Journal of Studies in Education
ISSN 2162-6952
2017, Vol. 7, No. 4

Student Absences and Academic Achievement

Eric R. Eide, Ph.D. (Corresponding author)

Department of Economics, Brigham Young University, 130 FOB, Provo, Utah 84602, USA

Tel: 1-801-422-4883   E-mail: eide@byu.edu

Ian Fillmore, Ph.D.

Department of Economics, Washington University in St. Louis, St. Louis, MO 63130, USA

Tel: 1-314-935-9283   E-mail: ianfillmore@gmail.com

Mark H. Showalter, Ph.D.

Department of Economics, Brigham Young University, 130 FOB, Provo, Utah 84602, USA

Tel: 1-801-422-4645   E-mail: showalter@byu.edu

Received: Nov. 1, 2017   Accepted: Nov. 22, 2017   Published: November 22, 2017

doi:10.5296/jse.v7i4.12119   URL: https://doi.org/10.5296/jse.v7i4.12119

Abstract

We use nationally representative data on students in the United States who are enrolled in grades 7-12 together with the statistical techniques of Ordinary Least Squares (OLS), Quantile Regression (QR), and Instrumental Variables (IV) to explore how various types of student absences are related to test scores. Using a variable that measures total absences, the OLS results suggest that missing two weeks of school is associated with a one-tenth standard deviation reduction in math score. Estimates vary widely when allowing the relation to differ by type of absence and by quantile of the conditional test score distribution. Using absence due to injury as an IV for total absence leads to larger coefficients than OLS.

Keywords: Student Absences, Academic Achievement, Quantile Regression, Instrumental Variables, Economics of Education

1. Introduction

In the current climate of educational accountability where the national spotlight is on student achievement, education policymakers and researchers often overlook the importance of getting students into the classroom. Policy advocates tend to focus on reform measures such as class size reduction, increased spending, and teacher quality. However, these policies will be ineffective at improving student outcomes if students are not in the classroom.

Students miss school for a variety of reasons, including illness, injury, truancy, suspension, and expulsion. Students who are repeatedly absent from school tend to struggle academically and in other areas. There is a strong correlation between absences and negative outcomes such as poor academic performance and involvement in risky behavior. For example, Ginsburg et al. (2014) find that students who miss three or more days of school in the prior month had lower average math and reading scores than students with fewer absences. Sundius and Farneth (2008) find that among older children, those who are frequently absent are more likely than their regularly attending peers to perform poorly in academics, to drop out of school, to use alcohol and drugs, and to be involved in the juvenile justice system.

1.1Literature Review

Studies in the Economics of Education literature that examine the relation between student absences and student achievement often frame the research question and methodology within an “education production function” approach, where student achievement (e.g. test scores, grades, graduation rates, etc.) is the measured “output” of the education production process and factors that affect student achievement are the “inputs” (e.g. family background, school spending, teacher characteristics, etc.). Researchers such as Lamdin (1996), Borland and Howsen (1998), and Coates (2003) have argued that student absences or variants of absences such as instructional time should be included as inputs into the education production function. Gottfried (2009) takes this idea further by disaggregating student absences into excused and unexcused, thereby recognizing that students miss school for a variety of reasons, and these different types of absences may have different effects on students.

A few economics studies have attempted to use statistical techniques that will identify a causal (as opposed to correlational) relation between student absences and academic achievement. Each uses different data sources and methodology, but each confirms that increasing absences leads to lower achievement. For instance, Goodman (2014) uses data from Massachusetts, Aucejo and Romano (2016) data from North Carolina, Gottfried (2010) data from Philadelphia, and Gershonson, et al. (2017) national data on primary school students, and each finds an inverse relation between absences and achievement across a variety of statistical models.

In addition to the direct costs to the student, absences from school also carry indirect costs to teachers, other students, and taxpayers. When a student misses class, a teacher may spend time helping that student get up to speed on the missed material leading to less time spent with the absent student’s classmates (Goodman, 2014). Further, frequently absent students may cause more behavioral problems, leading to poorer performance of their classmates.

Research by Gottfried (2011) found negative effects of peer unexcused absences on student reading and math scores. Additionally, if more absences lead to a greater likelihood of the student repeating a grade, that additional year is paid for by the taxpayer (Eide and Showalter, 2001).

