School Year Length and Student Performance: Quasi Experimental Evidence

Benjamin Hansen[*][†]

October 20, 2011

Abstract

This paper investigates the impact of instructional days on student performance. Because school year length is endogenously determined, I estimate the causal impact of school year length through two quasi-experiments that exploit different sources of variation in instructional days. The first identifies school year length's effect through weather-related cancellations in Colorado and Maryland. Weather-related cancellations are made up at the end of school years, allowing relatively large fluctuations in instructional days within school districts prior to test administration. Because school cancellations are not recorded for past school years, this data limitation is overcome by using two-sample indirect least squares. The second identification strategy takes advantage of state-mandated changes in test-date administration in Minnesota, which moved 5 times in 5 years. The results are similar for either source of instructional day variation: more instructional time prior to test administration increases student performance. The effects are consistent across various thresholds of performance and grade levels.

[*]Address: 1285 University of Oregon, Eugene OR, 97403. Email: bchansen@uoregon.edu. Phone: 541-357-8395. Fax: 541-346-1243.

1 Introduction

The positive association between education and earnings is one of the most robust findings in labor economics. However, not all educations are created equal. Indeed, quality has varied historically across demographic groups both within the United States and across countries. As a result, some policy makers have suggested increasing the quality of education as a tool in reducing labor market gaps in wages and employment. Interestingly, the pupil teacher ratio and per student spending – common policy interventions – have respectively fallen and risen in recent years while school year length has been stable (see Figure 1). Longer school years provide the potential for increased instruction time, review, and attention for individual students. While school year length has largely not been a topic of serious discussion in education policy, it has recently drawn more attention. Hawaii has been scolded by Secretary of Education If increased school year length does improve student performance, it could also be an alternative strategy for schools which have trouble attracting new or better teachers. This paper offers quasi-experimental evidence on school year length and its consequences for student performance.

There is a continuing debate on whether educational quality has a bearing on student outcomes – with academics, educators, and policy makers on both sides. The discourse began with the Coleman Report (1966), which found that per pupil resources have little impact on student success. Since then, for every study refuting the Coleman Report's conclusions, another supports them. Hanushek (1981) shows increased expenditure on teachers is unlikely to improve performance. Meanwhile, Margo (1986) estimates that 27 percent of the black/white literacy gap from 1920 to 1950 can be explained by differences in educational inputs. Krueger (1999) finds Project STAR students randomly assigned to small classes do better on standardized exams, though the benefits may be temporary. Relatively little work has investigated the impact of school year length, but that done has continued in the same

1

spirit of discord.

Initial research on school year length focused on labor market outcomes, while later studies have investigated test scores. Card and Krueger (1992) compare workers raised in different states, finding those from states with relatively longer school years earn more. Pischke (2003) takes advantage of short school years mandated in Germany to unify their schooling system. He concludes that shorter school years increase grade repetition, but have no long-term effects on employment or wages. Through a similar regime change, Krashinsky (2006) studies the elimination of the fifth year of high school in Ontario, Canada. Cohorts with four years of high school had substantially lower grade point averages in college than those who attended high school for five years. Contrary to other research, recent international cross-section studies by Lee and Barro (2001) and Wobmann (2000) conclude school year length has no impact on test scores. Eren and Mittlemet (2005) study the National Longitudinal Survey of Youth, which asks whether an institution's school year is longer or shorter than 180 days. They find that the best performing students benefit from longer school years while low performing students do worse with increased instructional time. Marcotte (2007) investigates the reduced form relationship between yearly snowfall and test scores, finding years with substantial snowfall are associated with lower performance in Maryland. Like previous school quality research, a consensus has yet to be reached regarding school year length's effect on student outcomes.

Due to inclement weather, districts routinely cancel school to avoid the liability and danger of traveling on unsafe roads. These cancellations, commonly called "snow days", vary from year to year and across districts, causing states to adopt policies in order to guarantee school is in session sufficiently. For example, the state of Colorado mandates that schools must extend their school year into the summer if total instructional hours fall below 1040. Given current scheduling, this amounts to less than three cancellations for most districts.

2

Conveniently (for the purposes of this study), Colorado administers its standardized tests in March, months before any missed days are ever made up. The same can be said for Maryland, which administers its tests at the end of April while school releases in June.

Because histories of cancellations for Maryland schools are not maintained, Marcotte estimates the reduced form relationship between aggregate snowfall and student performance. Marcotte and Hemelt (2007) obtain partial cancellation histories for Maryland, finding instructional days have significant effects on performance. However, due to the incomplete nature of cancellation histories they pool together two testing regimes (MSPAP (1993-2002) and MSA (2003-2005)) which introduces a potential selection problem: districts with relatively few cancellations tend to not maintain cancellation histories as far back as other schools.[1] Colorado also fails to collect closure histories in any unified location. I overcome this obstacle for both states by using a two sample estimation technique similar to two-sample IV (Angrist and Krueger, 1995). For the 06/07 and 07/08 school years I have collected daily cancellation informationby surveying schools in Colorado and Maryland. I combine a first stage of weather's impact on cancellations (for 2006-2008) and a reduced form of weather's relationship with student performance (for 2002-2006 in CO and 1993-2002 for MD) and estimate the effect of school cancellations through indirect least squares. This approach allows me to study the effect of weather-related cancellations over long periods of time, even if cancellation histories are not maintained. Using this method, future studies can easily confirm the effect weather-related cancellations on student outcomes, even if information on cancellations is available only for the more recent school years.

A second identification strategy investigates test examination dates, which changed 5 times over 5 years in Minnesota. The changes in test dates alternated between moving the test earlier and later. They were moved earlier by 10 and 11 school days (in 2002 and 2004), and were scheduled later by 10 school days and 15 school days twice (in 2001, 2003, and

---

[1]Marcotte and Hemelt adjust for this using district-specific time trends.

2005 respectively). This created substantial variation both increasing and decreasing the amount of time students received prior to examination. In addition, using variation do to test-date shifts may offer more external validity over variation in instructional timing due to weather-related cancellations as snow days interrupt instruction in unanticipated fashion. However variation due to a test-date shift is known by the teacher in advance and they can make their plans according, perhaps replicating more closely how teacher would respond to an extended school year.

This paper identifies the effect of instructional days from two different sources of variation, yet they both yield similar evidence regarding school year length's effect on student performance. Because the available performance variables are proportions, the effects are estimated using familiar probability models for grouped data. In addition, because the latent variable is a test score, the estimated effects on the latent variable have a valuable economic interpretation: how many standard deviations average scale scores have changed. Both the response probabilities and the implied effect on latent test scores yield evidence that increased instructional days raise student performance. This suggests that extending the school year can be a method of increasing student performance, and perhaps with it, human capital accumulation.

2   School Year Length: Background and Identification

The education production function is a common model used to study the choices of administrators and their ultimate consequences. The administrators are free to pick the levels of various inputs in the educational process, subject to their budgets and state guidelines. Examples of inputs include teachers (in number or quality), textbooks, and the length of instruction time. Outputs of the educational process include test scores, grades, graduation, going to college, and finding jobs, among many others. Figure 1 compares the national

4

trends of the pupil-teacher ratio and real per pupil spending against school year length over the last century. Contrasting the trends, expenditures on teacher employment have risen considerably, while little funds have been devoted to extending the amount of instructional time students receive. If longer school years do improve student outcomes, they could be an alternative to other policies that influence school quality.

The magnitude of school year length's impact on student outcomes is largely an empirical question. However, comparing across states or nations to assess school year length's effect can introduce problems of bias. Actual instructional days can be divided into two parts: the planned instructional days and cancellations. Planned instructional days are under the control of the administrator, subject to budgetary constraints and time. Most previous studies have focused on differences in planned instructional days, identified by comparing across states or nations. However the differences in planned instructional days can be largely due to differences in budgets, introducing possible upward bias. Also one might have concern that struggling schools might extend their school year to improve performance on standardized tests, which would bias school year length's effect downward. Texas recently required all districts to begin every year on the last Monday in August for this reason.[2] Using planned instructional days can bias school year length's effect, and the sign of the bias is arguably indeterminate. Thus studying the component of school year length under the control of administrators – planned instructional days – may be counter productive.

