedules, such as professional development days. Under such conditions, small timing differences in survey administration should not bias comparisons among schools.

Although this paper focused on the method of data collection rather than the survey constructs directly, school leaders may seek to modify their practice in response to the above findings. Specifically these findings suggest that teacher perceptions of leadership support lags as the year progresses. Perhaps this is in response to greater responsibilities for both teachers and principals later in the year. If this were true, principals may be more likely to attend to other demands and their support may flag. It may be that levels of principal support remain fairly

## Timing Is Everything

constant over time, yet teachers feel more stress later in the year and need principals to ramp up their support accordingly. The above findings may also highlight a shifting of the principals' role within an accountability framework. Early in the year principals may adopt the role of coach or mentor as they work with teachers to build instructional capacity. As the year progresses, principals' role may shift to that of evaluator as they engage with teachers in a more summative manner (e.g., formal observations of classroom teaching). This shift from coach to evaluator may push some teachers to see principals as less supportive during the spring than in the fall.

The decline in perceived leadership support may also arise as a result of the accumulation of stress and challenges encountered throughout the school year. With summer is traditionally viewed as a period of rest and renewal for teachers it is reasonable for teachers to begin the year positive and optimistic and to have this positivity wane when faced with mounting adversity. This may be the case even in the absence of any discernable differences in leadership behaviors. Whether the decline lies in actual changes in leadership support or is strictly an artifice of perception, the message for school leaders is clear: teachers need more support as the school year progresses. Focused support of new and novice teachers during late spring may be a particularly strong investment on the part of school leaders as these teachers are more likely to leave the school or profession.

Although this research was not designed to provide a comprehensive portrayal of all perceptions elicited by teachers, it does provide some direction to leaders as to where their support may be needed. As teachers report student behavior to be more problematic as the year progresses, principals can focus a portion of their efforts to maintain clear and consistent behavioral expectations for students. Especially in elementary schools, the establishment of behavioral norms and values is a priority early in the school year. Principals can support teachers by ensuring that this focus remains strong through the winter and spring months as well. As novice teachers report greater challenges with classroom management, school leaders may want to develop strategies to ensure early career teachers have the supports they need to limit behavioral challenges and focus on instruction. The research we have presented here suggests that it is important for principals to maintain these behavioral supports through the end of the school year.

Researchers engaged in survey research may also benefit from these findings. Following the watershed research on the summer learning loss, several studies have structured their research design to collect data including achievement testing at the beginning of the school year and conclude in the late spring, such as the Early Childhood Longitudinal Study, Kindergarten cohort. For designs with a treatment and control group, any seasonal changes would be equally manifest in the control condition and are not a concern. However, studies investigating a given policy or intervention that lack a control or comparison group may wrongly attribute fall-to-spring differences to program effects rather than seasonal change.

We see three limitations to our study. First, the findings apply only to the measures included here; other constructs operationalized through other measures may experience more or less



### **Timing Is Everything**

seasonal change. We see this possibility as a fruitful line of subsequent inquiry. Second, most teachers responded to surveys during the fall semester while most state exams were conducted in spring. This causes a notable reduction in sample size and subsequent reduction in precision. The last limitation arises because the testing window period varies across states, making it difficult to precisely identify teachers' perceptions before and after the state exam. In any future work, district-level exam dates should be integrated to better estimate how high-stakes exams may impact measures of climate, leadership, and efficacy.

## Timing Is Everything

### References

- Ajdacic-Gross, V., Lauber, C., Sansossio, R., Bopp, M., Eich, D., Gostynski, M., Gutzwiller, F., & Rössler, W. (2007). Seasonal associations between weather conditions and suicide: Evidence against a classic hypothesis. *American Journal of Epidemiology*, *165*(5) 561–569.
- Alexandar, K. L., Entwisle, D. R., & Olson, L. S. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis*, *23*(2), 171–191.
- Anderson, C. A., & Anderson, D. C. (1984). Ambient temperature and violent crime: Tests of the linear and curvilinear hypotheses. *Journal of Personality and Social Psychology*, *46*(1), 91–97.
- Ary, D., Jacobs, L., Razavieh, A., & Sorensen, C. (2010). *Introduction to research in education*. Belmont, CA: Wadsworth Cengage Learning.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association.
- Bagozzi, R. P. (1984). A prospectus for theory construction in marketing. *Journal of Marketing*, *48*(1), 11–29.
- Baker, B. D., Oluwole, J. O., & Green, P. C. (2013). The legal consequences of mandating high stakes decisions based on low quality information: Teacher evaluation in the race-to-the-top era. *Education Policy Analysis Archives*, *21*(5), Retrieved from <http://epaa.asu.edu/ojs/article/view/1298>.
- Balch, R. T. (2012). *The validation of a student survey on teacher practice* (Doctoral dissertation, Vanderbilt University).
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.
- Barber, L. K., Barnes, C. M., & Carlson, K. D. (2013). Random and systematic error effects of insomnia on survey behavior. *Organizational Research Methods*, *16*(4) 616–649.
- Baron, R. A. (1978). Aggression and heat: The “long, hot summer” revisited. In A. Baum, S. Valins, & J. Singer (Eds.), *Advances in environmental research* (Vol. 1). Hillsdale, NJ: Erlbaum.
- Barrow, L., & McGranahan, L. (2000). The effects of the earned income credit on the seasonality of household expenditures. *National Tax Journal*, *53*(4), 1211–1243.
- Baxter, M., & King, R. G. (1999). Measuring business cycles: Approximate band-pass filters for economic time series. *Review of economics and statistics*, *81*(4), 575–593.
- Bergman, P., McLaughlin, M., Bass, M., Pauly, E., & Zellman, G. (1977). Federal programs supporting educational change: Vol. VII: Factors affecting implementation and continuation (ERIC Document Reproduction Service No. 140 432). Santa Monica, CA: RAND.