The studies mentioned above focus on data from the U.S. For examples based on data from other countries, see Bos, Ruijters, and Visscher (1992) who study Dutch schools; Hancock, et al. (2017) who study Western Australian schools; and Arulampalam, Naylor, and Smith (2012) who study university students in the UK.

1.2 Research Objectives

We view the primary value-added of this paper as 1) exploring how different types of absences (illness, injury, skipping) are related to math and reading achievement across the conditional test score distributions; 2) adding to the evidence on the causal relation between absences and student achievement; and 3) using national data on 7th to 12th graders to provide a broader perspective than geographically-specific data used in most previous studies on this age group.

2. Data

We use the Child Development Supplement (CDS) of the Panel Study of Income Dynamics (PSID). The PSID is a nationally representative panel of individuals and their families.(Note 1) Begun in 1968, sampled individuals and families provide information on family composition, wealth, earnings, expenditures, employment, and a variety of other data. In 1997, the CDS was initiated by supplementing the PSID with additional information on families with children ages 0-12. The intent was to gather information to add to our understanding of the early formation of knowledge and skills.

The initial sampling of the CDS selected 2,705 families from the PSID. 2,394 families participated (88 percent), providing information on 3,563 children ages 0-12. The information from this initial survey is known as CDS-I. A follow-up survey was done in 2002-2003 (CDS-II) on the CDS-I families. CDS-II includes a rich set of variables describing the home and learning environment of the child: standardized test scores in multiple subjects, behavioral assessments, learning resources, time use, and health status are a few examples. It also gives detailed information on the primary caregiver.

2,017 families (91 percent) were successfully interviewed in CDS-II, including 2,908 children or adolescents ages 5-18. For this study, we use CDS-II data, along with some background information that was gathered in the CDS-I round of interviews. We also incorporate family background data from the 2001 PSID interviews. The CDS is ideal for answering our research question due to the rich amount of information it contains on family background, student health, absences, and test scores. Further, the vast majority of studies on student absences for students in grades 7-12 are based on geographically-specific data and so these nationally representative data provide information on the relation between absences and test scores for the US more generally. While these data are several years old, the CDS is the only available nationally representative data set that contains the required information for our analyses, namely, specific types of absences, standardized test scores, and detailed family background variables.

2.1 Dependent Variables

We are interested in the relation between absences and student performance. Our unit of observation is a child and our dependent variables are standardized test scores in math and reading. (Note 2) Math and reading are two subjects emphasized in school accountability systems so these measures have important policy implications.

2.2 Independent Variables

Absences are measured both as the total number of days missed and also as days missed by absence type (due to injury, illness, skipping). The CDS does not include a variable for total absences. (Note 3) We therefore impute total absences from the three other variables available in the CDS-II: 1) absences due to illness in the past 12 months, 2) absences due to injury in the past 12 months, and 3) number of times the child skipped school in the past six months without permission. The respondent on the first two questions is the child's primary caregiver (usually the mother), while the third question is answered by the child.(Note 4) To clarify the meaning of each of these variables, we include the wording of the questions in Appendix1. To obtain total absences, we add these three variables together, weighting the third according to the month in which the question is asked. For example, the number of days the child skipped school in the past six months will be very different if the question is asked at the end of summer break as opposed to just before summer break. Details on how these weights are calculated are in Appendix2.(Note 5)

Our total absence variable has a mean of 3.86 and a standard deviation of 6.53 (Table 1). Thirty percent of the sample has no absences and eighty percent of the sample has six or fewer absences. The 95th percentile is 14 absences in the school year. The individual absence variables show considerable differences in means (Table 1). The average number of absences due to illness is 2.59, due to skipping is 1.00, and due to injury is 0.27.

We include a number of control variables that measure the home environment of the child, which is strongly correlated with student achievement. Including these variables in our regressions allows us to separate out the influence of absences from observable family background circumstances.

