

Kamhöfer, Daniel A.; Cattan, Sarah; Karlsson, Martin; Nilsson, Therese

**Conference Paper**

## The Effects of Sickness Absence in School on Educational Achievements, Mortality and Income

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Modelling Causal Effects, No. F08-V3

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Kamhöfer, Daniel A.; Cattan, Sarah; Karlsson, Martin; Nilsson, Therese (2015) : The Effects of Sickness Absence in School on Educational Achievements, Mortality and Income, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Modelling Causal Effects, No. F08-V3, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft

This Version is available at:

<https://hdl.handle.net/10419/113180>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# The Effects of Sickness Absence in School on Educational Achievements, Mortality and Income\*

Sarah Cattan

Institute for Fiscal Studies,  
London

Daniel A. Kamhöfer

University of Duisburg-Essen,  
and CINCH

Martin Karlsson

University of Duisburg-Essen  
and CINCH and University of Oslo

Therese Nilsson

Lund University and Research  
Institute of Industrial Economics

February, 2015

PRELIMINARY — do not cite without permission of the authors

## Abstract

This study investigates the effect of missed instruction time in school on short-term educational performance as well as long-term retirement income and mortality. Using self-gathered Swedish register data, we are able to distinguish total days of absence within a school year and missed instructional time due to sickness. Using various fixed effects strategies and an instrumental variables approach we find that individual absence leads on average to a decrease of 4.4% of standard deviation in educational performance per school year. Our results suggest that the negative effect of sickness is mainly driven by missed instructional time. The sign of the long-run consequences of absence are in line with economic theory but the effect size is rather small.

**Keywords:** Absence in school, educational performance, register data

**JEL Classification:** C23, I14, I21

---

All correspondence to Daniel A. Kamhöfer, University of Duisburg-Essen and CINCH-Health Economics Research Center, Essen; Mail: Schützenbahn 70, 45127 Essen, Germany; Phone: +49 (0)201 183-6831; Email: daniel.kamhoefer@uni-due.de.

\*For valuable comments we are grateful to Arnaud Chevalier, Martin Dribe and the participants of the CINCH Research Seminar. We particularly indebted to Nina Schwarz for assembling the different data sources into a customised format. Daniel Kamhöfer thanks the Institute for Fiscal Studies in London for hospitality.

# 1 Introduction

While most social scientists confirm that “healthier children are better learners”<sup>1</sup>, evidence on the consequences of sickness and sickness absence during school is still rare. Using representative self-gathered Swedish register data from the 1930s and 1940s this study analyzes the effect of missed instructional time due to sickness in elementary school on short-term academic performance and long-term income and morality. Doing so this study picks up two streams of ongoing research: (1) the literature analyzing the long-lasting consequences of early health conditions and (2) the literature on the returns to school length extension and instructional time.

Studies investigating the causal effect of education on health usually claim that a causation, running from health to education, challenges the identification (see e.g., [Clark and Royer, 2013](#) and [Lleras-Muney, 2005, 2006](#)). However, this conjecture is itself rarely analyzed so far ([Eide and Showalter, 2011](#)). Among the few studies which analyze the consequences of health, most look at early childhood interventions and their effects on later-life outcomes as educational attendance and mortality ([Currie, 2009](#)). In-utero health shocks persist for instance throughout the whole life (see e.g., [Andreella et al., 2014](#) and [Bruckner et al., 2014](#)).

This study aims at shedding light on the health-education nexus by investigating the effects of sickness absence in elementary school (as health measure) on test scores in the same school year and on retirement income and long-livingly. By analyzing the effects of health in grade 1 and 4 (children aged 7 and 11) we provide evidence on the health-education nexus for a new age group.

For a long time, the literature on the returns to instructional time focused on the effects of school year length. The extension of instructional time per school year – leaving the total number of years unaffected – comes frequently up in educational policy as a way to boost the abilities of students. Extending instructional time works either through more hours of schooling per days or through an increase in the number of weeks with instruction in the school year. Both strategies go along with massive additional expenditures. Evidence on the benefits of more instruction is ambiguous, however. [Lavy \(2012\)](#) uses variation in the instructional time of various subjects between OECD countries and finds substantial returns to additional time allocated to a subject. [Fitzpatrick et al. \(2011\)](#) and [Carlsson et al.](#)

---

<sup>1</sup>This statement goes back to Janet Currie at a seminar at *Institut zur Zukunft der Arbeit*, April 7, 2006, and was also cited by [Heckman \(2007\)](#).

(2012) employ variations in test dates in the USA and Sweden, respectively, and find that more instructional time in advance to a test increases student performance. [Pischke \(2007\)](#) analyses the introduction of so-called short school years in Germany. Students affected by the short school years have up to two-thirds less instructional time and exhibit a higher probability of grade repetition in primary school and are less likely to visit a higher secondary school track. On the other hand, a long-run earning effect is not detectable. To the best of our knowledge, so far, only two studies – [Goodman \(2014\)](#) and [Aucejo and Romano \(2014\)](#) – investigating the causal effect of individual absence on academic performance. Increasing school attendance picks up less resources than extending the school year and might be more effective when worse performing students are more likely to miss instruction.

Establishing a causal effect of instruction time on performance is a challenging task. Beside the legal and probably parental obligation to go to school, students may decide to cut school – e.g., missing school without excuse or pretending to be sick in order to avoid attending school. Moreover, such behavior is likely to be driven by performance in school and by attitudes which correlate with performance but are unobservable to the researcher, e.g., motivation.

The first study which focuses on the causal identification of absence in school is [Goodman \(2014\)](#). Using student-level data from Massachusetts he is able to distinguish individual days of absence and days of absence that apply to all students because of school closures. In order to overcome the endogeneity problem [Goodman \(2014\)](#) applies to two strategies: school and individual fixed effects (FE) estimation as well as an instrumental variables (IV) approach. FE estimation removes the effect of (unobservable) time-invariant factors. The IV strategy handles time-variant unobservables and reverse causation as well. However, this comes to the costs of a conclusion that might be narrower and only applies to some persons. The instrument for days of missed instructional time is days of snowfall within a school year. Employing FE estimation [Goodman \(2014\)](#) finds that an additional day of individual absence decreases math and language skills. When children miss instructional time because the school is closed, he finds no effect on performance. IV estimates a higher effect of individual absence but only provides the local average treatment effect (LATE; see [Imbens and Angrist, 1994](#)) and may not apply to children. [Goodman \(2014\)](#) explains his results with the teachers' behavior. They are maybe more able to cope with collective absence than with individual absence. Missed collective instructional time might be post-

poned. In the case of individual absence, teachers either take time to help the student who missed class to catch up, meanwhile allocating less time to the other students, or they go on with the instruction without paying more attention to the student who missed content.

The second study that goes beyond year-region variation in school length and looks at individual changes is by [Aucejo and Romano \(2014\)](#). Using data from North Carolina they are able to distinguish between excused and unexcused absence. Their data allow them to compare individual absence with collective changes in the school year length that took place at the time under review. They, too, use student FE and IV estimation. As instrument they use variations in the special density of infective diseases. Using triple-FE of student, school and teacher they provide evidence that an additional day of absence leads to a decrease in math and reading test scores. Like in [Goodman \(2014\)](#), the IV coefficient of [Aucejo and Romano \(2014\)](#) exceeds the FE result – although, it states another LATE. The positive effect of an additional day of instruction due to a longer school year accounts for only one-tenth of the negative effect of an individually missed day. Investigating whether nonlinearity of days of schooling plays a role, [Aucejo and Romano \(2014\)](#) find that the effect per day is nearly the same for the first 30 days.

Given the previous findings, the study at hand extends the literature on individual absence by three aspects: (A) We estimating the effect of individual absence using a new and unique data set from Sweden in the 1930s; thereby, analyzing individual absence in a setup never investigated so far. Doing so, this paper contributes to the external validity of the effect of missed instructional time. (B) Unlike the two previous studies, we are able to distinguish overall days of absence and absence due to sickness. Sickness is not only one of the main drivers of absence but also an important factor itself. So far, there is barely any evidence on the consequences of health shocks during elementary school age on short- and long-term consequences. (C) Using historical data allows us to investigate long-term consequences of absence (and sickness) in childhood. Considering outcomes throughout the whole life-cycle allows a more complete picture of the consequences of instructional time and, therefore, enables a more informed policy decision. For the same reason, it is crucial to distinguish the causal effect of instructional time from a plain correlation. Reliable policy can only be drawn based on causal inference. In order to identify causal effects we follow a similar approach to [Goodman \(2014\)](#) and [Aucejo and Romano \(2014\)](#). Our data source enables us to employ school, class and individual FE. Using detailed weather infor-

mation we also apply an IV strategy with the number of benign month in a school year as instrument of sickness absence.

Particularly, this study addresses four research questions: (1) What is the effect of days of absence in primary school on test performance in the same grade? (2) Has sickness a direct effect beside missed instructional time? (3) Is there a long-lasting effect of missed instruction time and sickness on income later in life and on mortality? (4) Is the effect heterogeneous towards the social status and, if so, which groups are stronger affected?

Our results for the effect of sickness absence on performance are in line with the previous literature: throughout different FE specifications we find a per-day effect of 0.4-0.8% of a standard deviation. Our IV result exceeds the FE coefficients, too. Moreover, we find that sickness in elementary school is negatively associated with income later in live. On the other hand, sickness absence has small but positive effect of the probability of being dead before aged 70.

This paper proceeds as follows: Section 2 introduces our data source. Section 3 presents the empirical strategy while Section 4 shows the results. Section 5 investigates the heterogeneity of the effects. Section 6 concludes.

## 2 Data

### 2.1 Social Background

We use self-gathered Swedish register data from different data sources. Our main data source includes all individuals born in 133 church parishes in all regions of Sweden between 1930 and 1934. The sample was originally drawn to evaluate a maternal and infant health program issued by the Swedish government at the end of the 1920s and is, therefore, representative for the Swedish population in the time under review. For more information on the sampling process see [Bhalotra et al. \(2014\)](#) who analyze the health program. The birth register includes the child's name and the exact parish and date of birth. These information are used to match the birth information the other register data.