## Timing Is Everything

- Boyd, D., Grossman, P., Ing, M., Lankford, H., Loeb, S., & Wyckoff, J. (2011). The influence of school administrators on teacher retention decisions. *American Educational Research Journal*, 48(2), 303–333.
- Brouwers, A., & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education*, 16, 239–253
- Buhyoff, G. J., & Wellman, J. D. (1979). Seasonality bias in landscape preference research, *Leisure Sciences: An Interdisciplinary Journal*, 2(2), 181–190.
- Capel, S. A. (1991). A longitudinal study of burnout in teachers. *British Journal of Educational Psychology*, 61, 36–45.
- Carlsmith, J. M., & Anderson, C. A. (1979). Ambient temperature and the occurrence of collective violence: A new analysis. *Journal of Personality and Social Psychology*, 37, 337–344.
- Carson, R. L. (2006). Exploring the episodic nature of teachers' emotions as it relates to teacher burnout (Unpublished doctoral dissertation). Purdue University, West Lafayette, IN.
- Chan, D. W. (2003). Hardiness and its role in the stress-burnout relationship among prospective Chinese teachers in Hong Kong. *Teaching and Teacher Education*, 19, 381–395.
- Chan, D. W. (2010). Teacher burnout revisited: Introducing positive intervention approaches based on gratitude and forgiveness. *Educational Research Journal*, 25(2), 165–186.
- Chew, K. S. Y., & McCleary, R. (1995). The spring peak in suicides: A cross-national analysis. *Social Science & Medicine*, 40(2), 223–230.
- Church, A. (1993). Estimating the effect of incentives on mail survey response rates: A meta-analysis. *The Public Opinion Quarterly*, 57, 62–79.
- Churchill, G. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 16(February), 64–73.
- Cole, D. N., & Stewart, W. P. (2002). Variability of user-based evaluative standards for backcountry encounters. *Leisure Sciences*, 24(3-4), 313–324.
- Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in web- or internet-based surveys. *Educational and Psychological Measurement*, 60(6), 821–836.
- Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, 66, 227–268.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
- Cunningham, W. G. (1983). Teacher burnout—Solutions for the 1980s: A review of the literature. *The Urban Review*, 15(1), 37–51.

## Timing Is Everything

- DeFronzo, J. (1984). Climate and crime: Tests of an FBI assumption. *Environmental and Behavior, 16*, 185–210.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The Tailored Design Method*. Hoboken, NJ: Wiley.
- Dillman, D. A. Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., & Messer, B. L. (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Social Science Research, 38*(1), 1–18.
- Downey, D., Hippel, P. Von, & Hughes, M. (2008). Are “failing” schools really failing? Using seasonal comparison to evaluate school effectiveness. *Sociology of Education, 81*(3), 242–270.
- Drake, Tl, Rubin, M., Neumerski, C., Goldring, E., Cannata, M., Grissom, J., & Schuermann, P. (2014). Timelines of talent management decisions and teacher effectiveness data availability. Working paper. Retrieved from <http://principaldatause.org/assets/files/timelines/Timeline-Overview-Report-201408.pdf>
- Entwisle, D. R., & Alexander, K. L. (1992). Summer setback: Race, poverty, school composition, and mathematics achievement in the first two years of school. *American Sociological Review, 57*, 72–84.
- Entwisle, D. R., & Alexander, K. L. (1994). Winter setback: School racial composition and learning to read. *American Sociological Review, 59*, 446–460.
- Feldman, J. M., & Lynch, J. G. (1988). Self-generated validity: Effects of measurement on belief, attitude, intention, and behavior. *Journal of Applied Psychology, 73*, 421–435.
- Fernet, C., Guay, F., Senécal, C., & Austin, S. (2012). Predicting intraindividual changes in teacher burnout: The role of perceived school environment and motivational factors. *Teaching and Teacher Education, 28*, 514–525.
- Foa, U. G., & Foa, E. B. (1974). *Societal structures of the mind*. Springfield, IL: Charles C Thomas.
- Foa, U. G., & Foa, E. B. (1980). Resource theory: Interpersonal behavior as exchange. In K. J. Gergen, M. S. Greenberg, & R. H. Willis (eds.), *Social exchange: Advances in theory and research*. New York, NY: Plenum.
- Goldring, E., Porter, A., Murphy, J., Elliott, S. N., & Cravens, X. (2009). Assessing learning-centered leadership: Connections to research, professional standards, and current practices. *Leadership and Policy in Schools, 8*(1), 1–36.
- Groves, R. M, Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology, Wiley Series in Survey Methodology*. Hoboken, NJ: Wiley & Sons.

## Timing Is Everything

- Groves, R. M., Singer, E., Corning, A. D. (2000). A leverage-saliency theory of survey participation: Description and illustration. *Public Opinion Quarterly*, 64, 299–308.
- Gulliksen, H. (1950). *Theory of mental tests*. New York, NY: Wiley.
- Gulliksen, H. (1964). The structure of individual differences in optimality judgments. In M. W. Shelly & G. L. Bryan (eds.), *Human judgments and optimality* (pp.72–85). New York, NY: Wiley.
- Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis* (3<sup>rd</sup> ed.). New York, NY: Macmillan Publishing Company.
- Halpern, S. D., Ubel, P. A., Berlin, J. A., & Asch, D. A. (2002). Randomized trial of \$5 versus \$10 monetary incentives, envelope size, and candy to increase physician response rates to mailed questionnaires. *Medical Care*, 40(9), 834–839.
- Halverson, R., Kelley, C., & Shaw, J. (2014). A CALL for improved school leadership. *Phi Delta Kappan*, 95(6), 57–60.
- Harris, I. A., Khoo, O. K., Young, J. M., Solomon, M. J., & Raea, H. (2008). Lottery incentives did not improve response rate to a mailed survey: A randomized controlled trial, *Journal of Clinical Epidemiology*, 61(6), 609–610.
- Hayns, S. N. (1978). *Principles of behavioral assessment*. New York, NY: Gardner Press.
- Heyns, B. (1987). Schooling and cognitive development: Is there a season for learning? *Child Development*, 58, 1151–1160.
- Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3), 412–482.
- Hipp, J. R., Bauer, D., Curran, P. J., & Bollen, K. A. (2004). Crimes of opportunity or crimes of emotion? Testing two explanations of seasonal change in crime. *Social Forces*, 82(4), 1333–1372.
- Hopkins, K. D., & Gullickson, A. R. (1992). Response rates in survey research: A meta-analysis of the effects of monetary gratuities. *Journal of Experimental Education*, 61, 52–62.
- James, J., & Bolstein, R. (1992). Large monetary incentives and their effect on mail survey response rates. *The Public Opinion Quarterly*, 56, 442–453.
- Kane, T. J., & Staiger, D. O. (2012). Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. Research Paper. MET Project. Bill & Melinda Gates Foundation.
- Kaplowitz, M. D., Hadlock, T. D., & Levine, R. (2004). A comparison of web and mail survey response rates. *The Public Opinion Quarterly*, 68(1), 94–101.