Using variation in instructional days due to weather-related cancellations can eliminate the selection problems associated with longer planned school years (which could indicate greater school resources or poor performance on prior exams). The part determined outside the control of administrators still informs about the general effects of increasing instructional days, as weather-related cancellations reduce the amount of time teachers have to instruct,

---

[2]Dallas Morning News, Thursday May 4, 2006

quiz, or meet with students.[3]

Cancellations due to weather identify the effect of instructional days based on yearly fluctuations due to weather, when the test date is fixed. Another possibility is to study situations were dates of examination are shifted. This approach could share some common advantages with weather related cancellations. The variation occurs within districts, and changing the date of test administration does not alter other school resources. Although schools might wish to move their date of examination for endogenous reasons, in Minnesota all the shifts were state-wide. In addition, from 2000 to 2005 the date of examination alternated being shifted later and earlier. So trends which are relatively smooth–such as changes in demographics or school quality–can be controlled for and thus prevent bias due to spurious correlation. Because a changes in instructional time due to a test date shift is known at the beginning of the school year, teachers have time to plan out their year accordingly. For this key reason, examination date changes may more closely resemble an extended or shortened school year, offering estimates with more external validity.

2.1 Exogeneity of Weather: Snowfall's Spatial Distribution

A critical assumption in order for cancellations due to weather to identify the causal effect of instructional days is that cancellations be randomly assigned to schools. Even though weather is exogenous, if it is correlated with unobserved elements that impact student performance, causal effects remain unidentified. Thus choosing the correct sample framework, cross-section or panel, can be vital to identifying a causal effect.

Snow accumulates heavily along the mountain range in the middle of Colorado, and neglects to impact the southeastern region. Income in Colorado follows nearly the same spatial pattern. Though not as clear as Colorado's, it seems the correlation between snowfall

---

[3]We discuss in Section IV reasons weather-related cancellations could under or overstate school year length's causal effect.

and income is reversed in Maryland, with the poorest regions in the western strip of Maryland receiving the most snow while the wealthiest regions receive only mild amounts.[4] Although snowfall is exogenous, choice of residence is not random throughout the two states.

To see the extent of correlation between snowfall and resources in Colorado and Maryland, cross-section regressions are estimated using weather as the dependent variable. These are done purely to measure correlation between levels of weather and levels of resources, with the results presented in Table 2. In Colorado, districts with substantial snowfall tend to be rich districts while the correlation between snowfall and student family income varies by year in Maryland. However running the regression as a panel and controlling for district fixed effects and year specific trends, none of the variables are by themselves or jointly significant. So although snowfall exhibits spatial correlation with student or school resources, schools experiencing variation in snowfall are not systematically experiencing changes in school resources. Controlling for school level fixed effects and yearly trends can eliminate the selection bias that would be introduced due to non-random selection of residence in Maryland and Colorado.

2.2 Minnesota: Examination Date Variation

Another source of variation in instructional days exploited is the shifts in scheduled test date administration for the Minnesota Comprehensive Assessment. Minnesota is one of six states which mandates that school start after a specific date, with the remaining states leaving it to the discretion of local school districts.[5] Its September 1 starting date is also tied for the latest.[6] Between the years 2000 and 2005, the Minnesota Department of Education moved the date for its assessment each year, and by several days each time. Because of the shared mandated starting time for schools, shifts in the test date create the potential for more or

---

[4]Figures in the appendix demonstrate the spatial patterns across the two states.

[5]The other five are Texas, Indiana, North Carolina, Virginia, West Virginia. Taken from Education Commission of the States.

[6]Minn. Stat. 120A.41. Also in consequence, most schools begin the day after labor day.

less instructional time. The trend of average test scores is plotted against the number of instructional days prior to examination in Figure 3. Every time the test date is moved earlier, the trend flattens out, while tests administered later in the year show considerably more improvement. This is mirrored when plotting the change in average test scores against the change in instructional days.

The same effects are observed at a more disaggregated level. Using school level average test scores in Figure 4, I plot the distribution of the change in average scale scores for years with tests earlier in the school year, contrasted with the distribution of the change in scores for tests administered later in the year. The year-to-year change in scale scores is shifted to the right for both grades when the test is administered later in the year. Lastly as a robustness check, Figure 5 demonstrates variables strongly correlated with student ability and school resources (fraction eligible for free lunch and teachers employed) were relatively smooth through the same time period.

3 Specification and Estimation

The student performance data are results from the Colorado, Minnesota, and Maryland State Assessments. Each of the tests has stakes for teachers and administrators, but not for students.[7] Mathematics exams are studied because they are relatively objective and cover a consistent curriculum. All 3 states publicly make available grouped averages of performance, which will be the dependent variables of interest when calculating school year length's effect. However, it is useful to consider a simple model of student performance at a micro level to accurately interpret the results and establish identification.

---

[7]Depending on how close students are to a threshold they may exhert more or less effort to pass an exam, which has stakes for the student. See Betts (1996).

3.1 Micro Model of Student Performance

Consider a model of testing where a student's performance depends on his or her observable characteristics, his or her school's resources, and instructional days. Instructional days are the planned instructional days less cancellations, where cancellations are influenced by weather (snowfall in particular) and planned instructional days depend on resources. With information on individual student test scores, one could estimate linear regressions for the following model.

$$T_{ist} = I_{st}\beta + X'_{it}\beta_X + R'_{st}\beta_R + s_s + \tau_t + \varepsilon_{ist}$$

$$I_{st} = P_{st} - C_{st}$$

$$P_{st} = R'_{st}\alpha_R + e_{st}$$

$$C_{st} = w_{st}\alpha + v_{st}$$

$\quad T_{ist} \quad : \quad$ Student $i's$ test performance at school s at year t

$\quad I_{st} \quad : \quad$ Actual Instructional days for school s at year t

$\quad P_{st} \quad : \quad$ Planned instructional days for school s at year t

$\quad C_{st} \quad : \quad$ Cancellations for school s at year t

$\quad X_{it} \quad : \quad$ Characteristics for student i at year t

$\quad R_{st} \quad : \quad$ Resources at school s at year t

$\quad w_{st} \quad : \quad$ Weather for school s at year t

$\quad s_s \quad : \quad$ School fixed effect

$\quad \tau_t \quad : \quad$ Year fixed effect

At this point a student's performance depends on resources, both at the individual and school level, and instructional time. The reduced form impact of weather on student performance is

$$\frac{dT_{ist}}{dw_{st}} = -\beta\alpha.$$

Ideally, we would construct a weather measure w_{st} such that $\alpha = 1$.[8] If $\alpha = 1$, then there is no need for a first stage as using the reduced form estimates of weather would be equivalent to the structural relationship instructional days and student performance.[9] Aggregate snowfall is likely to be correlated with closures but can be improved upon.[10] For instance 10 days where it snows 1 inch will probably not lead to any cancellations. However, one day with 10 inches of snow almost surely would. Aggregate snowfall would treat these realizations of weather the same. To more accurately assess weather likely to cancel school, I also construct measures of weather based on the number of days on which snowfall exceeded thresholds. Of course another trade-off exists. Thirty inches of snow on one day might cancel school for the next 3 or 4 days, but threshold variables would treat this as equivalent to one day with 4 inches of snow. For completeness, both weather measures are considered.[11] Consider now the reduced form representation, removing instructional days directly from the regression.

$$T_{ist} = w_{st}(-\beta\alpha) + X'_{ist}\beta_X + R'_{st}(\beta_R + \beta\alpha_R) + \beta v_{st} + s_s + \tau_t + \varepsilon_{ist}$$

One can rewrite the expression above, getting

$$T_{ist} = w_{st}\gamma + X'_{ist}\beta_X + R'_{st}\psi + s_s + \tau_t + u_{ist}, \tag{1}$$

where $(\beta e_{st} - \beta v_{st} + \varepsilon_{ist}) = u_{ist}$, $-\beta\alpha = \gamma$, and $(\beta_R + \beta\alpha_R) = \psi$.

---

[8]In a slight abuse of notation I refer to different weather measures having different α's. This could be more accurately respresented as $\alpha_{w_{st}}$. To avoid more cumbersome notation, we will refer to them all as α.

[9]The reduced form for any weather measure could be rescaled so that $\alpha = 1$. Essentially a first stage relationship tells one how to rescale the units on the weather variable so that $\alpha = 1$.

[10]Marcottes chosen regressor. Marcotte uses yearly snowfall, I remove snow during winter vacation, weekends, or school holidays.

[11]There are many other weather measures that could be used. However, any measure highly correlated with cancellations that doesn't impact students other than through cancellations is sufficient and necessary for identification.

This regression would be easy enough to run, were micro-level data on student performance available. However, state assessment results made publicly available contain grouped information for grade levels within schools. Maryland publishes the proportion proficient and advanced while Colorado releases the proportion partially proficient, proficient, and advanced. Minnesota reports the proportion partially proficient, proficient, advanced, and also average test scores.[12] One can still estimate the effects on student performance, but the data requires it be done in the context of probability models.