Beside the name, place and date of birth, the birth register also includes information on socioeconomic indicators. The gender and whether the child was born in a hospital or at home is included. Moreover, we know if it was a twin birth, i.e.,

if the child has a twin. If the child was born in wedlock, we have information on the mother's and father's days of birth and their occupation at the time of birth. Since women barely work in the time under review, only a binary indicator if the mother was employed is used. For the father's employment, we order the occupations according to the five-digit Historical International Classification of Occupations (HISCO). In order to make sure that all groups of occupations have a sufficient number of observations, we only use the first digit of the HISCO code. It ranges from 0 (professional or technical worker) to 9 (unskilled laborer). To compress this information we assign families to one out of three socioeconomic groups according to the HISCO code. The lowered socioeconomic status (SES) is a father occupied as production worker (HISCO 7 to 9). Worked the father in agriculture (HISCO 6), the SES is classified as middle. Service workers (HISCO 0 to 5) have a high SES. For more information see [van Leeuwen et al. \(2002\)](#). Our data also consist information whether the parents belong to the minority of Sami and Meänkieli. Sami and Meänkieli live in Northern Sweden, have their own languages and are Finnish descending. Table 1 summarizes all variables.

## 2.2 Educational Performance

We combine the birth register data with information on educational achievements taken from so-called examination records ("Dagbok med examenskata"). For 11,000 children born between 1930 and 1934 in the 133 parishes we also have information on schooling achievements. We only cannot follow children who died or moved before entering school. Since those children who died before enrollment in elementary school at the age of 7 had a weaker health and are likely to have had a lower social and genetic endowment, our sample is only representative for persons who survived the age of seven. The parental decision to move might be correlated with endowment. However, moving was less like in the time under review than it is today. The effect on our results should be low. Due to our FE strategy, we need schooling information for at least two grades. We use only grades 1 and 4. For this age group we assume absence to be most important because of dynamic complementarities which may affect later years, see [Heckman \(2006, 2007\)](#). For only 3,000 children we could find examination records in the regional archives. We only used information on those children. Since we consider the availability of examination records after a period of 70 years as uncorrelated with health and educational performance, this should not threaten the represen-

Table 1: Summary statistics

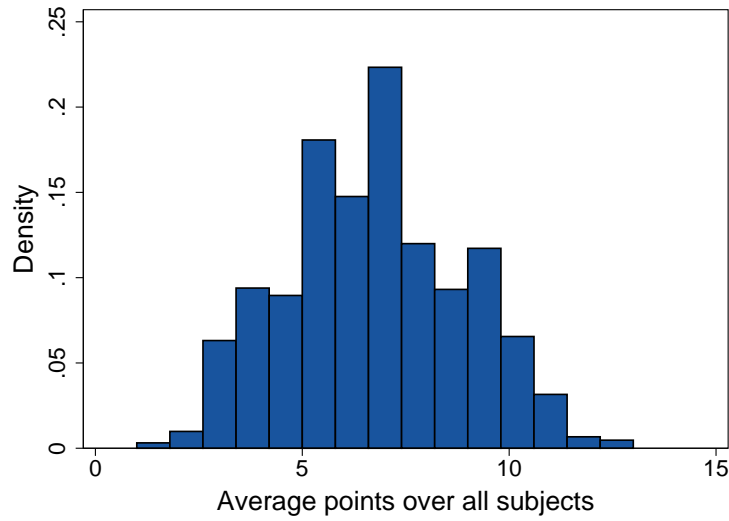
Variable	Mean grade 1	Mean grade 4
Absence (in days)		
Total	11.28	11.26
Sickness	9.64	8.38
Other reason	1.65	2.91
Points		
Average	4.60	6.66
Math	4.84	6.65
Reading & speaking	5.00	7.02
Writing	4.52	6.45
Religious education	4.40	6.84
Physical education	4.18	6.29
Long-term outcomes		
Retirement income (in SEK)	166,839	
Dead at age 70 (in %)	0.12	
Controls (in %)		
Female	0.50	
SES: low	0.45	
SES: middle	0.42	
SES: high	0.13	
Sami and Meänkieli minority	0.10	
Born in wedlock	0.94	
Twin	0.03	
Born in hospital	0.21	
Observations	2,997	2,997

tative of our sample. Since all persons were born between 1030 and 1934 and were usually aged 11 in grade 4, the data cover the school years 1936/37 to 1945/46.

The examination records were issued on classroom level and filled out by the teacher responsible for the class. Figure A1 in the Appendix shows an example of an examination record. Every student in the classroom<sup>2</sup> occupies one line in the record and the columns give the specific dates of the school year. The right columns of the examination records include summaries on performance and absence. We digitized the grades math, read & speak, writing, sports and Christianity education. Since we consider the first three subjects as more important than

<sup>2</sup>The students in the classroom are not necessarily in the same class, i.e., in the same grade. When the number of student in a school was too low to build a class, students of different grades received instruction time in the same classroom. This was especially the case in small school in rural areas.

Figure 1: Distribution of average points



the others, we focus them and only use the others in the deepened analysis of the mechanisms.

Originally, the grades were assigned according to a six-point scale ranging from A (“passed with great distinction”) to C (“not passed”). Each grade had either two or three subcategories. In order to use the information, we linearly transform each grade into a point scale ranging from 1 (worst performance) to 15 (best performance); see Figure A2 in the Appendix. Students needed at least 3 points to pass a subject. We use the average points of math, reading & speaking and writing to measure educational performance. Differentiated results are given in the Appendix. Figure 1 depicts the distribution of the average points (see Figure A4 for the distribution by subject). Most students received the mark BA (7 points). Teachers were encouraged to assign BA for intermediate students. In the time under review, the Swedish schooling system was highly selective and only extraordinary students should receive top grades. The best marks A (“passed with great distinction”) and a (“passed with distinction”) were rarely awarded. In fact only three of over 3,000 students in our sample passed in all three subjects with distinction (average points > 12) and no one passed all subjects with great distinction in either grade 1 or 4.

Unlike Goodman (2014) and Aucejo and Romano (2014), we use the grades assigned by the teacher and not standardized test scores. This has the advantage that the result is may not be driven by from of the day the test was conducted. The subjective judgment should never the less be comparable. First, as discussed

above, teachers were encouraged to assign 7 points for mediocre students. Second, because we condition on the class teacher in some models.

## 2.3 Absence

The examination records also include information on days of absence. We distinguish between total days of absence and days of sickness absence. Besides sickness absence, total days of absence include absence due to natural reasons that does not allow someone to reach to school (snow as used by [Goodman, 2014](#) would fall into this category), inappropriate clothes and shoes, other excused reasons and missing school unexcused. Because of the amount of (manpower to digitize the) data and because other reasons but sickness are often missing, we cannot disentangle other reasons for absence. Figure 2 shows the distribution of average days of absence in grades 1 and 4. About 5% of all students have no day of absence in at least one grade. For the sake of simplicity we pool 40 and more days of absence in the figure; 3.7% of (student-grade) observation are in this category. The mean number of days of absence is 11.3 days and the median is 9 days.<sup>3</sup> Figure A3 shows the distribution of other days of absence by reason and grade. Sickness absence is the main driver of total days of absence with a mean value of 9 days. 37.3% of all students miss class either because of sickness or not at all.

## 2.4 Income and Mortality

In order to assess the long-term effects of absence we use income and mortality as additional outcome variables. We use income data from the tax register from 2002 onwards. Earlier data are not available; however, the (retirement) income in 2002 seems to be a suitable proxy for prime age income. We only consider retirement income and neglect other income categories like capital income. Figure 3 displays the distribution of retirement income in 2002 in Swedish krona (SEK).

The average retirement income in 2002 is 166' SEK (17,700 EUR, 19,800 USD)<sup>4</sup> and the median 154' SEK. The SD is 130'. From here on we use standardized

---

<sup>3</sup>[Goodman \(2014\)](#) and [Aucejo and Romano \(2014\)](#) use rather recent data for students in the US and the distribution of days of absence is comparable. However, their mean values are somewhat lower. See Figure 2 in [Goodman \(2014\)](#) (mean 6.5 days for non-poor student and 9.5 days for poor) and Figure 1 in [Aucejo and Romano \(2014\)](#) (mean 6.1 days).

<sup>4</sup>Using current exchange rates.

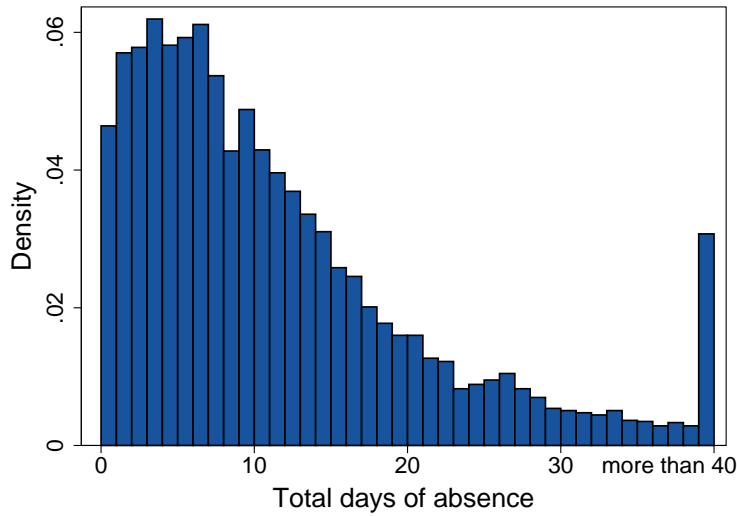


Figure 2: Distribution of total days of absence

income. The income information is only available for 2,400 persons because we lose persons migrated from Sweden or died before 2002.<sup>5</sup>

The second long-term outcome is an indicator whether the person died before reaching the age of 70. 11.6% of the sample died aged 69 or younger (indicator is 1). The exact date of death is taken from the Swedish death book.