## Timing Is Everything

- Keating, N. K., Zaslavsky, A. M., Goldstein, J., West, D. W., & Ayanian, J. Z. (2008). Randomized trial of \$20 versus \$50 incentives to increase physician survey response rates. *Medical Care, 46*(8), 878–881.
- Kitamura, T., Shima, S., Sugawara, M., & Toda, M. A. (1999). Temporal variation of validity of self-rating questionnaires: Improved validity of repeated use of Zung's Self-Rating Depression Scale among women during the perinatal period. *Journal of Psychosomatic Obstetrics & Gynecology, 20*(2), 112–117.
- Knowles, E. S. (1988). Item context effects on personality scales: Measuring changes the measure. *Journal of Personality and Social Psychology, 55*, 312–320.
- Kruger, L. J., Wandle, C., & Struzziero, J. (2007). Coping with the stress of high stakes testing. *Journal of Applied School Psychology, 23*(2), 109–128.
- McDowall, D., & Curtis, K. M. (2014). Seasonal variation in homicide and assault across large U.S. cities. *Homocide Studies*. Retrieved from <http://hsx.sagepub.com/content/early/2014/05/30/1088767914536985>.
- McLean, I. (2007). Climatic effects on incidence of sexual assault. *Journal of Forensic and Legal Medicine, 14*(1). 16–19.
- Michael, R. P., & Zumpe, D. (1983a). Annual rhythms in human violence and sexual aggression in the United States and the role of temperature. *Social Biology, 30*, 263–278.
- Michael, R. P., & Zumpe, D. (1983b). Sexual violence in the United States and the role of season. *American Journal of Psychiatry, 140*, 883–886.
- Morken, G., & Linaker, O. M. (2000). Seasonal variation of violence in Norway. *The American Journal of Psychiatry, 157*(10), 1674–1678.
- Murray, G. (2003). The Seasonal Pattern Assessment Questionnaire as a measure of mood seasonality: A prospective validation study. *Psychiatry Research, 120*(1), 53–59.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.
- Perry, J. D., & Simpson, M. E. (1978). Violent crimes in a city: Environmental determinants, *Environment and Behavior, 19*, 77–90.
- Porter, A., Goldring, E., Stephen, E., Murphy, J., Polikoff, M., & Xiu, C. (2008). Setting performance standards for the VAL-ED: Assessment of principal leadership (ERIC Number: ED505799). Retrieved from ERIC website: <http://eric.ed.gov/?id=ED505799>.
- Porter, S. R., & Whitcomb, M. E. (2003). The impact of lottery incentives on student survey response rates. *Research in Higher Education, 44*(4), 389–407.
- Rastogi, A. K., Singh, B. K., Dadu, S. K., Thakur, P. S., Lanjewar, A. K., & Raput, P. P. (2013). Trends of homicidal deaths in Indore (M.P.) region: One year retrospective study. *Journal of the Indian Academy of Forensic Sciences, 35*(4), 343–345.
- Rosenthal, R. (1987). *Judgment studies*. New York, NY: Cambridge University Press.

## Timing Is Everything

- Simmons, C. J., Bickart, B. A., & Lynch, J. G. (1993). Capturing and creating public opinion in survey research. *Journal of Consumer Research*, 20(2), 316–329.
- Singer, E., van Hoewyk, J., & Masher, M. (2000). Experiments with incentives in telephone surveys. *The Public Opinion Quarterly*, 64, 171–188.
- Sisti, D., Rocchi, M. B. L., & Preti, A. P. (2012). The epidemiology of homicide in Italy by season, day of the week and time of day. *Medicine, Science, and the Law*, 52(1), 100–106.
- Stock, J., & Yogo, M. (2005). Asymptotic distributions of instrumental variables statistics with many instruments. In D. Andrews & J. Stock (Eds.), *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*. New York, NY: Cambridge University Press.
- Tourkin, S., Thomas, T., Swaim, N., Cox, S., Parmer, R., Jackson, B., Cole, C., & Zhang, B. (2010). *Documentation for the 2007–08 Schools and Staffing Survey* (NCES 2010-332). National Center for Education Statistics, Institute of Education Science, Washington, DC: U.S. Department of Education.
- Viswanathan, M. (2005). *Measurement error and research design*. Thousand Oaks, CA: Sage Publication, Inc.
- Walford, G., Tucker, E., & Viswanathan, M. (2010). *The SAGE handbook of measurement*. Thousand Oaks, CA: SAGE Publications Inc.
- Weisberg, H. F. (2005). *The Total Survey Error Approach: A guide to the new science of survey research*. Chicago, IL: The University of Chicago Press.
- Weiss, C. H. (1993). Shared decision making about what? A comparison of schools with and without teacher participation. *Teachers College Record*, 95(1), 69–92.
- Weiss, E. M. (1999). Perceived workplace conditions and first-year teachers' morale, career choice commitment, and planned retention: A secondary analysis. *Teaching and Teacher Education*, 15(8), 861-879.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817–838.
- Willimack, D., Schuman, H., Pennell, B., & Lepkowski, J. (1995). Effects of a prepaid nonmonetary incentive on response rates and response quality in a face-to-face survey. *The Public Opinion Quarterly*, 59, 78–92.
- Young, M. A., Blodgett, C., & Reardon, A. (2003). Measuring seasonality: Psychometric properties of the seasonal pattern assessment questionnaire and the inventory for seasonal variation. *Psychiatry Research*, 117(1), 75–83.