Table 1. Descriptive statistics

	Obs.	Mean	Std. Dev.	Min	Max
Math Score	1,053	0.00	1.00	-2.78	2.90
Reading Score	1,049	0.00	1.00	-3.84	2.93
Total Absences	1,053	3.86	6.53	0	91.8
Injury Absences	1,052	0.27	1.18	0	18
Illness Absences	1,052	2.59	4.55	0	90
Skipped Absences	1,053	1.00	4.25	0	56.25
Age	1,053	15.23	1.82	12.0	19.1
Gender (1=Male)	1,053	0.48	0.50	0	1
Black	1,053	0.43	0.49	0	1
Hispanic	1,053	0.07	0.25	0	1
Other	1,053	0.01	0.11	0	1
North Central	1,015	0.28	0.45	0	1
South	1,015	0.28	0.45	0	1
West	1,015	0.14	0.35	0	1
Years of Education-H. of H.	993	12.93	2.76	0	17
Log(Family Income)	1,053	10.76	1.16	0	14.5
Family Size	1,053	4.28	1.31	1	13
College-Bound (1=Yes)	1,035	0.69	0.46	0	1

Note: All statistics except for Reading Score are from the sample used to compute OLS Math regression. Zero years of education for head of household appear to be for immigrant families (3 observations). There is one observation with a family size of 1--an 18-year-old female. There was one 19-year-old in the sample despite the description by the data collectors that the maximum age is 18. Math and Reading scores have been normalized based on the regression samples.

We include binary measures for gender (male), whether the child wants to graduate from college (Note 6), race/ethnicity (black, Hispanic, and other, with white omitted), and region of the country (north central, south, and west, with northeast omitted). We use continuous measures for child's age, head of household's years of schooling (Note 7), log of family income in 2001(Note 8), and family size. Low income, minority children are overrepresented in the sample due to the sampling strategy of the PSID; however, we control for family income and race/ethnicity in the regressions so the estimates should not be affected by the sample composition. Table 1 presents descriptive statistics of the main analysis variables.

We include in the estimation sample students who are at least 12 years old and who are enrolled in grades 7-12 at the time of the CDS-II survey. We do not include younger children because the absence questions were not asked to elementary school students. To be included in the sample, students must have data on both absences and test scores. We drop students who were enrolled in special education. For control variables with missing values, we include a dummy variable equal to one if the observation has a missing value and zero otherwise.

3. Methods

We employ three statistical techniques to estimate the relation between absences and student achievement.

3.1 Ordinary Least Squares

Ordinary Least Squares (OLS) estimates the relation between absences and the conditional mean of the test score distribution; that is, the mean of the test score distribution conditional on the control variables. This helps us see, on average, how costly absences are to students in terms of learning.

3.2 Quantile Regression

Quantile Regression (QR) allows us to estimate the relation between absences and test scores at different points (or "quantiles") of the conditional test score distribution, e.g. at the 0.25 quantile (lower achieving students), 0.50 quantile (median achievers), and 0.75 (higher achieving students). This suggests how absences differentially influence relatively high and low performing students. Quantile regression therefore provides a broader set of estimates of the relation between test scores and absences than OLS, which only estimates the relation at the conditional mean.

3.3 Instrumental Variables

Instrumental Variables (IV) is a statistical technique commonly used in economics to estimate a causal effect of an independent variable on the dependent variable in instances where OLS would only estimate a correlational relation. (Note 9) In our case, we attempt to sort out the extent to which there is a causal versus a correlational relation between absences and test scores. More precisely, do more student absences *cause* lower test scores, or is it simply that students who tend to be absent more on average also tend to be lower achieving students?

The challenge with estimating the causal effect in our context is that absences to some extent are subject to the choice of the student, and if there are unobserved aspects of the student (e.g. motivation to learn) that are correlated with both absences and achievement then it is difficult to identify the independent effect of absence on test scores from the influence of the unobserved student characteristics on test scores.

For IV to produce reliable estimates, a variable called an "instrument" must be identified which is correlated with absences, but which is also uncorrelated with unobserved factors that affect both absences and test scores (such as a student's motivation to learn). The variation in the instrument is then used, through its correlation with total absences, to identify the causal effect of absences on test scores in the statistical analyses.

3.4 Estimation Models

We first estimate the following model using OLS and QR:

$$Y = X\delta + A\beta + \varepsilon \tag{1}$$

Where Y is a vector containing test scores, X is a matrix of control variables, A is a vector containing absence variables, and ε is the error term which captures unobserved factors that influence test scores but which are not included in the regression model (e.g. a student's motivation to learn). The parameter δ measures the relation between the control variables and test scores, and β measures the relation between absences and test scores. We are primarily interested in the estimates of β.