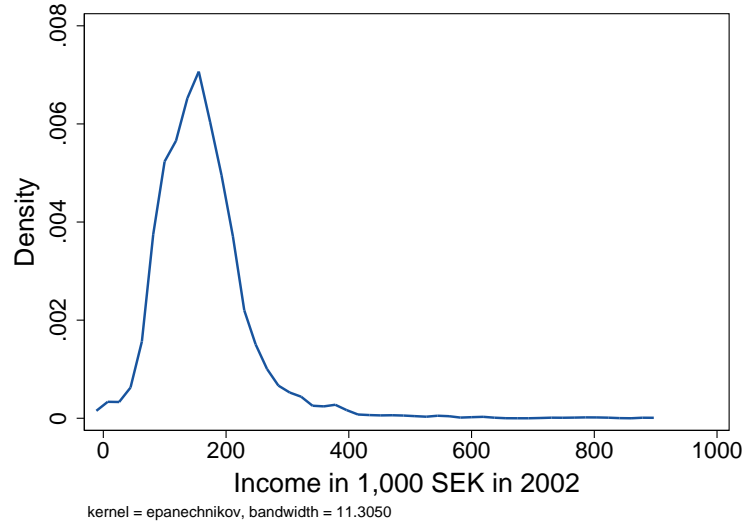
## 2.5 Weather Information

To instrument days of absence we use temperature information for 26 weather stations covering all parishes under review from 1936 to 1946. The average distance between the parishes in which we observe children to the next weather station on average less than 40 km. For 18 weather stations digital files including daily temperatures were provided by *Sveriges Meteorologiska och Hydrologiska Institut* (Swedish Meteorological and Hydrological Institute). For 8 weather stations digital files were not available and we instead use information on the five-day average temperature taken from the Swedish Weather Yearbooks, issues 1936 to 1946. Average temperatures over time and across regions exhibit a fair amount of variation, see Figure A5 in the Appendix for selected regions.<sup>6</sup> Our instrument is the number of benign months in a school year in the region of the weather station. A month in a certain school year is classified as being benign when the

<sup>5</sup> Additionally, we drop 3 persons with an income higher than 1 million.

<sup>6</sup> The pattern is the same for all 26 regions; however, for reasons of clarity we limit the figure to selected regions.

Figure 3: Distribution of income in 2002



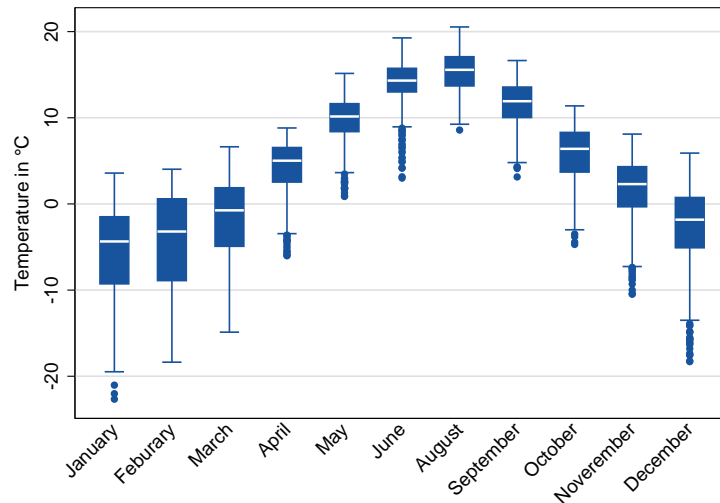
average in this temperature in a particular year is within the upper quintile of all mean temperatures for this month and this region over all years. This measurement of weather shocks was for instance used before by [Bruckner et al. \(2014\)](#). E.g., in the years under review, Malmö in Southern Sweden, has a 80th quantile temperature in August of  $18.1^{\circ}\text{C}$ . In April 1939, the average temperature was  $16.8^{\circ}\text{C}$ . We do not count this month as benign, therefore. In 1940, the average temperature in April was  $19.3^{\circ}\text{C}$ . The benign month variable takes value one. We follow this procedure for all regions and months. The instrument is the number of benign months in a school year in a region.<sup>7</sup> Figure 4 shows the boxplot of the temperature per month pooled for all 26 regions and all 10 school years. The figure also gives the interquartile temperature difference, i.e., the difference of the temperature between the fourth and the first quartile. Figure A6 in the Appendix presents similar boxplots by region.

## 2.6 Descriptives

Figure 5 states the relationship between days of sickness absence and average points over all subjects. The dots mark the mean number of points we observe per day of absence. The figure exhibits a clearly negative relationship: more days of absence are associated with a low performance in school. However, the slope of the relationship seems rather flat. While children without absence have on

<sup>7</sup>In order to make our calculations as precise as possible, we drop temperature information on holidays. The month July is for instance left out because it lies completely in the summer break.

Figure 4: Monthly box plots of temperatures



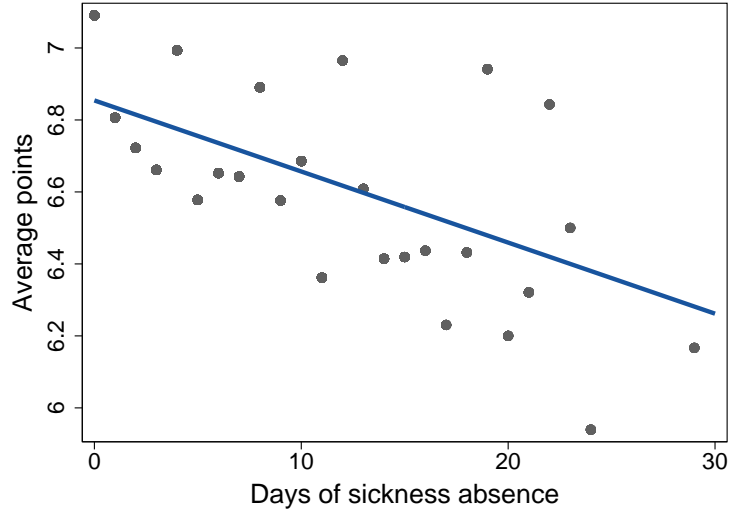
average 6.85 points, children who miss 10 days of absence score on average only 6.7 points. One reason for this might be that the variation of points this in general rather low in the 1930s and 1940s in Sweden. Furthermore, the points give the average of three subjects – math, reading & speaking and writing – changes in one subject might be canceled out. Figure A7 shows the relationship per subject. The effect of absence on math points is the most pronounced.

### 3 Empirical Strategy

#### 3.1 The Endogeneity Problem

Besides the clear graphical evidence, drawing causal conclusions based on the pattern presented in the descriptive statistics is not possible. On the one hand, there might be a causation running from performance in school to sickness absence. I.e., children miss instruction more often because they perform worse. On the other hand, so far not considered – and maybe even unobserved – factors could simultaneously affect absence and educational performance and, therefore, confound the graphical relationship.

Figure 5: Graphical relationship between days of sickness absence and average points



Starting point of our empirical strategy is a conventional ordinary least square (OLS) regression of the form

$$\begin{aligned}
 Points_{igp} = & \beta_0 + \beta_1 DA_{igp} + \beta_2 X_{1igp} + \beta_3 X_{2ip} + \beta_4 GR_{igp} + \beta_5 BY_{ip} + \beta_6 BM_{ip} \\
 & + \beta_7 AGE_{ipg} + \beta_8 BYM_{ip} + \beta_9 BYF_{ip} + \beta_{10} SES_{ip} + \beta_{11} PA_{igp} + \varepsilon_{igp}.
 \end{aligned}
 \tag{1}$$

$Points_{igp}$  is the average number of points person  $i$  in grade  $g$  (which is either 1 or 4) in parish  $p$  reaches in math, reading & speaking and writing. The variable of interest in  $DA$ , it states the number of days of absence  $i$  has in grade  $g$ .  $\beta_1$  gives the effect an additional days of absence has on the value of  $Points$ .  $\beta_0$  denotes the intercept.  $GR$  is a dummy variable that takes the value 1 if  $i$  visits grade 4, and 0 otherwise.  $X_1$  and  $X_2$ .  $X_1$  covers time-variant controls (school year length and grades range in classroom).  $X_2$  captures only time-invariant factors (gender, born in wedlock, twin birth, born in hospital, mother employed at time of birth, child belongs to Meänkieli and Sami minority). Since these factors do not change over time,  $X_2$  has only the subscript  $is$ .  $BY$ ,  $BM$ ,  $AGE$ ,  $BYM$ ,  $BYF$ ,  $SES$  and  $PA$  include dummies for the child's year of birth, the child's month of birth, the age in grade  $g$ , the year of birth of the father, the year of birth of the month, the socioeconomic status of the family (according to the first digit of the HISCO code of the occupation of the father) when the child was born, and the parish of residence. Only age and parish of residence are time-variant. All other vectors

have the subscript do not vary between grades 1 and 4. Since we control for the year of birth and the age in grade  $g$ , we indirectly also control for the year in which  $i$  was in grade  $g$ .

### 3.2 Fixed Effects and Omitted Confounders

Equation (1) may not cover all sources of selection. Within a parish parents may prefer to send their child to one school than another because of a better reputation of the school. Since those parents are more likely to put generally a stronger emphasis of their child's education, they might also take care that that child is less often absent. In order to capture such a selection, we use a school fixed effects (FE) estimation:

$$\begin{aligned} Points_{igs} = & \beta_0 + \beta_1 DA_{igs} + \beta_2 X_{1igs} + \beta_3 X_{2is} + \beta_4 GR_{igs} + \beta_5 BY_{is} + \beta_6 BM_{is} \\ & + \beta_7 AGE_{ips} + \beta_8 BYM_{is} + \beta_9 BYF_{is} + \beta_{10} SES_{is} + \psi_s S_s + \varepsilon_{igs}. \end{aligned} \quad (2)$$

Instead of parish FE we, here, include school FE  $S$ . Therefore, all other factors have the subscript  $s$ . We only compare children within the same school with each other. Since the schools were located in certain parishes and we do not observe schools that spread over more than one parish, Equation (2) no longer controls directly for parishes.

The structure of the data allows us to go beyond school FE and use class FE. Here we only compare children within the same class. We assign children to the same class if they are in the same school, visit the same grade in the same year, and have the same teacher. The resulting equation includes class FE  $C_c$  instead of school FE:

$$\begin{aligned} Points_{ic} = & \beta_0 + \beta_1 DA_{ic} + \beta_2 X_{1ic} + \beta_3 BY_{ic} + \beta_4 BM_{ic} + \beta_5 AGE_{ic} \\ & + \beta_6 BYM_{ic} + \beta_7 BYF_{ic} + \beta_8 SES_{ic} + \psi_c C_c + \varepsilon_{ic}. \end{aligned} \quad (3)$$

Since classmates have the grade in common, this equation does no longer control for the grade dummy  $GR$ . For the same reason the subscript  $g$  is neglected. The vector  $X_2$  of the time-variant variables school year length and grades range within the classroom is also dropped from the equation because it only varies on class level. This specification has the appealing feature over school FE (and even

over individual FE presented below) that time-invariant omitted variables were cancelled out as well.