## Appendix A

**Table A.1. Dependent Variable Constructs**

Variable Name	Question	Response Options	Recode	Cronbach's alpha
Student Poverty	Q56. To what extent is each of the following a problem in this school? <b>h. Poverty</b>	(1) Serious problem (2) Moderate problem (3) Minor problem (4) Not a problem	<i>Yes</i>	-
Classroom Control	Q54. How much actual control do you have in your classroom at this school over the following areas of your planning a teaching? <b>a.</b> Selecting textbooks and other instructional materials <b>b.</b> Selecting content, topics, and skills to be taught. <b>c.</b> Selecting teaching techniques. <b>d.</b> Evaluating and grading students. <b>e.</b> Disciplining students. <b>f.</b> Determining the amount of homework to be assigned.	(1) No control (2) Minor control (3) Moderate (4) A great deal of control	No	0.722 (U); 0.745 (S)
Leadership Support	Q55. To what extent do you agree or disagree with each of the following statement? <b>g.</b> My principal enforces school rules for student conduct and backs me up when I need it. <b>j.</b> The principal knows what kind of school he or she wants and has communicated it to the staff. <b>l.</b> In this school, staff members are recognized for a job well done.	(1) Strongly agree (2) Somewhat agree (3) Somewhat disagree (4) Strongly disagree	<i>Yes</i>	0.728 (U); 0.839 (S)
Student Behavior	Q56. To what extent is each of the following a problem in this school? <b>a.</b> Student tardiness <b>b.</b> Student absenteeism <b>c.</b> Student class cutting	(1) Serious problem (2) Moderate problem (3) Minor problem (4) Not a problem	<i>Yes</i>	0.727 (U); 0.728 (S)

Notes: U and S refer to unstandardized and standardized alphas. We standardized the response scales to have within-sample mean 0 and variance 1 after taking averages over the respective sets of scales.



**Table A.2. State Testing Windows**

<b>State</b>	<b>Begin</b>	<b>End</b>	<b>Name</b>	<b>State Test Sample</b>
AL	March 31, 2008	April 11, 2008	Alabama Reading and Mathematics Test	No
AK	March 31, 2008	April 14, 2008	Standards Based Assessment	No
AZ	April 7, 2008	April 18, 2008	Arizona Instrument to Measure Standards (AIMS)	No
AR	April 14, 2008	April 18, 2008	Benchmark Exams	No
CA	April 15, 2008	April 28, 2008	California Achievement Test CAT/6	No
CO	March 10, 2008	April 11, 2008	Colorado Student Assessment Program	No
CT	March 3, 2008	March 31, 2008	Connecticut Mastery/Connecticut Academic Performance Test	No
DE	March 5, 2008	March 14, 2008	Delaware Student Testing Program	No
DC	April 22, 2008	May 2, 2008	Stanford Achievement Test	No
FL	March 12, 2008	March 25, 2008	Florida Comprehensive Assessment Test	No
GA	April 2, 2008	May 2, 2008	Criterion Referenced Competency Tests	No
HI	March 31, 2008	April 18, 2008	Hawaii Content and Performance Standards II (HCPS)	No
ID	April 14, 2008	May 16, 2008	Idaho Standards Achievement Tests	No
IL	March 3, 2008	March 14, 2008	Illinois Standards Achievement Test	No
IN	March 3, 2008	March 13, 2008	Indiana Statewide Testing for Educational Progress-Plus (ISTEP)	No
IA	October 22, 2007	October 26, 2007	Iowa Tests of Basic Skills (ITBS)	<b>Yes</b>
KS	February 18, 2008	April 14, 2008	Kansas State Assessment	No
KY	April 21, 2008	May 2, 2008	Kentucky Core Content Test	No
LA	March 10, 2008	March 14, 2008	Louisiana Educational Assessment Program (LEAP)	No
ME	March 3, 2008	March 21, 2008	Maine Educational Assessment	No
MD	April 1, 2008	April 10, 2008	Maryland School Assessment	No
MA	March 24, 2008	April 4, 2008	Massachusetts Comprehensive Assessment System	No
MI	October 8, 2007	October 26, 2007	Michigan Educational Assessment Program	<b>Yes</b>
MN	April 14, 2008	May 2, 2008	Minnesota Comprehensive Assessments - Series II	No
MS	May 13, 2008	May 14, 2008	Mississippi Curriculum Tests 2	No
MO	March 31, 2008	April 25, 2008	Missouri Assessment Program	No
MT	March 3, 2008	March 26, 2008	Criterion-Referenced Tests (CRT)	No
NE	January 28, 2008	February 8, 2008	STARS Writing Assessments in grades	No
NV	January 22, 2008	February 22, 2008	Iowa Tests of Basic Skills/Iowa Tests of Educational	No

**Timing Is Everything**

<b>State</b>	<b>Begin</b>	<b>End</b>	<b>Name</b>	<b>State Test Sample</b>
			Development	
NH	May 5, 2008	May 22, 2008	New England Common Assessment	No
NJ	March 10, 2008	May 8, 2008	New Jersey Skills and Knowledge Assessment (NJ ASK)	No
NM	February 25, 2008	March 21, 2008	The New Mexico Standards Based Assessment	No
NY	March 3, 2008	March 12, 2008	Mathematics Assessment Tests	No
ND	October 22, 2007	November 9, 2007	North Dakota State Assessment	<b>Yes</b>
OH	April 21, 2008	May 9, 2008	Ohio Achievement Test	No
OK	April 10, 2008	April 25, 2008	Oklahoma Core Curriculum Tests	No
OR	October 15, 2007	May 23, 2008	Knowledge and Skills Tests	No
PA	March 31, 2008	April 11, 2008	Pennsylvania System of School Assessment	No
RI	October 1, 2007	October 23, 2007	New England Common Assessment	<b>Yes</b>
SC	May 13, 2008	May 22, 2008	Palmetto Achievement Challenge Test	No
SD	April 2, 2008	April 20, 2008	Dakota State Testing of Educational Progress	No
TN	March 31, 2008	April 23, 2008	Tennessee Comprehensive Assessment Program Achievement Test	No
TX	March 3, 2008	March 8, 2008	Texas Assessment of Knowledge and Skills	No
UT	September 18, 2008	October 5, 2008	Iowa Tests (/Utah Core Curriculum)	<b>Yes</b>
VT	October 2, 2007	October 19, 2007	The New England Common Assessment Program	<b>Yes</b>
VA	April 14, 2008	June 13, 2008	SOL Multiple Choice	No
WA	April 14, 2008	May 2, 2008	Washington Assessment of Student Learning	No
WV	May 12, 2008	May 16, 2008	West Virginia Educational Standards Tests	No
WI	October 22, 2007	November 23, 2007	Wisconsin Knowledge and Concepts Examinations	<b>Yes</b>
WY	March 12, 2008	April 16, 2008	Proficiency Assessments for Wyoming Schools (PAWS)	No

Notes: States offering a large-scale test during fall semester were included in our State Test Sample. Oregon was excluded because of a wide range of testing window period.