Then if we are interested in the probability that $T_{ist} \geq t^*$, where t^* is an academic standard, the partial effects include the reduced form effect on test scores, along with the density at the cut off point. To illustrate this point, let us rewrite the effect of weather on student performance, given the data refer to the probability of exceeding an academic standard.[13]

$$P(T_{ist} \geq t^*)$$

$$= P(Iw_{st}\gamma + X'_{ist}\beta_X + R'_{st}\psi + s_s + \tau_t + u_{ist} \geq t^*)$$

$$= P(u_{ist} \geq t^* - (w_{st}\gamma + X'_{ist}\beta_X + R'_{st}\psi + s_s + \tau_t))$$

$$\implies$$

$$\frac{dF}{dw_{st}} = \underbrace{f(t^* - (w_{st}\gamma + X'_{ist}\beta_X + R'_{st}\psi + s_s + \tau_t))}_{(I)} \quad \underbrace{(\gamma)}_{(II)}$$

or

$$\frac{dF}{dw_{st}} = \underbrace{f(t^* - (w_{st}\gamma + X'_{ist}\beta_X + R'_{st}\psi + s_s + \tau_t))}_{(I)} \quad \underbrace{(-\beta\alpha)}_{(II)}$$

---

[12]Maryland calls its middle category satisfactory, while Colorado and Minnesota refer to it as proficient. Maryland calls its highest category excellent, while Colorado and Minnesota name it advanced. In this paper, satisfory and excellent proportions in Maryland will be refered to as proficient and advanced for simplicity.

[13]We continue with the representation of a micro probability model. Grouped probability estimates have similar similar response probabilities, controlling for average student characteristics rather than particular traits.

The partial effects have two components: the effect of weather on latent test scores (II), and the density at the cutoff (I). One can estimate the impact at different academic standards, grades, or demographic groups. However, differences in the response probabilities reflect both variation in the latent effect on student performance or different densities of students at the chosen standard. In addition, typical procedures used to scale effects can be problematic in probability models.

When interpreting the practical size of coefficients, one method is to compare the estimated effect to the mean of the dependent variable. This can give the researcher an approximation of the "percentage" effect. With probability models this is problematic because the choice of failure and success is arbitrary. For this problem, I could have chosen to examine the probability of being below an academic standard. The new estimated effect is the same in magnitude, only with the sign reversed. However, the proportion of students below the standard is by definition one less the proportion above. By comparing the partial effect to the mean of the dependent variable, one can either inflate or deflate the "percentage" effect depending on the arbitrary definition of success and failure.[14] In other words, using the mean to scale the effect is not invariant to the researcher's choice of success in probability models. This obstacle can be partially overcome depending upon which probability model is implemented.

I proceed now to the estimators used in grouped probability models. Probability is replaced by its sample next-of-kin, the proportion of the students exceeding a threshold Minimum chi-square methods provide several different well studied estimators from which to choose.[15] I examine the linear probability model and normit presented below (with a general dependent variable and vector of regressors).

---

[14] This is similar to estimating elasticities. We could use the initial or the end point to scale the change. For this reason it is common to used the midpoint to get an average elasticity. In our case, the midpoint is always .5 by definition of probability.

[15] See Madalla for an extensive chapter on micro and grouped probability models (1983).

$$\text{Linear Probability Model} \quad P_t = X_t'\beta + u_t$$

$$\text{Normit} \quad \Phi^{-1}(P_t) = X_t'\beta + u_t$$

The linear probability model is familiar from binary outcomes, and the normit, the grouped version of a probit, is a reasonable choice for test scores well-approximated by a normal distribution.[16]   Also the β estimated by the normit regression has a valuable interpretation. When taking the normit (inverse cumulative normal) transformation of the proportion variable, the transformed variable is a standard normal variable. Due to the normit transformation, the estimated β indicates how many standard deviations latent test scores have shifted. Focusing on the impact on the latent variable (which is called the the latent effect from this point on in the paper) rather than response probabilities also provides an invariant way to compare partial effects, as the transformation eliminates the density component. This is one of the few situations where the untransformed coefficient in a probability model has a valuable economic interpretation. This is useful for comparing effects across grades, proficiency standards, or states, which have both different standards and densities of students.[17]

3.2   First Stage: Weather and Cancellations

Up to this point, the focus of the discussion has been on estimating the reduced-form effect of weather on student performance, γ. In order to place a magnitude on how additional school days affect student performance, the reduced-form effect needs to be scaled by the relationship between weather and cancellations, α. Because cancellation histories are not maintained, this data challenge is overcome by estimating a first stage equation for the 2006/2007 school year. The weather variables previously discussed are included as regressors

---

[16]Typically there is some skewing in tests. Early grades they are skewed right, and scores skewed left for later grades. Grades in the middle are typically those most symetrically distributed.

[17]If the data are truely generated by a normal distribution and the effect of instructional days is linear, then the latent effect will be the same across across thresholds.   If the effect differs across performance standards, this could be both due to non-normality or non-linearity of the effect.

in the first stage regressions. Two possible specifications for a first stage are explored. A high frequency approach estimates how well the weather variables predict closures on a particular day. A low-frequency analysis estimates how weather over the course of the school year predicts the number of cancellations occurring within that year.

The first specification's dependent variable is an indicator for whether school is open or cancelled at a particular district on a given day. Because the data are measured at the daily level, there are likely to be some matching problems. For example, snowfall on a Monday night would cancel school on Tuesday but is matched with Monday's school closure status. This measurement error will likely attenuate α towards zero. In addition, because the threshold variables are indicators, α will be naturally bounded between zero and one. Because β is the parameter of interest, attenuation of α would bias the estimate of β away from from zero as the reduced form effect γ is divided by $-\alpha$ to recover β (cancellations refer to lost days, so dividing by $-\alpha$ yields a β that corresponds to the effect of an additional day of schooling). For this reason, it may useful to think of the indirect least squares estimates as upper bounds.

$$Cancellation_{sd} = \alpha_o + \alpha w_{sd} + d_s + v_{sd}, \qquad (2)$$

where s indicates district and subscript d denotes the day and d_s is a district fixed-effect.

The low frequency approach uses the number of cancellations as the dependent variable, aggregating equation (2). The true population parameters remain unchanged with this aggregation for the population model, due to linearity. However misclassifications of weather due to calendar effects may be reduced as a lot of snow on Monday evening or Tuesday morning would both aggregate to 1 day with a lot of snow.[18]

---

[18]Notice for either specification, the other controls have been ommitted from the first stage. This is mainly due to the fact that the regressors are not yet available for the 2006/2007 school year. In addition, in the high frequency approach any variables that are time constant are absorbed because of the fixed effects. Because this includes any regressors that don't vary throughout a school year, the fixed effects are collinear

For completeness, both the low-frequency and high-frequency methods to estimate α are computed and if there is little mismatching of the weather variables, the estimates will be similar. Regardless, after estimating α the standard errors need to be adjusted to account for both the randomness of $\hat{\gamma}$ and $\hat{\alpha}$. The limiting distribution of $\frac{\hat{\gamma}}{\hat{\alpha}}$ is approximated using the delta-method. Recall for an estimated parameter vector $\hat{\theta}$, $g(\hat{\theta})$ has the following limiting distribution where $G(\theta)$ is the matrix of partial derivatives with respect to θ.

$$N(g(\hat{\theta}), G(\theta)'V(\hat{\theta})G(\theta))$$

In our case, the form of $g(\hat{\theta})$ is $\frac{\hat{\gamma}}{\hat{\alpha}}$.[19] The reduced forms for Colorado and Maryland are estimated respectively for 2002-2006 and 1993-2002. For both states the first stage is estimated for the 2006/2007 school year. Because the parameters are estimated from separate samples, it is assumed that the off-diagonal elements of the variance-covariance matrix are zero.

$$g(\gamma, \alpha) = \frac{\gamma}{\alpha}$$
$$G(\gamma, \alpha) = \begin{pmatrix} \frac{1}{\alpha} \\ \frac{-\gamma}{\alpha^2} \end{pmatrix}$$

$$\frac{\hat{\gamma}}{\hat{\alpha}} \to^d N\left(\frac{\gamma}{\alpha}, \frac{var(\hat{\gamma})}{\alpha^2} + \frac{var(\hat{a})\gamma^2}{\alpha^4}\right) \tag{3}$$

Notice if $\alpha = 0$, the mean and variance will be infinite, making the distribution undefined. This makes a powerful first stage critical to this study, like any instrumental variables approach.

---

with all school and student characteristics recorded at the yearly level.

[19]This typically requires continuity of the function of the parameters. This function is continuous everywhere, except where $\alpha = 0$. This is a common problem of exactly identified instrumental variables equations. In the results section the first stage is sufficiently powerful to reject the null that $\alpha = 0$.