Another threat to the identification are individual-specific unobservable factors. Neither school FE nor class FE would capture those factors. Examples are inherited genes which drive intelligence and attributes or parental behavior beyond the school choice. When a higher innate intelligence goes also along with a higher motivation or better innate health, the estimate of  $\beta_1$  would be biased in all equations above. The same is true for parental behavior. If parents who put a stronger emphasis on good grades also take care of a more frequent attendance, the estimators presented so far would fail. We overcome this problem using individual fixed effects:

$$Points_{ig} = \beta_0 + \beta_1 DA_{ig} + \beta_2 X_{2g} + \beta_3 AGE_{ig} + \psi_i I_i + \varepsilon_{ig}. \quad (4)$$

The vector  $I$  includes an indicator for every person.  $\psi$  gives the individual effects. All other FE for which we control above are cancelled out in the individual FE model because they do not vary between grades 1 and 4. The grade indicator  $GR$  also disappears from the equation. Although  $GR$  varies for given individual (it takes the values 1 in grade 4 and 0 in grade 1), the difference is the same for every person. This is also the reason why we are not able to include teacher FE in the individual FE regression. In the years under review it was common in Swedish school system that the teacher changed between grades 1 and 4. Including age is still possible because some students need to repeat a class. Therefore, the interpretation of this variable changed in the individual FE model. The identifying variation used in Equation (4) is the within-student difference in the number of missed days and the average points scored. In other words, we compare each student in grade 4 with herself in grade 1. As long as the unobservables do not vary between the grades, those factors were cancelled out here and do no longer cause a bias of the estimation of  $\beta_1$ .

Because individual FE estimation does not allow to control for time-invariant confounders but most of the social background variables are either time-invariant by nature (e.g., sex) or measured only once (e.g., SES of the family), we additionally use a cross sectional model which enables us to control of lagged factors. Starting point is the class FE specification of Equation (3). We only use information on

grade 4 and additionally include days of absence and average points as observed in grade 4. Formally, it follows:

$$\begin{aligned}
Points_{ic,g=4} = & \beta_0 + \beta_1 DA_{ic,g=4} + \beta_2 DSA_{ic,g=1} + \beta_3 Points_{ic,g=1} \\
& + \beta_4 X_{1ic,g=4} + \beta_5 X_{2ic} + \beta_6 BY_{ic} + \beta_7 BM_{ic} + \beta_8 AGE_{ic} \\
& + \beta_9 BYM_{ic} + \beta_{10} BYF_{ic} + \beta_{11} SES_{ic} + \psi_c C_c + \varepsilon_{ic}.
\end{aligned} \tag{5}$$

Time-variant variables are conditioned on the fourth grade. The equation is analogous to the class FE Equation (3) but also includes absence ( $DA_{ic,g=1}$ ) and performance ( $Points_{ic,g=1}$ ) in grade 1 as control variables. The lagged model identifies a causal effect of sickness absence when the time-invariant unobservable confounders are reflected in absence and performance in grade 1. Since the individual FE model only includes to other variables and because we only observe two points in time, the models of Equations (4) and (5) should not differ much.

### 3.3 Reverse Causation

Besides omitted factors another concern regarding the identification of a causal effect is a reverse causation running from performance in school to days of absence. I.e. bad students miss class more often because of their performance. They are maybe less motivated due to the bad grades in a given school year. If such a behavior is not consistent in grades 1 and 4, individual FE will not take the effect into account. When bad grades and cutting class are introduced by the teacher, the class FE may capture the effect and provides an unbiased estimation of absence. If the correlation of performance and absence can neither be explained by the individual's attributes nor by the teacher, the above methods will be produced biased estimations.

However, we would argue that children aged 7 or 11 are probably too young to cut class systematically because they are unsatisfied with the reward for their efforts in school. This is probably even truer in the 1930s and 1940s, where the possibility of attending school was not taken for granted. Even if some children cut school systematically and this is not captured by individual or class FE, this group of children should numerically rather be low. Thus, such behavior should not affect our results. On the other hand, there might be a positive reverse causation because better students are also healthier. The theoretical foundation for this is given by [Grossman \(1972\)](#). Better educated individuals possess more knowl-

edge which may make them better producers of their own health – either by combining inputs more efficiently or by achieving greater health improvements out of a given set of inputs.

We cannot completely rule out this relationship but we think of it as rather unlikely. While a positive correlation between education and health is well-established, identifying causal inference is a challenging task. [Fischer et al. \(2013\)](#) use the same data source as we do, i.e., they observe the same persons, and find that the introduction of an additional seventh grade reduces the probability of dying at younger ages. The shorter the time horizon is they analyze the bigger the effect of the reform. This indicates that risk behavior may drive the results. Using information on more than 1.4 million observations [Meghir et al. \(2012\)](#) analyze the introduction of the eighth grade in Sweden and find no economic meaningful effect on mortality. In general, even if there is an effect, it seems to be rather small; see [Lleras-Muney \(2005, 2006\)](#) for evidence on the US and [Clark and Royer \(2013\)](#) for UK. Looking at biomarkers [Jürges et al. \(2013\)](#) find no effect for the UK. For Germany, [Kemptner et al. \(2011\)](#) find a positive effect of compulsory schooling on health behavior but no direct effect on health. Even the health returns to college education seem rather small; see [Kamhöfer et al. \(2014\)](#) for evidence on Germany.

Given all these points, we think it is fair to argue that reverse causation is likely to play only a minor role in our application. When even the expansion of instructional time by a whole school year has no or only a very small effect on health, single days should barely affect the results. Moreover, the effect seems mainly to be driven by health behavior. Since we analyze first and fourth grades, their health behavior should be determined by their parents and the education of their parents. This factor is, however, reflected in the family's SES – for which we control in our analysis.

As an additional check whether reverse causation matters here, we instrument days of sickness absence with the number of cold months in the school year. In the first step, we measure the effect of an additional benign month on the days of absence in this grade. Formally, the first stage is

$$\begin{aligned}
 DA_{igs} = & b_0 + b_1 HM_{igs} + b_2 X_{1igs} + b_3 X_{2is} + b_4 GR_{igs} + b_5 BY_{is} + b_6 BM_{is} \\
 & + b_7 AGE_{ips} + b_8 BYM_{is} + b_9 BYF_{is} + b_{10} SES_{is} + p_s S_s + \zeta_{igs},
 \end{aligned}
 \tag{6}$$

where  $HM$  is the number of (relatively) hot month defined as in Section 2. All other terms are as in the school FE model of Equation (2).<sup>8</sup> Based on the estimated coefficients of the first stage we assign each student to a predicted number of days of sickness absence in grade 1 and 4. In the second stage we then use predicted days of absence instead of the actual (endogenous) number of days.

## 4 Estimation Results

### 4.1 The Effect on Educational Performance

Table 2 presents the baseline estimation results of the effect of days of absence on average points over math, reading & speaking and writing. Each column presents a separate estimation where the column number corresponds to the model discussed in Section 3. The outcome is measured in units of a standard deviation (SD) in order to simplify interpretation.<sup>9</sup>

Column 1 of Table 2 states the result of conventional pooled OLS estimation. One additional day of sickness is associated with a decrease in performance by 0.39% of a SD. Assuming a linear effect (we investigate this later), the mean number missed days (11) goes along with a performance decrease of 4.4% of a SD. This result does not take the panel structure of the data into account – although, the standard errors (SE) are clustered at the individual level. Column (2) presents the school FE model. When comparing only children within the same school, an additional day of absence leads to 0.44% of a SD few points. The same size of OLS and school FE results indicates that parents either do not select their children into specific schools or that the selection, if any, is independent from their emphasis of sickness absence. Column (3) compares only children within the same class. The effect of sickness absence is  $-0.4\%$  of a SD. The effect is about the same as before. This hints that children do not simultaneously adjust performance and absence according to disruptions on class level, e.g., teaching style. The individual FE result of column (4) only uses the within student variation between grades 1 and 4. If a student misses one day more compared to the number of days she is on average absent, her performance drops by 0.74% of a SD. Compared to the models

---

<sup>8</sup>We prefer the school FE model because unlike class and individual FE it allows us to compare grade 1 and 4. Given the results of other models, the choice of the specification does not seem to be crucial.

<sup>9</sup>If the grade of one of the three subjects is missing, we neglect it when building the average. When grades of two or all three subjects are missing, we treat the average points as missing.

(1) to (3) the effect size nearly doubles. This tells us two things: First, unobservable time-invariant factors seem to matter; otherwise the coefficient should not differ. Second, it looks like those unobservables lead to an overestimation<sup>10</sup> of the true effect by OLS, school and class FE. In order to be able to control for other confounders, column (5) presents the results of the lagged specification. Here, days of absence were not compared to the mean value over grade 1 and 4 but to grade 1.<sup>11</sup> The coefficients of specifications (4) and (5) barely differ. This suggests that days of sickness absence and average points in grade 1 are valid proxies for time-invariant unobservables.

The difference between the results of models (1) to (3), which do not overcome time-invariant unobservables, and models (4) and (5) that cancel the effect of time-invariant factors out, is remarkable. That indicates the naive OLS estimations are indeed biased for the reasons discussed in Section 3. Moreover, the bias goes in the same direction we would expect when factors like innate abilities and motivation play a role: the conventional strategies overestimate the true effect.<sup>12</sup> Individual FE might not be the true effect, but given the assumptions presented above, it is closer to the true effect than the other methods.

Tables A1 to A3 in the Appendix give the results by subject. The effect of days of absence is nearly always between 0.3% and 0.4% of a SD in models (1) to (3) and about 0.7% of a SD in models (4) and (5). Only one of the coefficients fails the 10% significance level, class FE for writing points. Because writing is only observed of half of the children in grade 1, this might also be due to the smaller sample size. The jump in the sickness absence coefficient between the class and the individual FE model is persistent for each subject.

<sup>10</sup>An overestimation in relative terms. In absolute terms the other methods overestimate the effect (because of the negative sign).

<sup>11</sup>Since the mean values in days of absence are nearly the same in both grades (compare Table 1), the reference value should not matter.