**Table A.3. Descriptive Statistics: Scale Means for Poverty Related Challenges by Day, Season, and Testing Period**

Temporal variation	Average teachers' perceptions: Student poverty						
<i>Full Sample:</i>							
Day of the week	Sun	Mon	Tue	Wed	Thu	Fri	Sat
	-0.121	-0.059	-0.025	-0.038	-0.038	-0.006	-0.008
	(0.054)	(0.031)	(0.029)	(0.033)	(0.034)	(0.031)	(0.065)
Season	Fall	Winter	Spring				
	-0.100	0.061	0.015				
	(0.024)	(0.022)	(0.084)				
Pre- & post-testing	Prior	Post					
	-0.051	-0.067					
	(0.021)	(0.044)					
<i>State Test Sample:</i>							
Day of the week	Sun	Mon	Tue	Wed	Thu	Fri	Sat
	-0.141	-0.197	-0.142	-0.151	-0.053	-0.111	-0.134
	(0.138)	(0.113)	(0.096)	(0.087)	(0.105)	(0.098)	(0.156)
Season	Fall	Winter	Spring				
	-0.135	-0.128	-0.237				
	(0.061)	(0.103)	(0.233)				
Pre- & post-testing	Prior	Post					
	-0.447	-0.091					
	(0.135)	(0.057)					

Note: All data weighted using adjusted the balanced repeated replication (BRR) weights and the standard deviation are presented in parentheses.

**Timing Is Everything**

**Table A.4. Descriptive Statistics: Scale Means for Classroom Control by Day, Season, and Testing Period**

<b>Temporal variation</b>	<b>Average teachers' perceptions: Classroom control</b>						
<i>Full Sample:</i>							
Day of the week	Sun	Mon	Tue	Wed	Thu	Fri	Sat
	-0.282	-0.230	-0.180	-0.226	-0.192	-0.204	-0.253
	(0.058)	(0.029)	(0.023)	(0.035)	(0.031)	(0.025)	(0.070)
Season	Fall	Winter	Spring				
	-0.196	-0.240	-0.249				
	(0.018)	(0.028)	(0.070)				
Pre- & post-testing	Prior	Post					
	-0.231	-0.044					
	(0.018)	(0.034)					
<i>State Test Sample:</i>							
Day of the week	Sun	Mon	Tue	Wed	Thu	Fri	Sat
	-0.189	-0.203	-0.115	-0.041	0.011	-0.172	-0.029
	(0.115)	(0.073)	(0.067)	(0.076)	(0.095)	(0.083)	(0.149)
Season	Fall	Winter	Spring				
	-0.104	-0.118	-0.214				
	(0.043)	(0.077)	(0.201)				
Pre- & post-testing	Prior	Post					
	-0.220	-0.124					
	(0.139)	(0.041)					

Note: All data weighted using adjusted the balanced repeated replication (BRR) weights and the standard deviation are presented in parentheses.

**Timing Is Everything**

**Table A.5. Descriptive Statistics: Scale Means for Leadership Support by Day, Season, and Testing Period**

<b>Temporal variation</b>	<b>Average teachers' perceptions: Leadership support</b>						
<i>Full Sample:</i>							
Day of the week	Sun	Mon	Tue	Wed	Thu	Fri	Sat
	0.140	0.045	0.087	0.036	0.060	0.014	0.057
	(0.044)	(0.025)	(0.023)	(0.025)	(0.027)	(0.032)	(0.054)
Season	Fall	Winter	Spring				
	0.093	0.013	-0.222				
	(0.015)	(0.019)	(0.067)				
Pre- & post-testing	Prior	Post					
	0.078	-0.082					
	(0.013)	(0.040)					
<i>State Test Sample:</i>							
Day of the week	Sun	Mon	Tue	Wed	Thu	Fri	Sat
	0.016	0.075	0.064	0.157	0.051	0.011	-0.132
	(0.156)	(0.083)	(0.070)	(0.066)	(0.105)	(0.099)	(0.173)
Season	Fall	Winter	Spring				
	0.107	-0.024	-0.371				
	(0.045)	(0.083)	(0.332)				
Pre- & post-testing	Prior	Post					
	0.369	0.031					
	(0.074)	(0.055)					

Note: All data weighted using adjusted the balanced repeated replication (BRR) weights and the standard deviation are presented in parentheses.

**Timing Is Everything**

**Table A.6. Descriptive Statistics: Scale Means for Student Behavior Problem by Day, Season, and Testing Period**

<b>Temporal variation</b>	<b>Average teachers' perceptions: Student behavior</b>						
<i>Full Sample:</i>							
Day of the week	Sun	Mon	Tue	Wed	Thu	Fri	Sat
	-0.334	-0.238	-0.185	-0.167	-0.197	-0.153	-0.236
	(0.040)	(0.023)	(0.022)	(0.023)	(0.026)	(0.026)	(0.069)
Season	Fall	Winter	Spring				
	-0.290	-0.059	-0.048				
	(0.016)	(0.024)	(0.086)				
Pre- & post-testing	Prior	Post					
	-0.209	-0.183					
	(0.015)	(0.041)					
<i>State Test Sample:</i>							
Day of the week	Sun	Mon	Tue	Wed	Thu	Fri	Sat
	-0.595	-0.594	-0.583	-0.575	-0.462	-0.427	-0.635
	(0.108)	(0.074)	(0.066)	(0.068)	(0.085)	(0.066)	(0.144)
Season	Fall	Winter	Spring				
	-0.565	-0.468	-0.848				
	(0.044)	(0.067)	(0.241)				
Pre- & post-testing	Prior	Post					
	-0.400	-0.545					
	(0.060)	(0.042)					

Note: All data weighted using adjusted the balanced repeated replication (BRR) weights and the standard deviation are presented in parentheses.