Ordinary Least Squares estimates of the association between absences and test scores may be biased for two reasons. First, absences may be measured with error because the answers to the survey questions are subject to recall error and we have no information on missing school due to suspension or expulsion. (Note 10) Second, because missing school can be a matter of choice on the part of each student, absences may be correlated with unobserved student characteristics. This will make the absence variables correlated with the error term ε which means OLS estimates will be biased.

To obtain valid estimates of the causal effect of absences on test scores, we employ an IV approach. We use the number of absences due to injury during the past 12 months as reported by the primary caregiver as an instrument for total absences. This instrument is based on the idea that injuries are likely to be random events thus creating a quasi-experiment.(Note 11) In order to be valid, the injury instrument should meet two conditions: 1) be correlated with total absences, and 2) be uncorrelated with the unobserved factors that affect both absences and test scores, i.e. uncorrelated with the error term ε. We address these two conditions in section 4.

We estimate our IV model using the Two-Stage Least Squares approach, given by Equations (2) and (3):

$$\hat{A} = X\hat{\theta} + I\hat{\gamma} \tag{2}$$

$$Y = X\delta + \hat{A}\beta + \varepsilon \tag{3}$$

Equation (2) is called the "first stage" equation and gives the predicted value of total absences ($\hat{A}$) as a function of I, the instrumental variable. Equation (3) is the primary equation of interest. It is called the "second stage" equation, and the important aspect of Equation (3) is that, relative to Equation (1), $\hat{A}$ is substituted for A. By construction $\hat{A}$ is uncorrelated with ε because variation in $\hat{A}$ is driven by the exogenous variation in I. Instrumental Variables estimates are obtained by OLS estimation of Equations (2) and (3). The estimated value of β from Equation (3) is the IV estimate for the causal effect of absences on test scores.

4. Results

4.1 Ordinary Least Squares Results

Our first set of results comes from OLS estimation of Equation (1). A summary of the key results is listed in Table 2. Column (1) presents results for math and column (5) results for reading. The reported standard errors are clustered at the household level. Focusing first on results for total absences, found in row (1), we see the coefficient in the math regression is

-0.011 and is statistically significant at the 5 percent level. This implies that missing 10 days of school (or two weeks) is associated with a decline in expected math score of just over 0.10 standard deviations. The result for reading is smaller, -0.005, and statistically insignificant.

While understanding the relation between absences and test scores is helpful in understanding the costs of missing school, students can miss school for a variety of reasons. To gain a clearer understanding of how different types of absences are related to math and reading scores, in columns (1) and (5) of the lower part of Table 2 we present OLS estimates where the total absence variable is replaced with three separate indicators for absence type. The absence type coefficients in column (1) come from the same regression, as do the analogous estimates for reading in column (5). The math results show a large and statistically significant association between being absent due to injury and math scores, a modest relation between skipping school (Note 12) and math scores, and no association between absence due to illness and math scores. For reading, there is a large and statistically significant coefficient for the injury absence variable, but the other coefficients are effectively zero both in magnitude and significance. These findings highlight something that was masked with the total absence variable: the association between absence and math and reading scores differs considerably according to the reason for the absence, with absences due to injury having the highest correlation.

The OLS estimates provide information about the relation between absences and student achievement on average, that is, at the conditional mean of the test score distributions. We next turn to quantile regression to see how the relation differs for lower and higher performing students.

Table 2. OLS and quantile regression results

	Math				Reading			
		Quantile Regressions				Quantile Regressions		
	OLS	0.25	0.50	0.75	OLS	0.25	0.50	0.75
Absence type	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total	-0.011**	-0.014**	-0.008*	-0.007	-0.005	-0.006*	-0.009	-0.001
	(0.004)	(0.007)	(0.004)	(0.006)	(0.004)	(0.003)	(0.006)	(0.004)
Injury	-0.058***	-0.041**	-0.043	-0.073***	-0.047***	-0.032*	-0.045***	-0.04***
	(0.015)	(0.018)	(0.028)	(0.016)	(0.017)	(0.018)	(0.017)	(0.014)
Illness	-0.005	-0.006	-0.005	-0.002	-0.002	-0.005	-0.009	0.0002
	(0.006)	(0.011)	(0.011)	(0.008)	(0.006)	(0.003)	(0.007)	(0.009)
Skipping	-0.013**	-0.018	-0.007*	-0.015	-0.004	-0.0006	0.002	-0.002
	(0.006)	(0.015)	(0.004)	(0.011)	(0.006)	(0.009)	(0.007)	(0.003)