<sup>12</sup>It is easy to see why we would expect an overestimation when innate abilities or motivation is omitted. Generally, the true effect is  $\beta_1^{\text{true}} = \hat{\beta}_1^{\text{OLS}} - \text{bias}$  and we estimate  $\hat{\beta}_1^{\text{OLS}} = \beta_1^{\text{true}} + \text{bias}$ , where  $\hat{\beta}_1^{\text{OLS}}$  is the OLS estimated effect of sickness absence on performance (although we speak of OLS, the same applies to all models in which time-invariant unobservables matter – therefore, school and class FE, too). Using the common textbook formula for omitted variables, it follows (Wooldridge, 2010, p.67):

$$\hat{\beta}_1^{\text{OLS}} = \beta_1^{\text{true}} + \frac{\beta_{OV} \text{cov}(DSA, OV)}{\text{var}(DSA)}.$$

The omitted factor is denoted with  $OV$  and its effect on *Points* is  $\beta_{OV}$ . In the case of abilities or motivation, the omitted factor has a positive effect on performance ( $\beta_{OV} > 0$ ). On the same time, it is negatively correlated with absence ( $\text{cov}(DSA, OV) < 0$ ). Thus,  $\beta_1^{\text{true}} = \hat{\beta}_1^{\text{OLS}} - (\text{bias} < 0)$  if follows that  $\hat{\beta}_1^{\text{OLS}} > \beta_1^{\text{true}}$ . Therefore, the models that do not take time-invariant omitted factors into account would overestimate the true effect.

Table 2: Baseline estimation results of absence in school on educational performance

	Conventional OLS (1)	School FE (2)	Class FE (3)	Individual FE (4)	Lagged variables (5)
<b>Effect of total days of absence on average points</b>					
Coefficient	−0.0039***	−0.0044***	−0.0040***	−0.0074***	−0.0066***
S.E.	(0.0011)	(0.0011)	(0.0012)	(0.0013)	(0.0016)
<b>Effects of selected control variables</b>					
Female	0.2694*** (0.0254)	0.2760*** (0.0334)	0.2745*** (0.0240)	0.2881***	0.2695*** (0.0329)
SES: middle	−0.0094 0.1982***	0.0070 0.1996***	−0.0065 0.2058***		0.0555 0.2814***
SES: high	0.1930*** (0.0428)	0.1961*** (0.0578)	0.2013*** (0.0400)		0.2732*** (0.0578)
Grade 4	−0.1275** (0.0621)	−0.1371 (0.1101)			
Lagged absence					0.0055*** (0.0016)
Lagged points					0.7738*** (0.0288)
<b>Other conditioning variables</b>					
Time-variant	✓	✓	✓	✓	
Time-invariant	✓	✓			✓
<b>Fixed effects</b>					
Birth year & month	✓	✓			✓
Age	✓		✓	✓	✓
Birth year mother/father	✓	✓			✓
Parish	✓				✓
School		✓			
Class			✓		
Individual				✓	
Lagged absence & points					✓
<b>Sample size</b>					
Observations	5,994	5,994	5,994	5,994	2,997
Persons	2,997	2,997	2,997	2,997	2,997

Notes: Dependent variable: average points over math, writing, reading and speaking in standard deviations. Explanatory variable: number of days of sickness absence. Time-variant conditional variables: range of grades instructed in the same classroom, length of the school year in weeks. Time-invariant conditional variables: female, born in wedlock, twin birth, mother employed at the time of birth, born in hospital, mother or father Finnish descent. Baseline category of father's occupation (occu.): unskilled or low-skilled production worker. Clustered standard errors (S.E.) in parentheses. Clustering was conducted on the level of the fixed effects (FE) in columns (2) to (4) and on individual level in columns (1) and (5). Significance: \* $p \leq 0.1$ , \*\* $p \leq 0.05$ , \*\*\* $p \leq 0.01$ .

Table 2 also gives the coefficients of some of the conditioning variables. Girls score on average one quarter better than boys. Between the baseline category of low SES (father is unskilled production worker) and middle SES (father is agricultural worker) is no difference on performance. When the family has a high SES (father is service worker), children have on average one-five of a SD better grades. Children in grade 1 have on average better grades than children in grade 4. This might be due to the increased level of difficulty in the curriculum. In specification (5), average points in grade 1 have a big effect on performance in grade 4. Surprisingly, the effect of lagged absence is also positive. This could be due to catch-up behavior of students who missed much instructional time. All in all, our results are in line with the two studies that analyzed the relationship between individual absence and educational achievement before. In his individual FE specification Goodman (2014, Table 2) finds effects of  $-0.008$  SD for math and language skills. Using individual FE Aucejo and Romano (2014, Table 3) find an effect of  $-0.0066$  for math and  $-0.0035$  for reading per day of absence. Employing a lagged specification and adding school and teacher FE, they estimate a coefficient of  $-0.0076$  for math and  $-0.0039$  for reading. That our results are similar is particularly noteworthy for two reasons: First, our data cover children born in Sweden in the 1930s while Goodman (2014) and Aucejo and Romano (2014) look at children born in the US after 2000. Second, we only have 6,000 observations and Goodman (2014) and Aucejo and Romano (2014) use 1.7 million and 1 million observations, respectively.

## 4.2 Sickness vs. Missed Instructional Time

This subsection investigates two questions: (1) Does the effect of sickness absence differ from the overall effect? This question allows us to investigate the role of health. (2) Do students learn missed content at home? Firstly, we disentangle total days of absence from sickness absence. Since we can expect both effects to go in the same direction, the distinction leaves the overall effects unaffected; however, both effects lead to different policy implications. If the sickness effect drives the result, parents should be encouraged to let children stay at home even for less severe reasons, e.g., a flu. The child could learn topics she missed using her textbooks. Staying at home would avoid that she infects her class mates and that the sickness becomes more severe. In one the other hand, missed instruction drives the overall result, students should be incentivized and enabled to learn missed context at home.

In order to get a separate effect of sickness absence, Table 3 disentangles the overall effect. Panel A shows the effect of total days of absence. Panels B and C split the number of total days in days of sickness and other days of absence, respectively. The odd numbered lines give, as before, the effect on the average points over math, reading & speaking and writing. All kinds of absence have a negative effect on grades. The effects of total days and days of sickness absence are nearly the same – although, the OLS estimator is higher for sickness absence. This may seem unsurprising since sickness absence drives the total days of absence, see Figure A3. However, given that sickness may determine educational performance indirectly through instructional time and possibly also direct, this result is remarkable. It hints that sickness influences education only through instruction. Panel C shows the effect of missed days for other reasons than sickness. The effect size is approximately the same as for sickness absence. This provides additional evidence that missed instruction drives the negative effect of sickness absence. Since we observe fewer days of absence for other reasons (mean 2.9 days) and less children who miss class for other reasons, the statistical significance is somewhat lower. Moreover, the jump between class FE and individual FE does not appear for other days of absence. We would interpret the results in Table 3 as indicator that given a child is absent, sickness (besides being the reason for absence) has no or only a minor effect on performance.

In order to investigate if children learn at home when they miss instruction, we compare the effect (sickness) absence has on math, reading & speaking as well as writing with the effect on religious and physical education. We follow the idea that students (and their parents) put more emphasis on math, reading & speaking and writing because these subjects are considered more important. On the other hand, religious and physical education are considered less important for further success. In order to draw inferences about learning at home, we need two assumptions: First, without learning, absence has the same effect on both groups of subjects. Second, when children miss class, they spend more time learning for the more important subjects. The even numbered lines in Table 3 give the effects absence on less important subjects. The effect of (sickness) absence seems to be the same for both groups of subjects. Given the assumption that absence affects both subjects in the same way without learning, the findings of Table 3 indicate that students do not learn for the important subjects when they miss instructional time. Given those results (and assumptions), the opportunities for children who missed class to learn at home should be improved. Teachers should

Table 3: Comparison of kinds of absence and groups of subjects

	Conventional OLS (1)	School FE (2)	Class FE (3)	Individual FE (4)	Lagged variables (5)
<b>Panel A: Total days of absence</b>					
<b>(1) Effect of total days of absence on important subjects (equal to Table 2)</b>					
Coefficient	−0.0039***	−0.0044***	−0.0040***	−0.0074***	−0.0066***
S.E.	(0.0011)	(0.0011)	(0.0012)	(0.0013)	(0.0016)
<b>(2) Effect of total days of absence on <i>less</i> important subjects</b>					
Coefficient	−0.0038***	−0.0040***	−0.0041***	−0.0048***	−0.0064***
S.E.	(0.0010)	(0.0011)	(0.0010)	(0.0013)	(0.0017)
<b>Panel B: Days of sickness absence</b>					
<b>(3) Effect of days of sickness absence on important subjects</b>					
Coefficient	−0.0073***	−0.0041***	−0.0035***	−0.0075***	−0.0073***
S.E.	(0.0018)	(0.0012)	(0.0012)	(0.0014)	(0.0018)
<b>(4) Effect of days of sickness absence on <i>less</i> important subjects</b>					
Coefficient	0.0018	−0.0001	−0.0042	0.0044	0.0076**
S.E.	(0.0027)	(0.0028)	(0.0032)	(0.0029)	(0.0038)
<b>Panel C: Other days of absence</b>					
<b>(5) Effect of other days of absence on important subjects</b>					
Coefficient	−0.0041	−0.0062**	−0.0081**	−0.0068**	−0.0042
S.E.	(0.0026)	(0.0029)	(0.0033)	(0.0029)	(0.0034)
<b>(6) Effect of other days of absence on <i>less</i> important subjects</b>					
Coefficient	0.0018	−0.0001	−0.0042	0.0044	0.0076**
S.E.	(0.0027)	(0.0028)	(0.0032)	(0.0029)	(0.0038)

Notes: Each cell gives the coefficient of a separate regression. Specifications and control variables as in Table 2. Clustered standard errors (S.E.) in parentheses. Clustering was conducted on the level of the fixed effects (FE) in columns (2) to (4) and on individual level in columns (1) and (5). Significance: \*  $p \leq 0.1$ , \*\*  $p \leq 0.05$ , \*\*\*  $p \leq 0.01$ .

advice students on what need to be learnt at home. Moreover, textbooks should be provided to enable students to learn outside the classroom.