**Table A.7. Descriptive Statistics of Survey Measures: Full Sample and State Test Sample**

	<i>Full Sample</i>						<i>State Test Sample</i>					
	Mean	<i>Percentiles</i>					Mean					
		10%	25%	50%	75%	90%		10%	25%	50%	75%	90%
<i>Entire Sample:</i>												
- Student poverty	-0.038 (0.018)	-1.708	-0.665	0.379	0.379	1.422	-0.136 (0.051)	-1.708	-0.665	-0.665	0.379	1.422
- Classroom control	-0.213 (0.017)	-1.731	-0.748	-0.092	0.564	1.219	-0.110 (0.038)	-1.404	-0.748	-0.092	0.564	1.219
- Leadership support	0.056 (0.012)	-1.506	-0.472	0.046	1.081	1.081	0.060 (0.047)	-1.506	-0.472	0.563	1.081	1.081
- Student behavior	-0.200 (0.013)	-1.623	-0.763	-0.333	0.528	0.958	-0.547 (0.037)	-1.623	-1.193	-0.763	0.097	0.528
<i>First Respondents:</i>												
- Student poverty	-0.020 (0.024)	-1.708	-0.665	0.379	0.379	1.422	-0.130 (.0560)	-1.708	-0.665	-0.665	0.379	1.422
- Classroom control	-0.266 (0.025)	-1.731	-1.075	-0.092	0.564	1.219	-0.137 (.0493)	-1.404	0.748	-0.092	0.564	1.219
- Leadership support	0.064 (0.018)	-1.506	-0.472	0.046	1.081	1.081	0.098 (.0620)	-1.506	-0.472	0.563	1.081	1.081
- Student behavior	-0.294 (0.021)	-1.623	-0.763	-0.333	0.097	0.958	-0.568 (.0472)	-1.623	-1.193	-0.763	-0.333	0.528

Note: All data weighted using adjusted the balanced repeated replication (BRR) weights and the standard deviation are presented in parentheses.

Timing Is Everything

**Table A.8. Descriptive Statistics—Sample Means Calculated Using Teacher Weights**

	Full Sample		State-Test Sample	
	Entire Sample	First Respondents	Entire Sample	First Respondents
No. observations	33,258	10,548	2,102	831
<i>School Factors</i>				
Average total enrollment	845.09 (13.79)	713.53 (12.22)	473.77 (15.53)	460.10 (17.67)
Average percentage of teacher's students with an IEP (%)	15.44 (0.41)	15.86 (0.66)	15.17 (0.78)	16.25 (1.17)
Average percentage of teacher's students with an LEP (%)	8.42 (0.30)	8.80 (0.55)	4.73 (0.54)	5.29 (0.93)
Urbanicity (%):				
- City	26.16	26.05	21.52	21.34
- Suburb	34.68	34.02	31.77	32.40
- Town	13.99	13.99	15.35	15.71
- Rural	25.16	25.93	31.35	30.55
Region (%):				
- Northeast	19.46	20.03	8.31	8.70
- Midwest	22.40	21.75	83.21	82.49
- South	39.53	38.78	0.00	0.00
- West	18.62	19.44	8.47	8.81
Charter School (%):	2.08	2.07	5.26	4.56
School level (%):				
- Primary school	46.65	55.28	64.78	69.47
- Middle school	18.84	18.95	24.62	22.02
- High school	29.80	21.46	0.00	0.00
- Combined school	4.71	4.30	10.60	8.51
Program type (%):				
- Regular	93.52	93.44	96.93	97.06
- Special program emphasis	2.90	2.82	1.54	1.39
- Special education	0.80	0.88	0.73	0.94
- Career/technical/vocational education	1.07	0.81	0.00	0.00
- Alternative	1.72	2.05	0.80	0.61



### Timing Is Everything

	Full Sample		State-Test Sample	
	Entire Sample	First Respondents	Entire Sample	First Respondents
<i>Teacher Factors</i>				
Female (%)	75.32	77.74	82.18	85.08
Union (%)	76.58	76.47	88.64	88.72
Average years of teaching experience	13.52 (0.16)	13.94 (0.21)	14.37 (0.47)	14.42 (0.74)
Ethnicity (%):				
- Hispanic ethnic origin	7.31	6.51	2.30	2.04
Race (%):				
- White	90.31	91.30	97.34	97.95
- Black/African American	7.68	6.51	2.23	2.02
- Asian	1.56	1.61	0.47	0.33
- Native Hawaiian/Pacific Islander	0.33	0.30	0.00	0.00
- American Indian/Alaska Native	1.20	1.23	0.67	0.97
Highest degree earned (%)				
- Associate's	0.76	0.54	0.35	0.38
- Bachelor's	48.08	47.21	47.52	43.54
- Master's	44.24	45.64	47.15	51.66
- Education specialist	6.08	5.90	4.71	4.07
- Doctoral or professional	0.84	0.70	0.27	0.34
Currently teaching grades of students (%)				
- Pre-kindergarten	2.01	2.20	4.38	3.98
- Kindergarten	12.81	14.47	19.86	16.97
- 1st	15.02	16.35	21.55	21.38
- 2nd	14.96	16.36	21.00	19.99
- 3rd	15.32	17.31	21.49	20.40
- 4th	14.85	16.53	20.24	18.43
- 5th	14.71	16.90	21.47	21.03
- 6th	15.84	16.92	24.59	21.87
- 7th	16.81	16.47	21.87	19.79
- 8th	17.11	17.02	22.18	17.28
- 9th	23.67	17.73	6.35	4.20
- 10th	26.75	19.71	6.46	4.38
- 11th	27.14	19.86	6.78	4.65
- 12th	26.15	19.23	6.71	4.65
- Ungraded	3.38	2.82	1.56	1.79

Note: All data weighted using adjusted the balanced repeated replication (BRR) weights and the standard deviation are presented in parentheses.