3.3 Minnesota

For Minnesota, similar regressions are estimated, albeit without some of the complications using weather to generate random variation in instructional days. The regressor of interest is simply the number of days prior to examination. This is found by calculating the number of potential school days between the first of day of school and the test date (removing holidays, weekends etc.). Because historical school schedules are not maintained, winter break is defined to be between December 23 and January 3. Though there might differences in winter break length, the fixed effects will capture any time constant discrepancies. So even if some schools have more instruction (due to winter break differences) than others, the deviation in instructional days from the mean will be the same for all school districts. If there are changes in winter break length over time (or weather-related cancellations), this would introduce measurement error, attenuating the estimates. One caveat is that because schools are experiencing the same deviation from their mean instructional time, instructional days would be correlated with year effects. In order to adjust for trends (which Figure 5 strongly suggests exist), school specific quadratic trends are included in the regressions.

4 Results

4.1 Data Sources

The performance data are taken from mathematics results made publicly available from the Maryland, Minnesota, and Colorado Departments of Education. The Maryland assessment results are from 1993-2002, Colorado's cover 2002-2006, and Minnesota's span 2000-2005. The 3rd, 5th, and 8th grades are studied in Maryland, the 8th grade is explored in Colorado, and the 3rd and 5th grades are examined in Minnesota.[20] Maryland and Minnesota also

---

[20] In Colorado, schools following a year-round schedule or 4-day school were excluded. This because details regarding breaks for year-round schools were not maintained, and 4-day schools report which weekday they have off since 2003, but not prior. Although the exam began administration in 2000, 2002 on is studied

16

make available the variables used as controls, while the control characteristics for Colorado are taken from the National Center of Educational Statistics. The weather data are daily surface observations from the National Climatic Data Center (details contained in the appendix on the linkage). Data on cancellations for Colorado and Maryland were obtained by calling school districts at the end of the 06/07 school year and checking web sites that announce cancellations.[21] The summary statistics are found in Table 3.

Control characteristics of the schools and their student bodies are included in the regressions. Even though weather and the test date changes are plausibly exogenous events, including the controls can prevent spurious correlation and also reduces sampling error. Common controls to all regressions run include the fraction of students eligible for reduced price lunches and the pupil teacher ratio. School fixed effects are included in all regressions, while year dummies account for trends in Colorado and Maryland and quadratic school specific trends are included for Minnesota. Maryland and Minnesota have a few unique controls not available through the National Center of Educational Statistics.[22] Colorado and Maryland also both report information on teaching assistants per pupil. Maryland and Minnesota both record the proportion of students which are limited-English proficient. Maryland provides yearly data on per capita wealth and the fraction of students which are Title I eligible and Minnesota has information on the average experience of teachers. Excluding or including these additional variables in Maryland or Minnesota has little impact on the results. Also, all the reduced-form regressions are weighted by the number of students taking the test, and because the level of snowfall is shared by all schools within a district in a year, standard errors are clustered by district and year.[23]

---

because the scale scoring changed to a new regime in 2002.

[21] In Colorado 107/178 districts provided cancellation for 2006/2007 school year, while in Maryland 19/24 responded.

[22] For Colorado, controls for the 05/06 school year had not been released yet. The prior years values were imputed for these missing obsevations. Also the mean of previous values was tried. The results are robust to method of imputation, or excluding the controls.

[23] This is in part due to relatively few districts in Maryland. Only 24 exist, and some are quite large with dozens of schools. Thus if one clustered at the district level, some clusters could take up excessively large

4.2 First Stage Estimates

To infer the effect of additional instructional days on student performance, weather's reduced form effects need to be scaled by weather's relationship with cancellations for Colorado and Maryland. The high frequency approach employees a linear probability model and includes district level fixed effects. An observation is a day for a district. Meanwhile, the low frequency approach aggregates over the year and compares across districts.

Both approaches yield similar estimates for Maryland. Each additional inch of snowfall increases the odds of a cancellation by .16. The high frequency estimate is precise enough with an F-statistic of 20.15 to suggest the instrument is not weak. This suggests that the reduced form coefficients should be scaled up by a factor of 6 in Maryland. Colorado superintendents are more resistant to snow, as an additional inch of snow is estimated to raise the probability of cancellation by .05, somewhat smaller than in Maryland. For every day with snow greater than 4 inches, the probability that Colorado school districts cancel school increases by .37. The high frequency regressions provide the most precise estimates of the structural relationship between weather and cancellations, all passing weak instrument standards, and hence are used for the indirect least squares estimates of an instructional day's effect. Any measure of weather could be linked with its reduced form for Colorado. The number of days with snow greater 4 inches is used for Colorado and inches of snowfall is used for Maryland for the final estimates presented in the next section, and the results are similar across other weather measures.[24]

---

partitions of the variance-covariance matrix.

[24] The 4 inch threshold measure seems the most robust across the two frequencies. In addition, it may be less sensative than snowfall to outliers such as the large snow-storm which hit Colorado December 20-23, 2006.

4.3 Reduced Form Estimates

Both the linear probability model and normit will be used in estimating the reduced form effect of snowfall and performance. Recall $\frac{dF}{dw_{st}} = f(\cdot) \ (-\alpha\beta)$. If one uses the same weather variable and compares across performance measures, differential effects could reflect both differences in effects on latent performance β, as well the density of students at the cutoff. Because the density is always greater than or equal to zero we can identify the sign of $-\alpha\beta$ above, but relative magnitudes cannot be compared because of differences in densities. Two model specifications are used to estimate the response probabilities, and the effect on latent scale scores. The linear probability model is used in estimating response probabilities, as it does not require specification of $f(\cdot)$ thus offering some additional robustness properties.[25] The untransformed normit coefficients will provide estimates of the effect of weather on latent scale scores.

With more days with substantial snowfall, the proportion above each of the academic standards falls. In addition, the effects are strongest low in the test score distribution, as the impacts of all the weather variables on the proportion partially proficient are larger and more statistically significant than the effects at other proficiency cutoffs.[26] With each day of snowfall with more than four inches, the fraction partially proficient declines by .0056 From the normit regression, an additional day with snow greater than 4 inches decreases test scores by .015 standard deviations (at the partially proficient standard). The direction of the effects is clear, increases in snow is associated with lower student performance.

Maryland shows similar results to Colorado's, albeit with greater statistical precision. The results are presented in Table 4. Rows labeled "proportion proficient" and "proportion advanced" refer to the linear probability estimates, while "latent proficient" and "latent

---

[25]It should be noted that estimated normit response probabilities closely mirror those estimated by the linear probability model and are available upon request.

[26]This could also be due to a local effect, if districts that experience the most variation in snowfall are also those whose density is most concentrated around the partially proficient standard. See Angrist and Imbens (1994).

advanced" refer to the untransformed normit coefficients.[27]   An additional inch of snowfall decreases the proportion scoring proficient by .00073 for the 5th grade and .00053 in the 8th grade.   Likewise, an inch of snowfall is estimated to decrease latent scale scores by .0024 standard deviations (for the fifth grade at the proficient standard).   With the exception of the third grade, the estimated effects are all significant.   Once again, increased winter weather, in the form of inches of snowfall, is associated with reduced performance for all grades and proficiency levels.[28]

4.4   Final Estimates of the Effect of Additional Instructional Days

Because Colorado and Maryland's indirect least squares estimates refer to the effect of losing an instructional day, those estimates are multiplied by -1.[29] With this slight transformation in mind, Table 6 compares the estimates of the effect of an additional day of schooling for all three states across various grades and thresholds of proficiency. Rows with proportion variables are estimated using linear probability models, while rows denoted as latent refer to untransformed normit coefficients. All have similar qualitative implications: additional instructional days improve student performance. Most are highly significant, though there are some differences in magnitude.

Because the density of students varies across grades and academic standards, the best measures to compare across grades and states are probably the latent effects. The estimated effects derived from weather-related cancellations are in general larger than those from test-date changes. In Maryland, an additional day of schooling is estimated to improve test scores by as much as .016 standard deviations, while an additional day improves test scores by at as

---

[27]Because the normit function is not defined for proportions equal to zero or 1, these are replaced with small deviations, i.e. 0.01 and .99.

[28]This is somewhat different from Marcotte and Marcotte and Helmette, who find 3rd graders are most strongly effected by instructional days.

[29]This is because the fixed effect regressions refer to deviations from means, the estimated coefficents can refer to deviation above the mean (more snow days) or below (less snow days).   Linearity of the regression model allows this transformation.

much as .013 standard deviations in Minnesota. For Colorado, the largest estimate suggests an additional instructional day raises test scores by .039 standard deviations.