### 4.3 Long-term Effects

Tables 4 and 5 give the effect of absence in grade 4 of primary school on income and the mortality. The tables show the results for sickness absence in grade 4 because other days of absence (included in total days of absence) seem to be less

consistent also the results do not differ much. The choice of the grade does not matter. Since we observe both outcomes per persons at only one point in time, individual FE and the lagged specification are not applicable here. The models in columns (1) to (3) are as in Table 2.

Table 4: Income results

	Conventional OLS (1)	School FE (2)	Class FE (3)
<b>Effect of days of sickness absence on income (in SD)</b>			
Coefficient	−0.0040	−0.0045	−0.0030
S.E.	(0.0029)	(0.0035)	(0.0033)
<b>Effects of selected control variables</b>			
Female	−0.6720*** (0.0612)	−0.6917*** (0.0741)	−0.6331*** (0.0503)
Father occu.: agricul- ture	−0.0338 (0.0437)	−0.0456 (0.0450)	−0.0722 (0.0500)
Father occu.: service	0.3564*** (0.0928)	0.3546*** (0.0780)	0.3358*** (0.0871)
<b>Sample statistics</b>			
Observations	2,328	2,328	2,328
Persons	2,328	2,328	2,328

*Notes:* Explanatory variable: number of days of sickness absence. Control variables and fixed effects (FE) of year & month of birth, age, year of birth mother/father, oparish, school, and class as in Table 2. Clustered standard errors (S.E.) in parentheses. Baseline category of father's occupation (occu.): unskilled or low-skilled production worker. Clustering was conducted on individual level in the OLS model, on school level in the school FE model, and on class level in the class FE model. Significance: \* $p \leq 0.1$ , \*\* $p \leq 0.05$ , \*\*\* $p \leq 0.01$ .

Table 4 shows the effect of days of sickness absence on a person's income (in units of SD) in 2002. Throughout all models the effect of an additional day of sickness absence reduces the income between 0.3% and 0.45% of a SD. While the effect is economically plausible, it is statistically not different from zero. However, for three reasons we are not surprised that the effect is not sharper: First, the sample size of 2,300 observations is rather small. Second, we look at an intervention of just one additional day of sickness absence. Third, the time span between visting grade 4 (at age 11) and the year 2002 is on average 60 years. Therefore, we will not stress the statistical significance of the point estimators too much; instead we would like to pronounce that the coefficients of all models have the expected

sign and that the effect seems to remain over a long period of time. The effect is rather strong: the gender wage gap has the size of about two-thirds of a school year sickness absence. The difference in income between low-skilled production worker (baseline category) and higher skilled service workers equals the effect of 90 days of sickness absence.

Table 5: Mortality results

	Conventional OLS (1)	School FE (2)	Class FE (3)
<b>Effect of days of sickness absence on the probability of being dead before age 70</b>			
Coefficient	0.0003	0.0002	0.0001
S.E.	(0.0007)	(0.0006)	(0.0007)
<b>Effects of selected control variables</b>			
Female	−0.0504*** (0.0117)	−0.0441*** (0.0099)	−0.0457*** (0.0139)
Father occu.: agricul- ture	−0.0244* (0.0140)	−0.0021 (0.0155)	0.0036 (0.0159)
Father occu.: service	0.0008 (0.0200)	0.0115 (0.0186)	0.0118 (0.0229)
<b>Sample statistics</b>			
Observations	3,085	3,085	3,085
Persons	3,085	3,085	3,085

*Notes:* Explanatory variable: number of days of sickness absence. Control variables and fixed effects (FE) of year & month of birth, age, year of birth mother/father, oparish, school, and class as in Table 2. Clustered standard errors (S.E.) in parentheses. Baseline category of father's occupation (occu.): unskilled or low-skilled production worker. Clustering was conducted on individual level in the OLS model, on school level in the school FE model, and on class level in the class FE model. Significance: \* $p \leq 0.1$ , \*\* $p \leq 0.05$ , \*\*\* $p \leq 0.01$ .

Table 5 depicts the effect of sickness absence on the probability of being passed away the age of 70. The outcome variable takes the value 1 if a person is dead before reaching the 70th birthday; and otherwise 0. An additional day of sickness absence in school at age 11 increases the probability of being dead before age 70 between 0.03 percentage points in the OLS model and 0.01 percentage points in the class FE model. The baseline probability of dying before age 70 is 11.7%. The relative effect of the OLS coefficient is, therefore, 0.26%.<sup>13</sup> Besides the statistical insignificance, the effect size seems plausible. Assuming linearity, 160 days of

<sup>13</sup>Because  $0.0003/0.117 = 0.0026$ .

sickness absence equal the effect of being male instead of female, about half a school year. Again, the results match our expectations: The sign is line with the theory. Given the small intervention of one day of sickness absence, a stronger and highly significant effect would even cast doubts.

#### 4.4 Instrumental Variables Estimation

As discussed in Section 3, we cannot completely rule out that a causation running from education to health bias the results – although, this is highly unlikely, as we have argued. Therefore, we use an IV approach to overcome this weakness. The first line of Table 6 presents the effect of the number of benign months within a school year on the days missed due to sickness. Per additional month in a school year with a temperature within the upper quintile of the distribution the number of missed days decreases on average by 0.42 days. I.e., a benign month leads to half of a day less sickness. Differences in the first stage are only due to the number of observations; the first-stage specification is always Equation (6). The first-stage  $F$  statistic of the instrument at the bottom of Table 6 is always above 25. Since this clearly exceeds the [Staiger and Stock \(1997\)](#) rule-of-thumb value of 10, we consider the number of benign months as relevant for days of sickness absence.

The second line of Table 6 states the structural estimations. An additional (instrumented) day of sickness absence leads to a decrease in educational performance by 3.2% of a SD. The effect is statistically significant at the 10% level. This is about eight times the effect of the OLS estimation. However, the point estimators do not give the effects for the same groups of people. While the OLS estimator gives the average treatment effect over all persons, the IV effect of only true for the group of compliers. I.e., it only the effect depicts of sickness absence on performance for those children who miss instruction because of the weather.<sup>14</sup> [Goodman \(2014\)](#) uses with days of snow another weather condition and finds similar results. [Aucejo and Romano \(2014\)](#) instrument absence with differences in the infection probability across years and regions and also find that the IV estimation clearly exceeds the OLS/FE results. Moreover, unlike the FE estimations, the IV results exhibit a strong variation across subjects. Table A4 in the Appendix presents separate IV estimation for math, reading and speaking and

---

<sup>14</sup>To be precisely, it is the effect for children who miss class because in the year and the region in which they visited school the temperature was not as often in the upper quintile as in the case of their peers in other year and/or regions.

Table 6: IV estimation results by subject

	Average points	Income	Mortality
<b>Effect of days of instrumented sickness absence</b>			
Coefficient	-0.0319*	0.0065	0.0085*
S.E.	(0.0177)	(0.0107)	(0.0045)
<b>First-stage effect of benign months</b>			
Coefficient	-0.4271***	-0.4922***	-0.4136***
S.E.	(0.0830)	(0.0956)	(0.0818)
<b>Sample statistics</b>			
Observations	5,221	3,963	5,269
Persons	2,790	2,116	2,815
First-stage <i>F</i> statistic	26.46	26.49	25.56

*Notes:* Each column gives the estimation result of a separate regression. The dependent variable is given in the header. Explanatory variable (second stage): instrumented number of days of sickness absence. Explanatory variable (first stage): number of benign months in the school year. Clustered standard errors (S.E.) in parentheses. Clustering was conducted on the individual level. Significance: \* $p \leq 0.1$ , \*\* $p \leq 0.05$ , \*\*\* $p \leq 0.01$ .

writing. While the first stage is the same for each subject, the second of the average points seems to be driven by the second stage of the read and speak results. The effect of (instrumented) sickness read and speak points is 3.3% and statistically significant at the 10% level. The effect size on math points is somewhat and, therefore, not statistically significant (the standard error is the same). For writing the effect seems positive. However, the estimation is based on fewer observations and not the  $t$  statistic in below 1.

Table 6 also presents the IV effect of sickness absence on income and mortality. An additional day of sickness absence is associated with a 0.85 percentage points higher probability of being dead before the age of 70. This exceeds the OLS/FE results of Table 5 0.01-0.03 percentage points clearly. Again, this is the effect for compliers and does not necessary apply for the average person. The effect of sickness absence on income is positive but based on only three-quarter of the observations and is not statistically significant.

All in all, the IV results are in line with the literature on absence in school and provide evidence that the finding for absence in general also holds for sickness absence.

## 5 Heterogeneity

TO BE DONE.

## 6 Conclusions

This paper analyzes the effect of days of (sickness) absence during primary school on short-term educational performance. Utilizing self-gathered register data from Sweden in the 1930s and 1940s allows us to extend the literatures in two ways: First, unlike previous studies, we are able to distinguish between total days of absence and sickness absence. Second, we can also provide evidence of effect of absence on and long-term economic success and health. In order to overcome endogeneity problems we employ various fixed effects and instrumental variables strategies. Estimating the effect of absence on performance in school, we only use variation in days of absence and test scores between the grades at the individual level. Regarding income and mortality as long-term outcomes, we only compare people within the same class in grade 4.

For a long time, policy makers and researchers focused on the extension of the school year length as instrument to increase instruction time (and thus performance). However, evaluations of school year extension suggest that such interventions pick-up a lot of resources and improve education barely, if at all. More recently, [Goodman \(2014\)](#) and [Aucejo and Romano \(2014\)](#) investigate individual absence as reason for a shorter instructional time. [Aucejo and Romano \(2014\)](#) find that the decrease in performance due to one day of absence is ten times bigger than the return to an additional day of instruction. [Goodman \(2014\)](#) can only distinguish between collective absence due to school closures and individual absence. [Aucejo and Romano \(2014\)](#) know whether the student misses instruction excused or unexcused. We, on the other hand, discriminate between the effect of sickness (and not being able to learn) and the effect of missed instruction. From a policy perspective, this distinction might be of special interest when trying to implement strategies to reduce missed instruction time or enable students to learn missed content.