## Timing Is Everything

### Appendix B. Regression Outcomes

**Table B.1. Variation across Days of Week**

	Poverty Challenges		Classroom Control		Leadership Support		Student Behavior	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Sunday	-0.115+	-0.061	-0.078	-0.007	0.126*	0.117+	-0.181***	-0.065
	(0.063)	(0.054)	(0.064)	(0.056)	(0.053)	(0.060)	(0.047)	(0.042)
Monday	-0.053	-0.084*	-0.026	-0.040	0.031	0.042	-0.085*	-0.067*
	(0.040)	(0.037)	(0.037)	(0.032)	(0.039)	(0.041)	(0.032)	(0.031)
Tuesday	-0.019	-0.029	0.024	0.011	0.073+	0.088*	-0.032	-0.042
	(0.040)	(0.041)	(0.032)	(0.036)	(0.039)	(0.041)	(0.032)	(0.028)
Wednesday	-0.032	-0.054	-0.022	-0.007	0.021	0.041	-0.014	-0.040
	(0.043)	(0.043)	(0.043)	(0.036)	(0.038)	(0.039)	(0.034)	(0.035)
Thursday	-0.032	-0.030	0.012	0.007	0.045	0.075	-0.044	-0.052
	(0.043)	(0.045)	(0.041)	(0.042)	(0.045)	(0.046)	(0.036)	(0.032)
Saturday	-0.002	0.034	-0.049	0.005	0.042	0.037	-0.084	0.015
	(0.068)	(0.063)	(0.071)	(0.063)	(0.060)	(0.067)	(0.068)	(0.067)
School controls	No	Yes	No	Yes	No	Yes	No	Yes
Teacher controls	No	Yes	No	Yes	No	Yes	No	Yes
R-squared	0.001	0.118	0.001	0.164	0.001	0.030	0.002	0.256
No. of observations	33,258	29,260	33,258	29,260	33,258	29,260	33,258	29,260

Notes: The outcomes of OLS regression models are reported. The endogenous explanatory variables for school and teacher characteristics are included where indicated but not reported above. The standard errors are in parentheses. The model is fitted using the BRR weights. The symbols +, \*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, 1%, 0.1% levels, respectively.

## Timing Is Everything

**Table B.2. Variation across Days of Week: First and Later Respondents**

	Poverty Challenges		Classroom Control		Leadership Support		Student Behavior	
	First (1)	Later (2)	First (1)	Later (2)	First (1)	Later (2)	First (1)	Later (2)
Sunday	-0.061 (0.116)	-0.059 (0.061)	-0.046 (0.111)	0.006 (0.061)	0.266* (0.121)	0.062 (0.082)	-0.085 (0.091)	-0.058 (0.047)
Monday	-0.105 (0.082)	-0.075 (0.051)	-0.052 (0.063)	-0.031 (0.038)	0.060 (0.067)	0.039 (0.054)	-0.103 (0.064)	-0.051 (0.032)
Tuesday	-0.016 (0.081)	-0.044 (0.053)	0.050 (0.066)	-0.014 (0.046)	0.133+ (0.073)	0.068 (0.053)	-0.077 (0.058)	-0.023 (0.036)
Wednesday	-0.040 (0.082)	-0.062 (0.053)	-0.019 (0.067)	0.005 (0.047)	0.063 (0.080)	0.033 (0.057)	-0.062 (0.064)	-0.028 (0.038)
Thursday	-0.017 (0.075)	-0.040 (0.055)	-0.025 (0.083)	0.024 (0.049)	0.177* (0.083)	0.023 (0.064)	-0.076 (0.069)	-0.041 (0.033)
Saturday	0.103 (0.127)	0.001 (0.079)	0.072 (0.129)	-0.029 (0.076)	0.123 (0.119)	-0.009 (0.089)	-0.103 (0.113)	0.070 (0.073)
School controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.121	0.123	0.177	0.162	0.039	0.030	0.233	0.265
No. of observations	9,153	20,107	9,153	20,107	9,153	20,107	9,153	20,107

Notes: The outcomes of OLS regression models are reported. The endogenous explanatory variables for school and teacher characteristics are included where indicated but not reported above. The standard errors are in parentheses. The model is fitted using the BRR weights. The symbols +, \*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, 1%, 0.1% levels, respectively.

**Timing Is Everything**

**Table B.3. Poverty Related Challenges over the School Year**

	OLS	OLS	OLS	OLS	IV	IV
Sample	All	All	First	Later	All	All
	(1)	(2)	(3)	(4)	(5)	(6)
Days	0.001 (0.001)	0.001 (0.001)	0.002 (0.001)	0.001 (0.001)	0.002 (0.002)	0.002 (0.002)
Winter	0.062 (0.077)	0.078 (0.069)	0.067 (0.099)	0.072 (0.082)	0.048 (0.123)	-0.019 (0.118)
Spring	-0.102 (0.166)	-0.010 (0.161)	-0.408 (0.269)	0.068 (0.183)	-0.132 (0.273)	-0.225 (0.266)
School controls	No	Yes	Yes	Yes	No	Yes
Teacher controls	No	Yes	Yes	Yes	No	Yes
R-squared	0.006	0.120	0.125	0.126		
No. of observations	33,258	29,260	9,153	20,107	33,258	29,260

Notes: The outcomes of OLS (columns 1-4) and 2SLS (columns 5-6) regression models are reported. The endogenous explanatory variables for school and teacher characteristics are included but not reported above. The standard errors are in parentheses. The model is fitted using the BRR weights. The symbols +, \*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, 1%, 0.1% levels, respectively.

## Timing Is Everything

**Table B.4. Classroom Control over the School Year**

	OLS	OLS	OLS	OLS	IV	IV
Sample	All	All	First	Later	All	All
	(1)	(2)	(3)	(4)	(5)	(6)
Days	-0.001 (0.001)	0.000 (0.001)	0.001 (0.001)	0.000 (0.001)	-0.004* (0.002)	0.001 (0.002)
Winter	0.032 (0.069)	-0.031 (0.055)	-0.008 (0.085)	-0.036 (0.069)	0.218+ (0.118)	-0.089 (0.106)
Spring	0.113 (0.168)	-0.057 (0.146)	-0.016 (0.212)	-0.062 (0.177)	0.522* (0.237)	-0.185 (0.229)
School controls	No	Yes	Yes	Yes	No	Yes
Teacher controls	No	Yes	Yes	Yes	No	Yes
R-squared	0.001	0.164	0.176	0.162		
No. of observations	33,258	29,260	9,153	20,107	33,258	29,260

Notes: The outcomes of OLS (columns 1-4) and 2SLS (columns 5-6) regression models are reported. The endogenous explanatory variables for school and teacher characteristics are included but not reported above. The standard errors are in parentheses. The model is fitted using the BRR weights. The symbols +, \*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, 1%, 0.1% levels, respectively.