Note: For total absences, each coefficient comes from a separate regression. For the absence type indicators, each column comes from the same regression. Regressions include controls for age, gender, race/ethnicity, region, college bound, head of household education, log family income, family size, and indicator variables for missing observations. Standard errors (in parentheses) for quantile regression are robust to non-i.i.d. errors, and for OLS are corrected for clustering at the household level. *** denotes statistical significance at the 0.01 level; ** denotes 0.05; * denotes 0.10.

4.2 Quantile Regression Results

To examine how the association between absences and test scores may change across the conditional distribution of test scores we estimated the models using quantile regression at the 0.25, 0.50, and 0.75 quantiles. The math results are in Table 2, columns (2) – (4), and reading results in columns (6) – (8). Starting with the results for total absences in row (1),the math coefficient estimates are -0.014 at the 0.25 quantile and -0.008 for the median, which suggests that increased absences are disproportionately associated with lower math scores for the lowest performing students. For reading, there is a significant association of -0.006 for those at the bottom 0.25 quartile. Taken together, the quantile regression estimates based on the total absence variable suggest missing school is associated with significantly lower math and reading scores for students in the bottom half of the respective tests score distributions.

In the lower part of Table 2 we present quantile regression estimates using the indicators for absence type. A notable finding is that the strongest correlations for both math and reading are for the injury absence variable. For math, there are significant coefficients at the 0.25 and 0.75 quantiles, and for reading at the 0.25, 0.50, and 0.75 quantiles. The coefficients are large, ranging from -0.032 for the 0.25 quantile for reading up to -0.073 for the 0.75 quantile for math. One reason why the injury coefficients are large could be that injuries occur randomly compared to illness and skipping school. Hence, there may be less discretion on the part of the student and parent about missing school due to an injury compared to the other types of absences. For example, students who skip school may choose to skip on a low productivity day when they believe there is not important academic material or testing that may occur, whereas for injuries a student may be unable to attend class regardless of how important the day is.

The estimation results discussed thus far have provided correlational evidence on the relation between absences and test scores at both the conditional mean and at various quantiles of the conditional test score distribution, as well as how these relations vary by type of absence. We next turn to estimation of the causal effect of absences on test scores using IV regressions.

4.3 Instrumental Variable Results

Our instrument is the number of absences due to injury in the past 12 months. Our assumption is that injuries occur on an unpredictable basis, and this randomness overcomes the correlation of the absence variable with the error termε. To provide some evidence that injuries are random, in Table 3 we provide sample means of some of the covariates broken down by categories for number of absences due to injury and number of absences due to illness. If absences due to injury are random, then we would expect to see similar means in the covariates across the categories for absence frequency. However, if there is something systematic about the number of injuries, and they are correlated with observable variables, then we would expect to see different covariate means for different absence frequency categories. Note that we provide more categories for the illness absences because they account for the majority of student absences. Table 3 shows two panels, one for absences due to injury and one for absences due to illness. Comparing the two panels it is clear that there is more variation in means for the illness absences relative to injury absences. For example, the percentage of students who are black in each injury absence category is roughly the

Table 3. Average differences in characteristics by absence category

	Injury Absences			Illness Absences			
	0 Abs	1-2 Abs	3+ Abs	0 Abs	1-4 Abs	5-9 Abs	10+ Abs
Black	0.42	0.47	0.45	0.55	0.39	0.25	0.33
Hispanic	0.07	0.03	0.08	0.09	0.05	0.08	0.06
LogFamily Income	10.76	10.83	10.74	10.62	10.83	10.96	10.65
Observations	957	60	38	379	481	142	51
Head of Household Educ.	12.95	12.81	12.53	12.49	13.14	13.11	13.49
Observations	905	54	36	352	459	135	47

same, whereas it varies from 0.55 to 0.25 for illness absence. Similar findings are apparent for log family income and head of household education. While our main concern is with unobservable factors that influence both absences and test scores, which Table 3 doesn't address because it is based on observed factors, it nevertheless provides suggestive evidence that injury absences are random.