We find that test results decrease on average 4.4% of a standard deviation due to sickness absence.<sup>15</sup> Taking [Aucejo and Romano \(2014\)](#)'s returns to additional instruction for granted, this correspondent to a 30 days school year extension. Transferring [Aucejo and Romano \(2014\)](#)'s results to our sample might not be perfectly accurate; however, it gives an intuition of the importance of absence. Moreover, we find that the effect of sickness absence equals the effect of being absent for other reasons. This indicates that sickness itself may not cause the decrease in performance but missed instruction drives the results. Comparing the effects for math and language with less important subjects, the effect of (sickness) absence is about the same. We may interpret this (although, we have no direct evidence) as a hint that children do not learn (more) when they miss instruction. Besides reducing absence in school, education policy should aim at improving the opportunities and incentives to learn at home when not being in class.

Unlike previous literature, our data source allows to investigate long-term consequences of absence in school. Additional days of sickness absence have a negative effect on retirement income and a positive effect on the probability of being death before reaching the age of 70. Although, both effects are not statistical significantly different from zero we would argue that the economic theory conforming signs of the coefficients and the reasonable magnitude indicate long-term effects of absence in school.

---

<sup>15</sup>On average, students miss 11 days per school year due to sickness, where the per-day effect is 0.4% of a SD in the OLS model.

## References

- Andreella, C., Karlsson, M., Nilsson, T., and Westphal, M. (2014). Intergenerational Transmission of Health in Times of Crisis. Mimeo.
- Aucejo, E. and Romano, T. (2014). Assessing the Effect of School Days and Absences on Test Score Performance. CEP Discussion Papers 1302, Centre for Economic Performance, London School of Economics.
- Bhalotra, S., Karlsson, M., and Nilsson, T. (2014). Life Expectancy and Mother-Baby Interventions. CINCH Working Paper Series 1404, CINCH–Health Economics Research Center.
- Bruckner, T., van den Berg, G., Smith, K., and Catalano, R. (2014). Ambient temperature during gestation and cold-related adult mortality in a swedish cohort, 1915-2002. *Social Science & Medicine*, 119(0):191 – 197.
- Carlsson, M., Dahl, G., and Rooth, D.-O. (2012). The Effect of Schooling on Cognitive Skills. NBER Working Papers 18484, National Bureau of Economic Research, Inc.
- Clark, D. and Royer, H. (2013). The Effect of Education on Adult Mortality and Health: Evidence from Britain. *American Economic Review*, 103(6):2087–2120.
- Currie, J. (2009). Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development. *Journal of Economic Literature*, 47(1):87–122.
- Eide, E. and Showalter, M. (2011). Estimating the relation between health and education: What do we know and what do we need to know? *Economics of Education Review*, 30(5):778–791.
- Fischer, M., Karlsson, M., and Nilsson, T. (2013). Effects of compulsory schooling on mortality: Evidence from sweden. *International Journal of Environmental Research and Public Health*, 10(8):3596–3618.
- Fitzpatrick, M., Grissmer, D., and Hastedt, S. (2011). What a difference a day makes: Estimating daily learning gains during kindergarten and first grade using a natural experiment. *Economics of Education Review*, 30(2):269–279.
- Goodman, J. (2014). Flaking Out: Student Absences and Snow Days as Disruptions of Instructional Time. NBER Working Papers 20221, National Bureau of Economic Research, Inc.

- Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy*, 80(2):223–55.
- Heckman, J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312(5782):1900–1902.
- Heckman, J. (2007). The economics, technology, and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences*, 104(33):13250–13255.
- Imbens, G. and Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.
- Jürges, H., Kruk, E., and Reinhold, S. (2013). The effect of compulsory schooling on health – evidence from biomarkers. *Journal of Population Economics*, 26(2):645–672.
- Kamhöfer, D., Schmitz, H., and Westphal, M. (2014). The Non-Monetary Returns to College Education Using the Marginal Treatment Effect Approach. Mimeo.
- Kemptner, D., Jürges, H., and Reinhold, S. (2011). Changes in compulsory schooling and the causal effect of education on health: Evidence from Germany. *Journal of Health Economics*, 30(2):340–354.
- Lavy, V. (2012). Expanding School Resources and Increasing Time on Task: Effects of a Policy Experiment in Israel on Student Academic Achievement and Behavior. NBER Working Papers 18369, National Bureau of Economic Research.
- Lleras-Muney, A. (2005). The Relationship Between Education and Adult Mortality in the United States. *The Review of Economic Studies*, 72(1):189–221.
- Lleras-Muney, A. (2006). Erratum: The Relationship between Education and Adult Mortality in the United States. *The Review of Economic Studies*, 73(3):847.
- Meghir, C., Palme, M., and Simeonova, E. (2012). Education, Health and Mortality: Evidence from a Social Experiment. NBER Working Papers 17932, National Bureau of Economic Research, Inc.
- Pischke, J.-S. (2007). The Impact of Length of the School Year on Student Performance and Earnings: Evidence From the German Short School Years. *The Economic Journal*, 117(523):1216–1242.

- Staiger, D. and Stock, J. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3):557–586.
- van Leeuwen, M., Maas, I., and Miles, A. (2002). *HISCO: Historical International Standards Classification of Occupations*. Leuven University Press, Leuven.
- Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, Massachusetts.

## Appendix

Table A1: Estimation results for performance in math

	Conventional OLS (1)	School FE (2)	Class FE (3)	Individual FE (4)	Lagged variables (5)
<b>Effect of total days of absence on math points</b>					
Coefficient	−0.0039***	−0.0053***	−0.0056***	−0.0069***	−0.0078***
S.E.	(0.0011)	(0.0012)	(0.0012)	(0.0014)	(0.0019)
<b>Effects of selected control variables</b>					
Female	0.2769*** (0.0254)	0.0881*** (0.0274)	0.0906*** (0.0259)		0.0599 (0.0366)
SES: middle	−0.0051 (0.0315)	0.0312 (0.0371)	0.0200 (0.0343)		0.0871** (0.0418)
SES: high	0.1951*** (0.0433)	0.1839*** (0.0505)	0.2014*** (0.0403)		0.2970*** (0.0641)
Grade 4	−0.1144* (0.0623)	−0.1368 (0.0979)			
Lagged absence					0.0041** (0.0018)
Lagged points					0.7535*** (0.0315)
<b>Other conditioning variables</b>					
Time-variant	✓	✓	✓	✓	
Time-invariant	✓	✓			✓
<b>Fixed effects</b>					
Birth year & month	✓	✓			✓
Age	✓		✓	✓	✓
Birth year mother/father	✓	✓			✓
Parish	✓				✓
School		✓			
Class			✓		
Individual				✓	
Lagged absence & points					✓
<b>Sample size</b>					
Observations	5,978	5,978	5,978	5,978	2,989
Persons	2,989	2,989	2,989	2,989	2,989

*Notes:* Dependent variable: points over in math in standard deviations. Explanatory variable: number of days of sickness absence. Time-variant conditional variables: range of grades instructed in the same classroom, length of the school year in weeks. Time-invariant conditional variables: female, born in wedlock, twin birth, mother employed at the time of birth, born in hospital, mother or father Finnish descent. Baseline category of father's occupation (occu.): unskilled or low-skilled production worker. Clustered standard errors (S.E.) in parentheses. Clustering was conducted on the level of the fixed effects (FE) in columns (2) to (4) and on individual level in columns (1) and (5). Significance: \* $p \leq 0.1$ , \*\* $p \leq 0.05$ , \*\*\* $p \leq 0.01$ .

Table A2: Estimation results for performance in reading &amp; speaking

	Conventional OLS (1)	School FE (2)	Class FE (3)	Individual FE (4)	Lagged variables (5)
<b>Effect of total days of absence on reading &amp; speaking points</b>					
Coefficient	−0.0038***	−0.0032***	−0.0021*	−0.0070***	−0.0044**
S.E.	(0.0011)	(0.0011)	(0.0012)	(0.0014)	(0.0017)
<b>Effects of selected control variables</b>					
Female	0.2734*** (0.0254)	0.3139*** (0.0379)	0.3254*** (0.0253)		0.3159*** (0.0327)
SES: middle	−0.0067 (0.0312)	−0.0270 (0.0316)	−0.0370 (0.0301)	(0.0309)	0.0015 (0.0387)
SES: high	0.1961*** (0.0432)	0.1495*** (0.0553)	0.1413*** (0.0413)		0.1843*** (0.0559)
Grade 4	−0.1215* (0.0625)	−0.1344 (0.1269)			
Lagged absence					0.0056*** (0.0016)
Lagged points					0.5892*** (0.0258)
<b>Other conditioning variables</b>					
Time-variant	✓	✓	✓	✓	
Time-invariant	✓	✓			✓
<b>Fixed effects</b>					
Birth year & month	✓	✓			✓
Age	✓		✓	✓	✓
Birth year mother/father	✓	✓			✓
Parish	✓				✓
School		✓			
Class			✓		
Individual				✓	
Lagged absence & points					✓
<b>Sample size</b>					
Observations	5,980	5,980	5,980	5,980	2,990
Persons	2,990	2,990	2,990	2,990	2,990

*Notes:* Dependent variable: points in reading & speaking in standard deviations. Explanatory variable: number of days of sickness absence. Time-variant conditional variables: range of grades instructed in the same classroom, length of the school year in weeks. Time-invariant conditional variables: female, born in wedlock, twin birth, mother employed at the time of birth, born in hospital, mother or father Finnish descent. Baseline category of father's occupation (occu.): unskilled or low-skilled production worker. Clustered standard errors (S.E.) in parentheses. Clustering was conducted on the level of the fixed effects (FE) in columns (2) to (4) and on individual level in columns (1) and (5). Significance: \* $p \leq 0.1$ , \*\* $p \leq 0.05$ , \*\*\* $p \leq 0.01$ .