Several factors could explain these differences. First, the estimates reported for Colorado and Maryland use only cancellations in the first stage regression. If delayed starts and early releases (other events that disrupt class and reduce instructional time) are treated as cancellations in the first stage regressions, the final indirect least squares estimates decrease by 25 percent. Non-linearity could also play a role because of decreasing returns to instructional time (Minnesota had more variation than Maryland, which in turn had more than Colorado). Also the effect of additional school days could vary due to test difficulty, student ability, or teacher quality. A few large snow storms impacted Colorado in 2006/2007, which could have attenuated the first stage, and thereby biased the indirect least squares estimates away from zero. Furthermore, weather-related cancellations may reduce critical review time, whereas a moved test date allows teachers to reschedule their time to allow for proper review. These reasons suggest the estimates of instructional days' effect derived from weather-related cancellations could be considered as upper bounds.[30]

In Minnesota, bias could go in the other direction. Several of the test date changes postponed the test until after spring break. If students forget material while on vacation, the Minnesota estimates could understate the effect of additional day of instruction. Also the later dates may have allowed less time for post-assessment material. This could lead to spill overs reducing the amount of material learned before the fourth grade, which could potentially also affect test scores in the fifth grade. These factors suggest that the estimates due to changes in test-date administration could be thought of as a lower bound for instructional days' effect.[31]

---

[30] Teacher absences could be an additional concern (Miller et.al. 2007). Because teacher absences are excluded from my data (due to availability), the indirect least squares estimates would be upward biased. This supports the notion that the estimates due to weather related cancellation can be viewed as upper bounds.

[31] One additional factor that could play a role is absolute age, as students are either older or younger

Lastly, both identification strategies refer to the effect of a contemporaneous change in instructional days. In essence, they measure the temporary effects of increasing instructional days for a particular school year. If the school year were permanently longer, there could be positive spill-over effects. For this reason, the effect of a permanent increase in school year length could be greater than those estimated in this paper.

4.5   Robustness Checks

I proceed to investigate two robustness checks. As pointed out in the previous section, if the ability to remove snow is improving over time, the indirect least squares estimates would overstate the effect of additional instructional days. Another factor that could play a role is school attendance. If school is not cancelled when a snowstorm hits, students might miss school and fall behind their classmates. This creates bias as the original reduced form estimates of weather's effect on cancellations would also include the effect of weather on attendance, if there is one. These two sources of possible bias are investigated using additional data sources.

Parameter stability is an implicit assumption of the two sample indirect least squares estimates. If technology in snowfall removal has improved more snow will be required to cancel school, this would bias the indirect least squares estimates away from zero. This concern is likely to be most relevant for Maryland, as the reduced form data go back to the 92/93 school year, while the first stage is estimated for the 06/07 school year. Harford County School District in Maryland has maintained a rich history of weather-related cancellations. From September 1988 through today, they have recorded daily cancellation, delay, and early release information. In addition, total yearly cancellations have been recorded since 1975. These additional data sources offer two ways to test the structural stability of weather's re-

depending on the date of test administration. However students are only older when they take the test, not when they are learning the material throughout the year. For evidence regarding absolute age, see Bedard and Dhuey (2007).

22

lationship with cancellations. A high frequency analysis estimates the relationship between snowfall and daily cancellations for each school-year beginning in 1988. A low frequency approach will the effect of yearly snowfall on total cancellations for ten year windows, beginning with the 1974-1983 window and ending with 1998-2007 window. Figure 6 contains the estimated coefficients for both approaches.[32] Though the estimated relationship between snowfall and cancellations varies across years, it seems to be noise rather than a systematic trend. Also interestingly enough, the high frequency results from pooling across all years suggest that an additional inch of snow increases the probability of a cancellation by .11. This is similar to the earlier results for Maryland using all school districts with only the 06/07 school year.

A last robustness check explores whether snowfall impacts attendance. The hypothesis studied in this paper has concerned the number of days teachers have for instructing their students, not the number of instructional days students choose to attend. If snowfall causes truancy, the indirect least squares estimates will be biased. This is not because the first stage is invalid–rather it concerns the original structural model. Upward bias would occur as part of the reduced form effect is due to attendance, but the current indirect least squares estimates would attribute it all to cancellations. Beginning in 2005, the Colorado Department of Education has recorded and published the total percentage of hours missed, the percentage of hours missed and excused, and the percentage of hours missed and unexcused. The three measures are regressed on yearly snowfall and the number of days with snow greater than 4 inches in separate regressions–with the results contained in Table 7. There is not a statistical or practical relationship between snowfall and the total percentage of hours missed or the percentage of missed hours unexcused.

There is some evidence that greater snowfall increases the number of excused absences.

---

[32]The yearly regression could not be estimated in 1997, 1998, and 2002 as there were no cancellations, hence no variation in the depedent variable.

Although the effect is marginally significant, it is small in magnitude. For each day with snow greater than 4 inches, the proportion of hours excused increases by .0008. No matter how it is scaled, the effect is relatively small. In addition it is unknown when the days were missed. Because cancellations require make-up days in the summer, parents could be excusing their students from the make-up days at the end of the school year (because of previously arranged family vacations or other activities). So although there is some evidence of possible bias because winter storms cause excused absences, the correlation between snowfall and excused absences is small in magnitude and could be explained by scenarios that would not bias the results.

5   Conclusions

Prior research has been at odds over the effect of school inputs on student outcomes–both labor market and academic. I find evidence consistent with increased instructional days improving student performance. This supports Card and Krueger's findings that longer school years are associated with increased wages. Two different identification strategies are employed in calculating the effect of an additional day of schooling, taking advantage of exogenous variation in instruction due to both weather and state mandated shifts in test administration. Also, it is encouraging that the estimates are similar to those of Marcotte and Hemelt. This holds although the method used in Maryland is different along with additional data from Colorado. A different source of instructional day variation in Minnesota provide similar and even stronger results. Weather-related cancellations and test date shifts both offer statistically significant evidence that additional school days increase student performance.

The larger estimates suggest that 5 additional days of instruction would increase test scores by .15 standard deviations, while the smaller suggest it could improve test scores

by .05 standard deviations. It may be of use to compare these estimates to other policy interventions, such as decreasing the pupil-teacher ratio. Krueger finds that being in a small class increases a student's percentile ranking by roughly 4 points. Comparing these percentile effects to standard deviation shifts in scale scores offers rough comparison, and also requires a distribution and location. If the data were generated by a normal distribution and the student is at the median, increasing his or her percentile ranking by 4 percentile points implies a scale score shift of .1 standard deviations. If a student were at the 90 percentile, increasing by his or her percentile ranking by 4 points implies the scale score increased by .25 standard deviations. This is only a back-of-the-envelope comparison, but it seems that a couple weeks of additional school days is a reasonable substitute for smaller classes.

Although I find evidence of the potential benefits of extending the school year, this does not necessarily justify requiring all schools to do so. In part this is because the costs of lengthening the school year are not homogeneous across districts (due to air conditioning, teacher salaries, transportation). Thus, locations where it is expensive to lengthen their school year might optimally take advantage other policy interventions, such as reducing the pupil-teacher ratio. This would be consistent with efficient distribution of schooling resources.