Table 4presents the first stage and second stage IV results. Columns (1) and (3) show results where the instrument is the number of injury absences, and columns (2) and (4) are results where the instrument is composed of two dummy variables, one representing 1 or 2 absences and the other 3 or more absences (omitted category is no absences). The second stage results are in row (1), and the first stage results are in the lower rows. We only present the first stage results for the math sample because the results based on the reading sample are nearly identical.

Looking first at the results based on the "number of injury absences" instrument we see the first stage coefficient is significant at the 1 percent level. This suggests the instrument is correlated with total absences, as IV requires. The second stage estimates in columns (1) and (3), representing the causal effect of total absences on test scores, show the coefficient in the math regression is -0.045and in the reading regression is-0.036, both significant at the 1 percent level. The IV estimates are substantially larger than the OLS estimates for total absence and suggest a large effect on test scores. The combination of OLS and IV results has an interesting implication. Absences per se, as illustrated by the OLS results, have a relatively small association with math test scores and essentially no relation with reading scores. An explanation consistent with these findings is that students pick low productivity days for missing school (at least for truancy and to some extent illness). However, as the IV results suggest, a random absence is more costly.

Table 4. Instrumental variable results

	Math		Reading	
Instrument:	(1)	(2)	(3)	(4)
	#Injury Absences	#Injury Abs. Dummies	#Injury Absences	#Injury Abs. Dummies
Second Stage Results:				
Total Absences	-0.045***	-0.046**	-0.036***	-0.043***
	(0.013)	(0.019)	(0.012)	(0.016)
First Stage Results (Math only)				
#Injury absences	1.319***			
	(0.165)			
Binary instruments:				
1 or 2 injury absences		2.093**		
		(0.852)		
3 or more injury absences		6.876***		
		(1.063)		
F Test		25.99***		

Note: First stage results are presented only for math sample because results for reading sample are nearly identical. Regressions include controls for age, gender, race/ethnicity, region, college bound, head of household education, log family income, family size, and indicator variables for missing observations (head of household, college-bound, and region). Standard errors corrected for clustering at the household level are included in parentheses. *** denotes statistical significance at the 0.01 level; ** denotes 0.05; * denotes 0.10.

To account for a possible non-linear effect of injury on total absences, Table 4 also includes in columns (2) and (4) the results of the dummy variable specification of the injury absence instrument. This regression specification gives similar results to the continuous "number of injury absences" specification: strong first stage results based on an F-test of the instruments (F=25.99 with 2 and 880 degrees of freedom with a p-value of approximately 0.00) and the statistically significant second stage estimates of -0.046 for math and -0.043 for reading.

Our estimates are generally in-line with other studies that estimate the relation between student absences and achievement. Goodman (2014) uses data from Massachusetts students in grades 3-8 and 10 and finds that a one day increase in absence leads to a 0.01 standard deviation decrease in math score based on fixed effect models, and a 0.05 standard deviation decrease when using bad weather as an instrument for school closure. The magnitudes of these estimates are comparable to ours. Aucejo and Romano (2016) find qualitatively similar but smaller estimates. They use data on North Carolina elementary school students in grades 3-5 together with fixed effects models and find that a 10 day increase in absences leads to a 5 percent of a standard deviation reduction in math score and 2.9 percent of a standard deviation reduction in reading. Based on Philadelphia school district data, Gottfried (2010) uses geographical distance that an elementary student lives from school as an instrument for attendance and finds a positive effect of attending school on achievement, with IV estimates larger than OLS estimates, similar to what we find.

5. Conclusion

In this paper we explore the relation between absences and test scores using national data for US students in grades 7-12.On balance across the various models, we find that missing school is associated with significantly lower math and reading test scores, particularly in the bottom half of the test score distributions and particularly for injury-related absences.

Taken together, these findings are useful to education administrators and teachers because they suggest for whom absences may be most detrimental. It also suggests that schools should not only collect data on total absences (e.g. who is absent on a particular day) but also the reason for the absence so that a plan can be developed to help the student get caught up, especially during periods of extended absence. In future work, it would be useful to explore in more detail what types of absences (illness, injury, truancy) are most prevalent among students from different age and demographic groups, and how absences are associated with the educational performance of each of these sub-groups. It would also be interesting to understand the link between absences in school and job performance later in an individual's life. These analyses would help complete the picture of how the human capital disruption associated with absences is related to both short run and long run outcomes.