**Table B.5. Leadership Support over the School Year**

	OLS	OLS	OLS	OLS	IV	IV
Sample	All	All	First	Later	All	All
	(1)	(2)	(3)	(4)	(5)	(6)
Days	-0.002* (0.001)	-0.001+ (0.001)	-0.001 (0.001)	-0.002+ (0.001)	-0.000 (0.001)	0.000 (0.001)
Winter	0.027 (0.050)	0.040 (0.055)	0.066 (0.091)	0.038 (0.063)	-0.062 (0.087)	-0.061 (0.090)
Spring	-0.080 (0.119)	-0.034 (0.138)	0.245 (0.222)	-0.092 (0.158)	-0.274 (0.191)	-0.258 (0.209)
School controls	No	Yes	Yes	Yes	No	Yes
Teacher controls	No	Yes	Yes	Yes	No	Yes
R-squared	0.004	0.031	0.035	0.035		
No. of observations	33,258	29,260	9,153	20,107	33,258	29,260

Notes: The outcomes of OLS (columns 1-4) and 2SLS (columns 5-6) regression models are reported. The endogenous explanatory variables for school and teacher characteristics are included but not reported above. The standard errors are in parentheses. The model is fitted using the BRR weights. The symbols +, \*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, 1%, 0.1% levels, respectively.

**Timing Is Everything**

**Table B.6. Student Behavior Problems over the School Year**

Sample	OLS	OLS	OLS	OLS	IV	IV
	All (1)	All (2)	First (3)	Later (4)	All (5)	All (6)
Days	0.004*** (0.001)	0.003*** (0.001)	0.003** (0.001)	0.003** (0.001)	0.003* (0.001)	0.004** (0.001)
Winter	-0.041 (0.059)	-0.031 (0.055)	-0.034 (0.082)	-0.039 (0.062)	0.038 (0.091)	-0.116 (0.087)
Spring	-0.355* (0.147)	-0.185 (0.139)	-0.547** (0.194)	-0.108 (0.156)	-0.181 (0.205)	-0.375+ (0.194)
School controls	No	Yes	Yes	Yes	No	Yes
Teacher controls	No	Yes	Yes	Yes	No	Yes
R-squared	0.019	0.263	0.242	0.272		
No. of observations	33,258	29,260	9,153	20,107	33,258	29,260

Notes: The outcomes of OLS (columns 1-4) and 2SLS (columns 5-6) regression models are reported. The endogenous explanatory variables for school and teacher characteristics are included but not reported above. The standard errors are in parentheses. The model is fitted using the BRR weights. The symbols +, \*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, 1%, 0.1% levels, respectively.

**Table B.7. Variation before and after State Exams**

	Poverty Challenges		Classroom Control		Leadership Support		Student Behavior	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Time prior	-0.001 (0.001)	-0.001 (0.001)	-0.001+ (0.001)	-0.000 (0.001)	-0.001 (0.001)	-0.001 (0.001)	0.001 (0.001)	-0.001 (0.001)
Prior	-0.105 (0.219)	-0.038 (0.342)	0.181 (0.214)	0.347+ (0.187)	0.545** (0.196)	0.238 (0.282)	-0.045 (0.165)	-0.147 (0.158)
School controls	No	Yes	No	Yes	No	Yes	No	Yes
Teacher controls	No	Yes	No	Yes	No	Yes	No	Yes
R-squared	0.017	0.157	0.006	0.243	0.014	0.090	0.007	0.157
No. of observations	1,829	1,487	1,829	1,487	1,829	1,487	1,829	1,487

Notes: The outcomes of OLS regression models are reported. The endogenous explanatory variables for school and teacher characteristics are included but not reported above. The standard errors in parentheses are in parentheses. The model is fitted using the BRR weights. The symbols +, \*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, 1%, 0.1% levels, respectively.

## Timing Is Everything

**Table B.8. Variation before and after State Exams: Response Time**

	Poverty Challenges		Classroom Control		Leadership Support		Student Behavior	
	First (1)	Later (2)	First (1)	Later (2)	First (1)	Later (2)	First (1)	Later (2)
Time prior	-0.002 (0.003)	-0.001 (0.001)	-0.000 (0.001)	0.001 (0.001)	0.000 (0.002)	-0.002 (0.001)	-0.001 (0.002)	-0.001 (0.001)
Prior	0.053 (0.486)	-0.530+ (0.310)	0.304 (0.299)	0.082 (0.481)	0.029 (0.289)	0.847* (0.365)	-0.086 (0.209)	-0.408+ (0.237)
School controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.194	0.185	0.249	0.298	0.187	0.115	0.219	0.175
No. of observations	573	914	573	914	573	914	573	914

Notes: The outcomes of OLS regression models are reported. The endogenous explanatory variables for school and teacher characteristics are included but not reported above. The standard errors in parentheses are in parentheses. The model is fitted using the BRR weights. The symbols +, \*, \*\*, and \*\*\* indicate statistical significance at the 10%, 5%, 1%, 0.1% levels, respectively.

Copyright © 2015 by Peter T. Goff, Jihye Kam, and Jacek Kraszewski

All rights reserved.

Readers may make verbatim copies of this document for noncommercial purposes by any means, provided that the above copyright notice appears on all copies. WCER working papers are available on the Internet at <http://www.wcer.wisc.edu/publications/workingPapers/index.php>.

Correspondence concerning this article should be addressed to Peter T. Goff, Department of Educational Leadership & Policy Analysis, School of Education, University of Wisconsin-Madison, 253 Education Building, 1000 Bascom Mall, Madison, WI 53706. Any opinions, findings, or conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies, WCER, or cooperating institutions.