In conclusion, my final estimates are consistent with more instructional days raising student performance. Because total instructional days in a school year (pre and post test administration) are fixed despite changes in weather-related cancellations or test-date administration, the estimates relate to the effect of an increase in instructional days. Permanently longer school years could have positive spill-over effects not accounted for by either estimation strategy. The results in this paper suggests longer school years can improve student performance, and perhaps increase human capital accumulated.
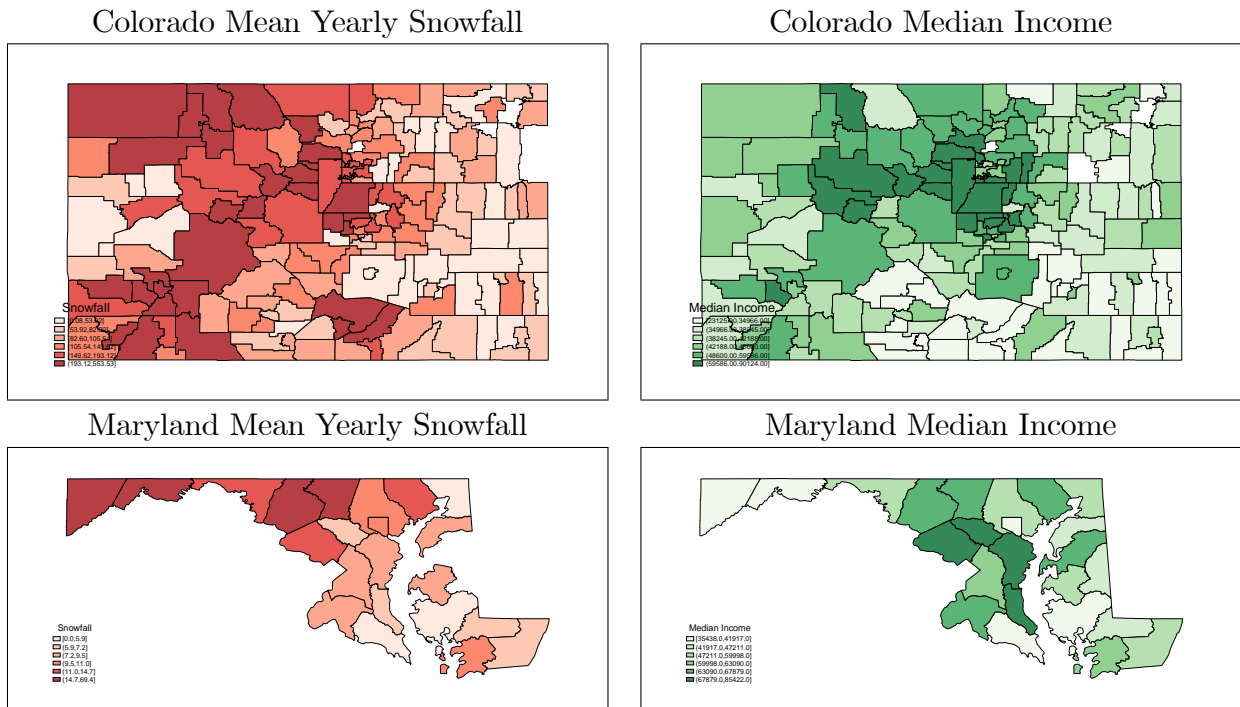
References

[1] Angrist, Joshua and Guido Imbens (1994). "Identification and Estimation of Local Average Treatment Effects." Econometrica, Mar., (1994).

[2] Angrist, Joshua D. and Alan B. Krueger. "Split Sample IV Estimates of the Return to Schooling." Journal of Business Economics and Statistics. Apr. (1995).

[3] Bedard, Kelly and Elizabeth Dhuey. "Is September Better than January? The Effect of Minimum School Entry Age Laws on Adult Earnings." Manuscript, (2007).

[4] Betts, Julian R. "The Impact of Educational Standards on the Level and Distribution of Earnings." The American Economic Review, Mar.(1998).

[5] Coleman, James S., E.Q. Campbell, C.J. Hobson, J. McPartland, A.M. Mood, F.D. Weinfeld, and R.L. York. Equality of Educational Opportunity. Washington, DC: U.S. Govt. Printing Office (1966).

[6] Eren, Ozkan and Daniel Mittlemet. "Time to Learn? The Organizational Structure of Schools and Student Achievement." Empirical Economics, (2007).

[7] Hanushek, Eric A. "Throwing Money at Schools" Journal of Policy Analysis and Management. Autumn (1981).

[8] Krashinsky, Harry. "How Would One Extra Year of High School Affect Academic Performance in University? Evidence from a Unique Policy Change." Manuscript. (2006)

[9] Krueger, Alan B. "Experimental Estimates of Education Production Functions" Quarterly Journal of Economics May (1999).

[10] Lee, Jong-Wha and Robert Barro. "School Quality in a Cross-Section of Countries." Economica, (2001)

[11] Madalla, G.S. "Limited-Dependent and Qualitative Variables in Econometrics." Cambridge University Press, (1993).

[12] Marcotte, Dave E. "Schooling and Test Scores: A Mother-Natural Experiment", Economics of Education Review, (Forthcoming).

[13] Marcotte Dave E. and Steven W. Hemelt. "Unscheduled School Closings and Student Performance." IZA DP No. 2923, July (2007).

[14] Margo, Robert A. "Educational Achievement In Segregated School Systems: The Effects of "Separate-but-Equal." American Economic Review. Sep. (1986)

[15] Miller, Raegen T. and Richard J. Murnane and John B. Willett. "Do Teacher Absences Impact Student Achievement? Longitudinal Evidence from One Urban School District." NBER Working Paper No. 13356.

[16] Pischke, Jorn-Steffen. "The Impact of Length of the Schooled Year on Student Performance and Earnings: Evidence From the German Short School Years." NBER Working Paper 9964, (2003).

[17] Wößmann, Ludger ."Schooling Resources, Educational Institutions and Student Performance: the International Evidence",. Oxford Bulletin of Economics and Statistics (2003)

-.5in-.5in

6 Appendices

6.1 Figures

Colorado Mean Yearly Snowfall

Colorado Median Income

Maryland Mean Yearly Snowfall

Maryland Median Income

6.2 Creation of Snowfall Variables

For both Maryland and Colorado, weather data was extracted from the National Climatic Data Center (NCDC) daily surface observations. In Maryland snowfall was taken as the average of snowfall observed in county, as districts and counties are the same. A few counties which did not maintain weather histories and were linked to the closest coop locations. Also days with missing observations were imputed using the closest coops higher and lower in elevation. In Colorado, school districts locations, in longitude and latitude, were extracted from the National Center of Educational Statistics. This was then used to determine the elevation of the school districts. Knowing both the latitude/longitude coordinates of the schools and their elevations, school districts were linked with the two nearest weather stations higher and lower. Then snowfall was computed the average of these nearby stations. The first stage estimates use the data available from the NCDC as of August 2007.

Figure 1

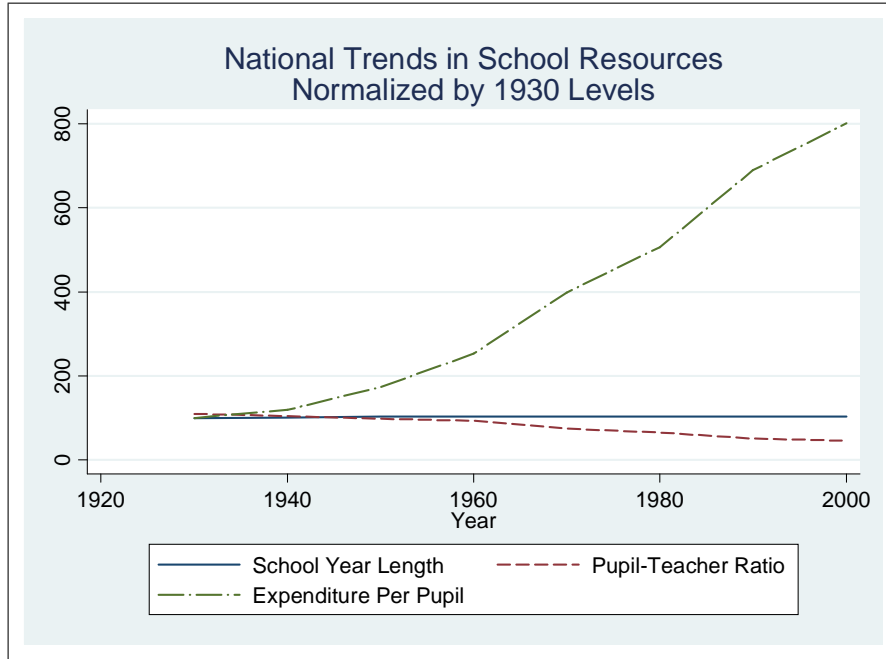National Trends in School Resources
Normalized by 1930 Levels

Figure 1 shows that on average in the United States, while the pupil teacher ratio has fallen and real expenditure per pupil has risen, school year length has remained relatively fixed.