Acknowledgement

We thank David Sims for comments on an earlier draft of this paper.

References

Arulampalam, W., Naylor, R., & Smith, J. (2012). Am I Missing Something? The Effects of Absence from Class on Student Performance. *Economics of Education Review*, *31*(4), 363-375. https://doi.org/10.1016/j.econedurev.2011.12.002

Aucejo, E.M., & Romano,T. F. (2016). Assessing the effect of school days and absences on test score performance. *Economics of Education Review*, *55,* 70-87. https://doi.org/10.1016/j.econedurev.2016.08.007

Borland, M. V., & Howsen, R. M. (1998). Effect of student attendance on performance: Comment on Lamdin. *Journal of Educational Research*, *91*, 195–197. https://doi.org/10.1080/00220679809597542

Bos, K. Tj., Ruijters, A.M., & Visscher, A.J. (1992). Absenteeism in secondary education. *British Educational Research Journal*, *18*(4), 381-395. https://doi.org/10.1080/0141192920180405

Chang, H. N., & Romero, M. (2008). Present, engaged, and accounted for: The criticalimportance of addressing chronic absence in the early grades. National Center forChildren in Poverty. Mailman School of Public Health, Columbia University.http://www.aecf.org/resources/present-engaged-and-accounted-for/

Coates, D. (2003). Education production functions using instructional time as an input. *Education Economics*, *11*(3), 273-292. https://doi.org/10.1080/0964529032000148809

Cobb-Clark, D.A., Kassenboehmer,S.C., Le, T.,McVicar, D., & Zhang, R.(2015). Is there an educational penalty for being suspended from school? *Education Economics*, *23*(4), 376-395. https://doi.org/10.1080/09645292.2014.980398

De Witte, K., & Csillag, M. (2014). Does anybody notice? On the impact of improved truancy reporting on school dropout. *Education Economics*, *22*(6), 549-568. https://doi.org/10.1080/09645292.2012.672555

Eide, E. R., & Showalter, M.H. (2001). The effect of grade retention on educational and labormarket outcomes. *Economics of Education Review*, *20,* 563-576. https://doi.org/10.1016/S0272-7757(00)00041-8

Gershonson, S., Jackowitz, A., &Brannegan, A. (2017). Are student absences worth the worry in U.S. Primary Schools? *Education Finance and Policy, 12*(2), 137-165. https://doi.org/10.1162/EDFP_a_00207

Ginsburg, A., Jordan, P., & Chang, H. (2014). Absences Add Up: How school influences student success.http://www.attendanceworks.org/wordpress/wp-content/uploads/2014/09/Absenses-Add-Up_September-3rd-2014.pdf.

Goodman, J. (2014). Flaking Out: Student absences and snow days as disruptions of instructional time. NBER Working Paper 20221.http://www.nber.org/papers/w20221

Gottfried, M. A. (2009). Excused versus unexcused: How student absences in elementary school affect academic achievement. *Educational Evaluation and Policy Analysis, 31*(4), 392–415. https://doi.org/10.3102/0162373709342467

Gottfried, M.A. (2010). Evaluating the relationship between student attendance andachievement in urban elementary and middle schools: An instrumental variables approach. *American Educational Research Journal,* *47*(2), 434-465. https://doi.org/10.3102/0002831209350494

Gottfried, M.A. (2011). Absent peers in elementary years: The negative classroom effects of unexcused absences on standardized testing outcomes. *Teachers College Record, 113*(8), 1597-1632.

Hancock, K.J., Lawrence, D., Shepherd, C.C.J., Mitrou, F., & Zubrick S.R. (2017). Associations between school absence and academic achievement: Do socioeconomics matter? *British Educational Research Journal*, *43*(3), 415-440. https://doi.org/10.1002/berj.3267

Lamdin, D. J. (1996). Evidence of student attendance as an independent variable in education production functions. *Journal of Educational Research*, *89,* 155–162. https://doi.org/10.1080/00220671.1996.9941321

Lefgren, L. (2014). Instrumental variables. In D. Brewer & L. Picus (Eds.), *Encyclopedia of Education Economics & Finance*. Vol. 1, pp. 407-409. Los Angeles, CA:SAGE.

Sundius, J., & Farneth, M. (2008). Missing school: the epidemic of school absence inBaltimore City. Baltimore, MD: Open Society Institute-Baltimore.https://www.opensocietyfoundations.org/reports/missing-school-epidemic-school-absence

U.S. Department of Education. (2016). School climate and discipline: Know the data.http://www2.ed.gov/policy/gen/guid/school-discipline/data.html#featured .

Woodcock, R. W., & Johnson, M.B. (1989). *WJ-R tests of cognitive ability*. Itasca, IL: Riverside Publishing.

Notes

Note 1. See http://psidonline.isr.umich.edu for more information on the PSID.

Note 2. The test scores are the Woodcock-Johnson Revised Tests of Achievement, Form B (Woodcock and Johnson, 1989). We standardize the test scores to have a mean of 0 and a variance of 1, based on the estimating sample.

Note 3. Most U.S. school systems track average daily attendance but do not carefully monitor absences for individual students (Chang and Romero, 2008).

Note 4. The achievement tests were given at the time of the survey. The retrospective questions about absences referred to the period prior to the interview date.

Note 5. One observation has an imputed value of over 180 total absences. Since most districts have only 180 school days, we drop this observation. However, removing this outlier does not affect our results in a qualitative way.

Note 6. This variable measures the student's motivation to perform well in school.

Note 7. Head of Household's years of schooling is topcoded at 17.

Note 8. Family income is a continuous variable that is allowed to take on non-positive values (there are five observations with non-positive values in our sample). We code all non-positive values to $1 so we can compute the log of income.

Note 9. See Lefgren (2014) for a primer on using IV in economics of education analyses.

Note 10. In 2011-12, about 7 percent of public school students in kindergarten through 12th grade were suspended or expelled (U.S. Department of Education, 2016). Cobb-Clark et al. (2015) find that suspension is strongly associated with education outcomes, although the relationship is unlikely to be causal.

Note 11. Our assumption of the randomness of injuries is assumed to hold conditional on all explanatory variables, including family background characteristics such as family income and parents' education, which are included in the regressions.

Note 12. De Witte and Csillag (2014) study the impact of truancy on school dropout. Using fixed effects and quasi-experimental methods they find that truancy is positively related to early school leaving, and that improved truancy reporting reduces school dropout rates.

Note 13. We also estimated the IV regressions including controls for risky behaviors (in the last six months was stopped and questioned by the police, or was arrested by the police) and health status of the child (primary caregiver report, and a hospitalization variable) because these factors could be correlated with both test scores and the likelihood of being absent. The results with these controls are not markedly different than the reported results, and hence are not reported here.

Appendix 1

This appendix provides the wording from the CDS questionnaire for the variables used to construct the total absence variable, and the instruments.

Q21B7A "SCH DAYS MISSED - ILLNESS 02"

B7a. How many days in the past 12 months did CHILD miss more than half of the day from (school/ child care center/ preschool/ Head start) because of illness?

Q21B7B "SCH DAYS MISSED - INJURY 02"

B7b. How many days in the past 12 months did CHILD miss more than half of the day from (school/ child care center/ preschool/ Head start) because of injury?

Q23L11G "SKIPPED SCHOOL 02"

L11g. In the last 6 months, about how many times have you skipped a day of school without permission?

Appendix 2

<u>Calculation of weights for "SKIPPED SCHOOL 02"</u>

If month of interview was Mar. through June, skipSch = 9/6 * "SKIPPED SCHOOL 02"

If month of interview was Feb. or July, skipSch = 9/5 * "SKIPPED SCHOOL 02"

If month of interview was Jan. or Aug., skipSch = 9/4 * "SKIPPED SCHOOL 02"

If month of interview was Sep. through Dec., skipSch = 9/3 * "SKIPPED SCHOOL 02"

This weighting scheme assumes that school ends in June and begins in September. So if a child is asked in September how many days of school she skipped in the last 6 months, her response would only cover 3 months of the school year. If, however, she were asked the same question in June, her response would cover a full six months as the school year. The weights reflect this difference in when the interview was conducted. That is, we divide the response by the number of months out of the last six which were during the school year. Then we multiply that number by 9 to get the number of times the student skipped school over an entire school year.