Figure 2

A   B
Colorado District-Level Variation in Instructional Time

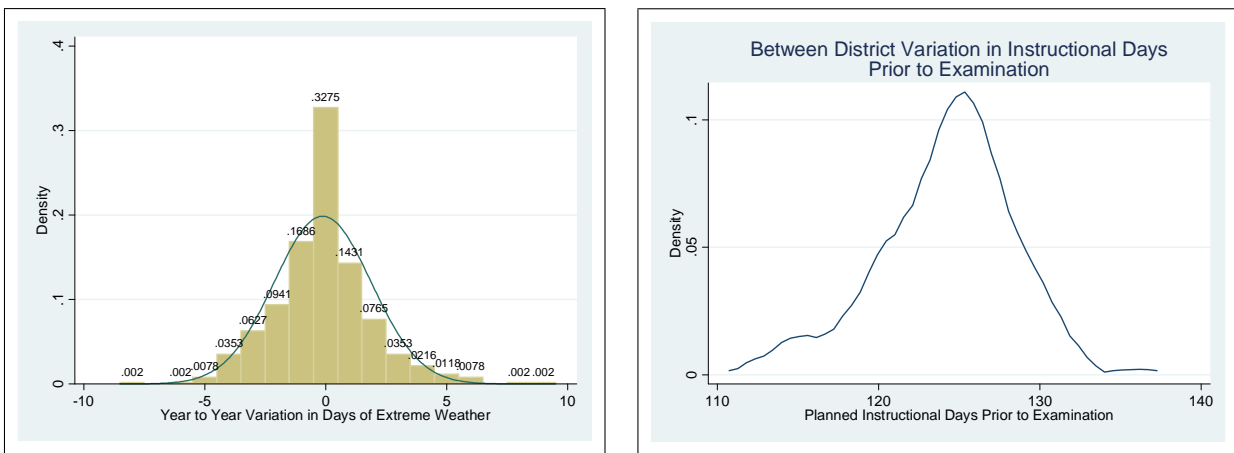Between District Variation in Instructional Days
Prior to Examination

Figure 2 shows that the variation of instructional days within school districts in Colorado is comparable to the between variation in planned instructional days in Colorado. This provides evidence there can be sufficient variation in instructional days to identify the effect comparing within school districts.

Figure 3
Trends in Minnesota Average Scale Scores and Instructional Days

A

B
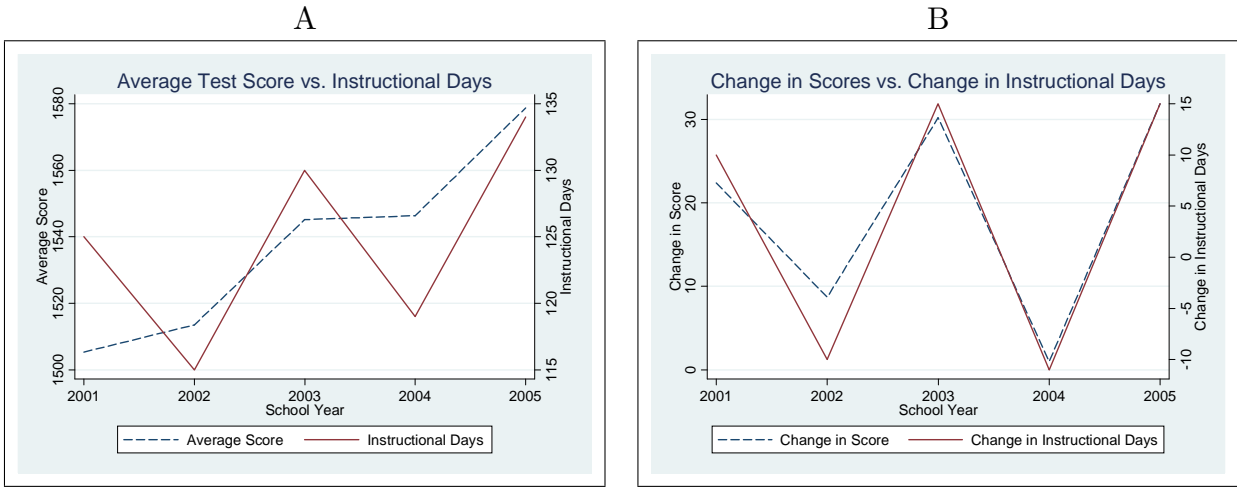
Figures 3 A and B show that the upward trend in average scale scores in Minnesota flattens out when the test is administered earlier, and steepens when exams are later.

Figure 4
Distribution of Minnesota School Average Scale Score Changes, by Date of Test

A

B
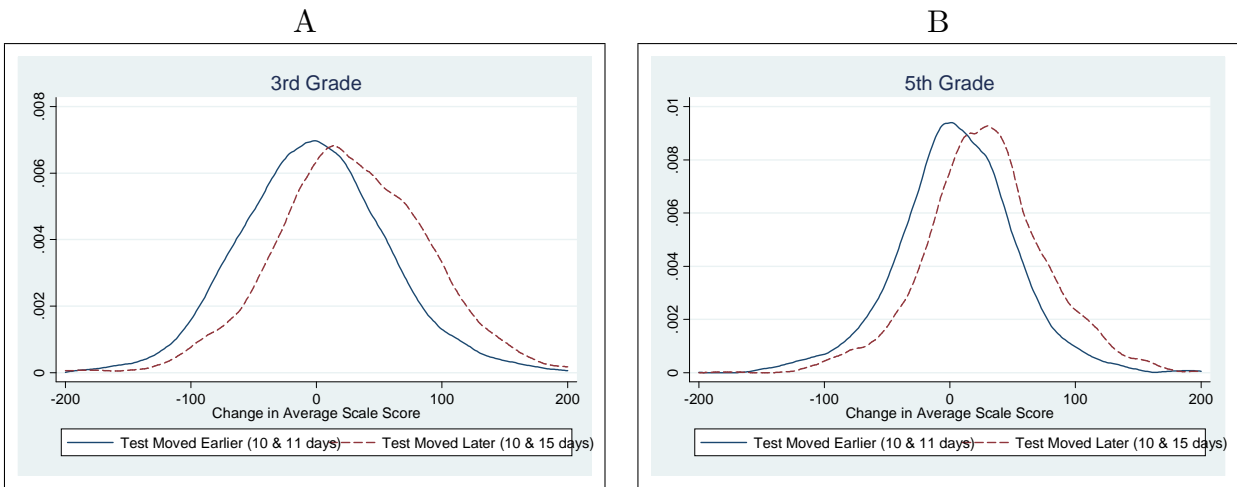
Figures 4 A and B demonstrate that the entire scale score (improvement) distribution is shifted to the right when exams are shifted later. This shows that the relationship observed in Figure 3 is not driven merely by outliers.

Figure 5
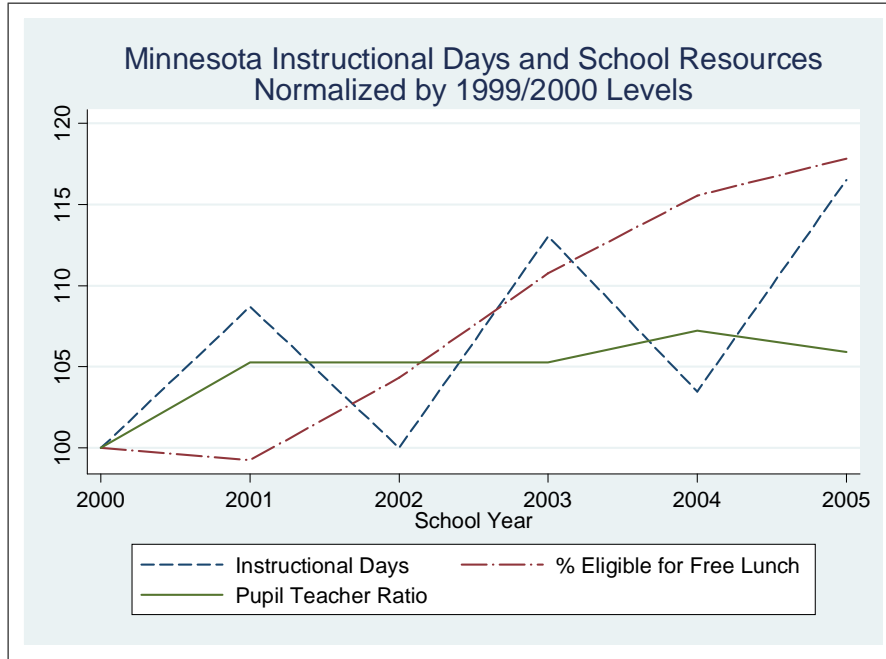
Figure 5 demonstrates that school characteristics such as the pupil/teacher ratio and the % of students receiving reduced price lunches change relatively smoothly compared to the changes in test administration date.

Figure 6
Historical Stability of First Stage, Harford County Maryland

Figure 6 shows that although there may be sampling variability using a particular school year to estimate the relationship between snow and cancellations, a systematic trend does not seem present.

31

Table 1
Correlation Between Snowfall and Resources

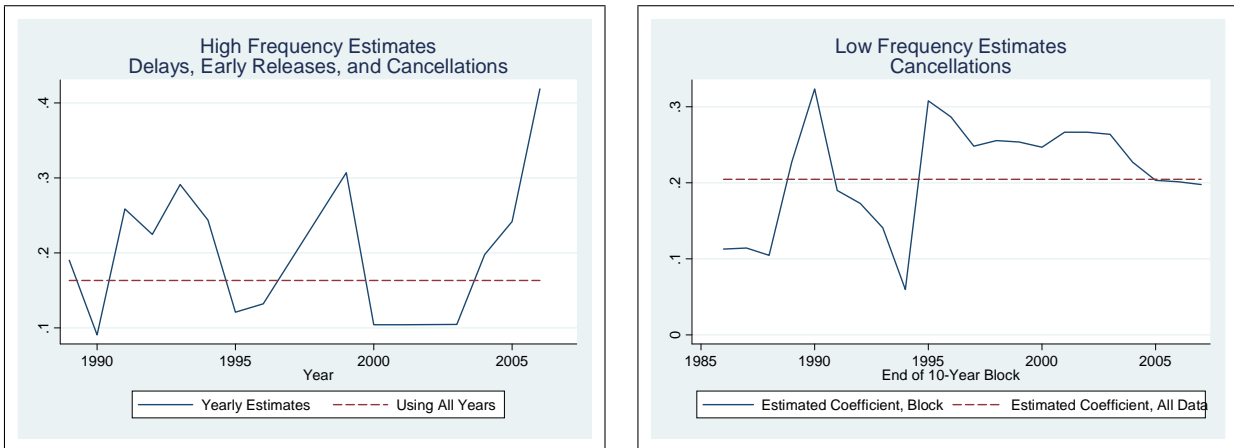| Year	99/00	00/01	01/02	02/03	03/04	Panel
Colorado						
Pupil Teacher Ratio	·	·	.28**	.17	.12	-.008
			(.14)	(.15)	(.08)	.159
% Reduced Price Lunch	·	·	-17.26***	-12.99***	-12.99***	1.79
			(4.62)	(4.11)	(4.11)	(3.034)
Fixed Effects	No	No	No	No	No	Yes
Year Dummies	No	No	No	No	No	Yes
F-Test	·	·	15.32***	7.92***	8.53***	0.17
Maryland						
Pupil Teacher Ratio	-.041	.054	.025	·	·	-.011
	(.043)	(.0493)	(.020)			(.018)
% Reduced Price Lunch	-29.87	-35.99	-12.98	·	·	1.027
	(35.74)	(34.67)	(14.40)			(14.05)
Fixed Effects	No	No	No	No	No	Yes
Year Dummies	No	No	No	No	No	Yes
F-Test	0.53	0.72	1.65	·	·	.17

Dependent Variable: Yearly Snowfall. All results clustered at district-year level.
***significant at 1%, ** significant at 5%, % significant at 10%. Parentheses indicate standard errors.

Table 1 establishes that within a cross section there is non-random selection of residence, particularly in Colorado. Controlling for fixed effects, changes in weather appear unrelated to changes in resources.

Table 2
Summary Statistics

Variable	Maryland	Colorado	Minnesota
% Advanced	.10 (.14)	.14 (.11)	.15 (.11)
% Proficient	.44 (.22)	.41 (.18)	.55 (.17)
% Partially Proficient	.	.72 (.16)	.71 (.15)
Snowfall	11.55 (13.55)	20.26 (10.80)	.
Pupil Teacher Ratio	16.5 (1.63)	16.3 (2.31)	16.0 (1.81)
% Eligible for Free Lunch	.33 (.25)	.37 (.24)	.33 (.22)
Teaching Assistants/1000 Students	10.91 (3.63)	17.97 (7.82)	.
Average Teacher Experience	.	.	16.85 (4.91)
Median Wealth Per Student	223, 785 (84, 966)	.	.
% Title One	.15 (.32)	.	.
% Limited English Proficient	.020 (.04)	.	.054 (.11)

Parentheses indicate standard errors.

Table 3
First Stage: Effect of Snowfall on Cancellations

	High Frequency		Low Frequency	
Weather Measure	Colorado	Maryland	Colorado	Maryland
Snowfall	.052***	.16***	.013	.14**
	(.004)	(.03)	(.010)	(.056)
t-statistic	11.52	4.49	1.33	2.5
f-statistic	132.80	20.15	1.76	6.23
# Days Snow > 4 inches	.37***	.	.27**	.
	(043)	.	(.109)	.
t-statistic	8.66	.	2.68	.
f-statistic	75.01	.	7.18	.
# Days Snow >1 s.d.	.37***	.	.23**	.
	(.044)	.	(.11)	.
t-statistic	8.44	.	2.05	.
f-statistic	71.27	.	4.21	.
Fixed Effects	yes	yes	no	no
Number of Observations	17716	2681	107	18

***significant at 1%, ** significant at .5%, * significant at 10%

Results clustered at district-year level. Parentheses indicate standard errors.

Table 3 contains the estimated relationship between weather and cancellations. Snowfall–in various measures– is strongly related with cancellations.

Table 4

Colorado Reduced Form

Effect of an Additional Day with Snow>4 inches on Performance

Dependent Variable	Grade 8
Response Probabilities	
% Partially Proficient	−.0056**
	(.0024)
% Proficient	−.0030
	(.0032)
% Advanced	−.0043**
	(.0021)
Shift of Scale Scores in Standard Deviations	
Latent Partially Proficient	−.015*
	(.008)
Latent Proficient	−.0053
	(.009)
Latent Advanced	−.0073
	(.009)
Fixed Effects	Yes
Controls	Yes

***significant at 1%, ** significant at 5%, * significant at 10%

All results clustered at district-year level. Parentheses indicate standard errors.

Table 4 shows that additional days with snow exceeding four inches is associated with lower performance in Colorado.

Table 5

Reduced Form Maryland

Effect of An Additional Inch of Snowfall on Performance

Dependent Variable	Grade 3	Grade 5	Grade 8
Response Probability			
Proportion Advanced	−.000052	−.00048***	−.00046***
	(.00013)	(.00015)	(.00016)
Proportion Proficient	−.00050	−.00073**	−.00053**
	(.00038)	(.00032)	(.00022)
Shift of Scale Scores in Standard Deviations			
Latent Advanced	−.00049	−.0027***	−.0021***
	(.00098)	(.00081)	(.00068)
Latent Proficient	−.0018	−.0024**	−.0014**
	(.0012)	(.0010)	(.0006)
School Fixed Effects	Yes	Yes	Yes
Controls	Yes	Yes	Yes

***significant at 1%, ** significant at 5%, * significant at 10%

All results clustered at district-year level. Parentheses indicate standard errors

Table 5 shows that additional inches of snowfall in Maryland reduce performance, regardless of threshold or grade studied.

Dependent Variable	Colorado	Maryland			Minnesota	
	Grade 8	Grade 3	Grade 5	Grade 8	Grade 3	Grade
Response Probability						
Proportion Advanced	.011** (.0055)	.00038 (.00081)	.0030*** (.0011)	.0029** (.0012)	.0028*** (.00027)	.0011** (.0001(
Proportion Proficient	.0081 (.0084)	.00031 (.0025)	.0045** (.0022)	.0033** (.0015)	.0045*** (.00025)	.0031*' (.0002(
Proportion Partially Proficient	.014** (.0069)	.	.	.	.0022*** (.00023)	.0020*' (.0001'
Shift of Scale Scores in Standard Deviations						
Latent Advanced	.019 (.024)	.003 (.0063)	.016*** (.0062)	.013** (.0051)	.013*** (.0012)	.0042*' (.0006'
Latent Proficient	.013 (.024)	.011 (.0077)	.015** (.0069)	.0090** (.0042)	.012*** (.00069)	.0089*' (.0005!
Latent Partially Proficient	.039* (.021)	.	.	.	.0074*** (.00071)	.0070*' (.0006:
Percentage Shift in Scale Scores						
Log Average Score	.	.	.	.	.00082*** (.000085)	.00062* (.00006
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes		Yes		Yes	
Trend	Year Dummies	Year Dummies			Quadratic By Schoo	

***significant at 1%, ** significant at .5%, * significant at 10%.

All regression clustered at district-year level. Parentheses indicate standard errors.

Table 6 presents the estimates of the effect of an additional day of instruction on performance. The results all share the same sign (improvement), and many are highly significant. Estimates derived from weather-related cancellations are relatively larger, while those due to test date shifts are the most statistically significant.

Table 7
Attendance and Snowfall, Colorado

	% Hours Missed	% Hours Missed Unexcused	% Hours Missed Excused
Snowfall	$-.00022$	$-.0001$	.00007
	(.00014)	(.0001)	(.00007)
Days w/Snow>4	.0006	$-.0003$	.00085*
	(.001)	(.0006)	(.0005)
Fixed-Effects	Yes	Yes	Yes
Mean	.057	.0135	.044

***significant at 1%, ** significant at .5%, * significant at 10%

All results clustered at district-year level

Table 7 contains estimates of the effect of snowfall on attendance in Colorado. For most measures of attendance and weather, snowfall appears unrelated to truancy. The one measure which exhibits statistical significance, is small regardless of how it is scaled.