Table A3: Estimation results for performance in writing

	Conventional OLS (1)	School FE (2)	Class FE (3)	Individual FE (4)	Lagged variables (5)
<b>Effect of total days of absence on writing points</b>					
Coefficient	−0.0037***	−0.0029**	−0.0027*	−0.0051***	−0.0045**
S.E.	(0.0013)	(0.0012)	(0.0016)	(0.0018)	(0.0023)
<b>Effects of selected control variables</b>					
Female	0.3295*** (0.0336)	0.3980*** (0.0437)	0.4139*** (0.0316)		0.4388*** (0.0478)
SES: middle	0.0235 (0.0437)	0.0430 (0.0446)	0.0328 (0.0415)		0.0722 (0.0547)
SES: high	0.2403*** (0.0580)	0.2316*** (0.0869)	0.2294*** (0.0547)		0.2495*** (0.0826)
Grade 4	−0.2593*** (0.0850)	−0.1938*** (0.0687)			
Lagged absence					0.0037* (0.0022)
Lagged points					0.6385*** (0.0398)
<b>Other conditioning variables</b>					
Time-variant	✓	✓	✓	✓	
Time-invariant	✓	✓			✓
<b>Fixed effects</b>					
Birth year & month	✓	✓			✓
Age	✓		✓	✓	✓
Birth year mother/father	✓	✓			✓
Parish	✓				✓
School		✓			
Class			✓		
Individual				✓	
Lagged absence & points					✓
<b>Sample size</b>					
Observations	3,486	3,486	3,486	3,486	1,743
Persons	1,743	1,743	1,743	1,743	1,743

*Notes:* Dependent variable: points in writing in standard deviations. Explanatory variable: number of days of sickness absence. Time-variant conditional variables: range of grades instructed in the same classroom, length of the school year in weeks. Time-invariant conditional variables: female, born in wedlock, twin birth, mother employed at the time of birth, born in hospital, mother or father Finnish descent. Baseline category of father's occupation (occu.): unskilled or low-skilled production worker. Clustered standard errors (S.E.) in parentheses. Clustering was conducted on the level of the fixed effects (FE) in columns (2) to (4) and on individual level in columns (1) and (5). Significance: \* $p \leq 0.1$ , \*\* $p \leq 0.05$ , \*\*\* $p \leq 0.01$ .

Table A4: IV estimation results by subject

	Math	Read & speak	Writing
<b>Effect of days of instrumented sickness absence</b>			
Coefficient	−0.0285	−0.0332*	0.0124
S.E.	(0.0186)	(0.0181)	(0.0183)
<b>First-stage effect of benign months</b>			
Coefficient	−0.4246***	−0.4243***	−0.5292***
S.E.	(0.0830)	(0.0830)	(0.1097)
<b>Sample statistics</b>			
Observations	5,219	5,221	3,126
Persons	2,789	2,790	1,671
First-stage <i>F</i> statistic	26.19	26.17	23.26

*Notes:* Each column gives the estimation result of a separate regression. The dependent variable is given in the header. Explanatory variable (second stage): instrumented number of days of sickness absence. Explanatory variable (first stage): number of benign months in the school year. Clustered standard errors (S.E.) in parentheses. Clustering was conducted on the individual level. Significance: \* $p \leq 0.1$ , \*\* $p \leq 0.05$ , \*\*\* $p \leq 0.01$ .

För folkskolan.

*A. V. Vahlström*  
(Ordförande)

*B. J.*  
(Lärare)

Under redovisningsåret 1917/18 för skolexamen i  
*Teori*  
(Skolexamen)

Under redovisningsåret 1917/18 för skolexamen i  
*Praktik*  
(Skolexamen)

**DAGBOK MED EXAMENSKATA**

för klass 3-4.

under redovisningsåret 1917/18 — 30/6 1918 omfattande höstterminen  
och vårterminen 1918 — 1/6

antal läseveckor 39

Lärosvardningen har haft eget klassrum — delat klassrum med—

**Anvisningar.**

1. Elefanten är uppdelad i artiklar för en läsningstid på 10 minuter.  
2. Den lärare, som har den kommande undersökningen kom-  
mande dag på följande, för att kunna förbereda sig på den.  
3. Elefanten är uppdelad i artiklar för en läsningstid på 10 minuter.  
4. Den lärare, som har den kommande undersökningen kom-  
mande dag på följande, för att kunna förbereda sig på den.  
5. Elefanten är uppdelad i artiklar för en läsningstid på 10 minuter.  
6. Den lärare, som har den kommande undersökningen kom-  
mande dag på följande, för att kunna förbereda sig på den.  
7. Elefanten är uppdelad i artiklar för en läsningstid på 10 minuter.  
8. Den lärare, som har den kommande undersökningen kom-  
mande dag på följande, för att kunna förbereda sig på den.  
9. Elefanten är uppdelad i artiklar för en läsningstid på 10 minuter.  
10. Den lärare, som har den kommande undersökningen kom-  
mande dag på följande, för att kunna förbereda sig på den.  
11. Elefanten är uppdelad i artiklar för en läsningstid på 10 minuter.  
12. Den lärare, som har den kommande undersökningen kom-  
mande dag på följande, för att kunna förbereda sig på den.  
13. Elefanten är uppdelad i artiklar för en läsningstid på 10 minuter.  
14. Den lärare, som har den kommande undersökningen kom-  
mande dag på följande, för att kunna förbereda sig på den.  
15. Elefanten är uppdelad i artiklar för en läsningstid på 10 minuter.  
16. Den lärare, som har den kommande undersökningen kom-  
mande dag på följande, för att kunna förbereda sig på den.  
17. Elefanten är uppdelad i artiklar för en läsningstid på 10 minuter.  
18. Den lärare, som har den kommande undersökningen kom-  
mande dag på följande, för att kunna förbereda sig på den.  
19. Elefanten är uppdelad i artiklar för en läsningstid på 10 minuter.  
20. Den lärare, som har den kommande undersökningen kom-  
mande dag på följande, för att kunna förbereda sig på den.  
21. Elefanten är uppdelad i artiklar för en läsningstid på 10 minuter.  
22. Den lärare, som har den kommande undersökningen kom-  
mande dag på följande, för att kunna förbereda sig på den.  
23. Elefanten är uppdelad i artiklar för en läsningstid på 10 minuter.  
24. Den lärare, som har den kommande undersökningen kom-  
mande dag på följande, för att kunna förbereda sig på den.  
25. Elefanten är uppdelad i artiklar för en läsningstid på 10 minuter.  
26. Den lärare, som har den kommande undersökningen kom-  
mande dag på följande, för att kunna förbereda sig på den.  
27. Elefanten är uppdelad i artiklar för en läsningstid på 10 minuter.  
28. Den lärare, som har den kommande undersökningen kom-  
mande dag på följande, för att kunna förbereda sig på den.  
29. Elefanten är uppdelad i artiklar för en läsningstid på 10 minuter.  
30. Den lärare, som har den kommande undersökningen kom-  
mande dag på följande, för att kunna förbereda sig på den.

	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62
--	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Figure A1: Example of an examination record

Passed...													Not passed		
with great distinction		with distinction		with great credit			with credit			without credit					
A	A-	a	a-	AB+	AB	AB-	BA+	BA	BA-	B+	B	B-	BC	C	
↓	↓												↓	↓	
15	14	...											...	2	1

Figure A2: Grading scheme

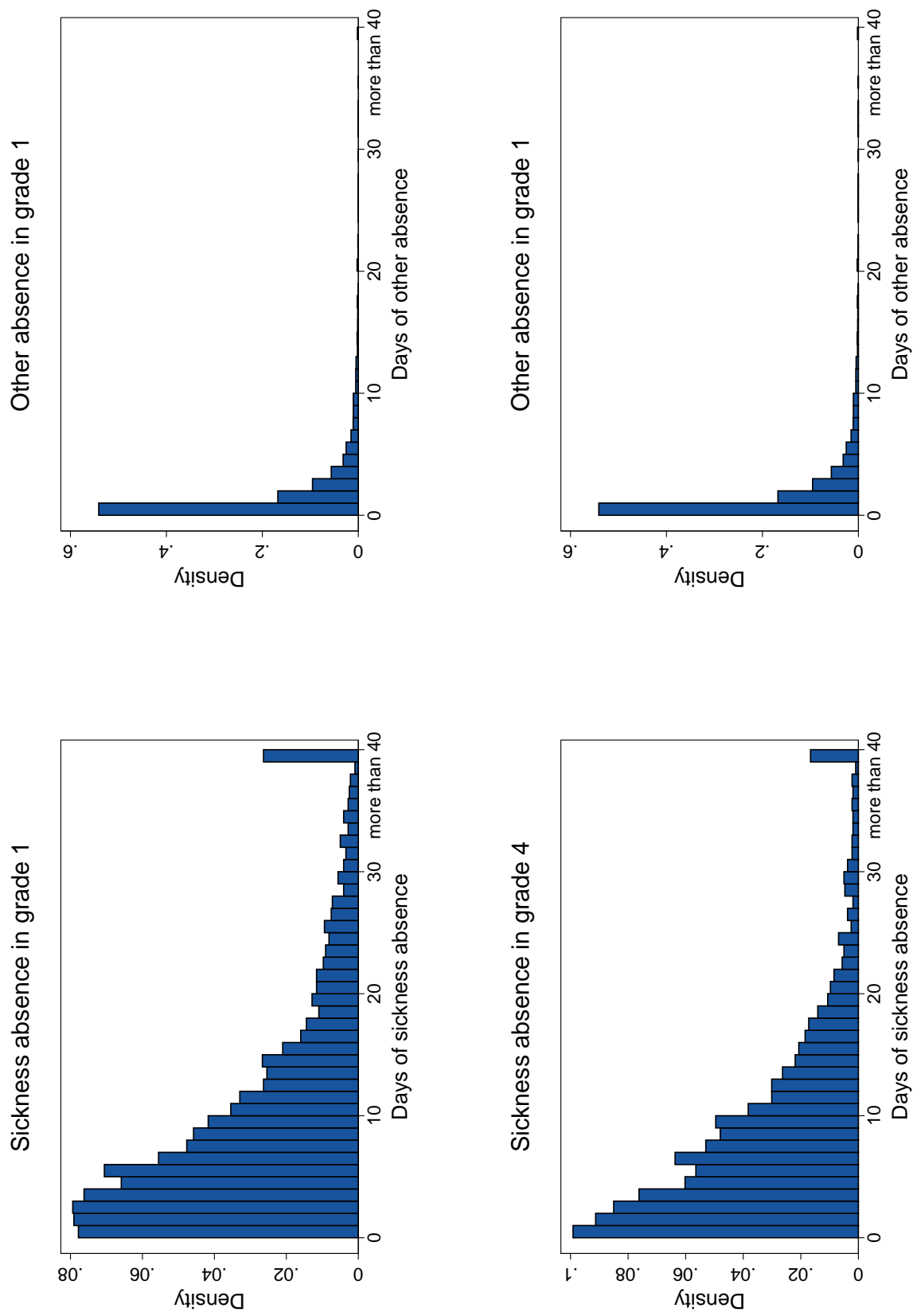


Figure A3: Distributions of absence by grade and reasons

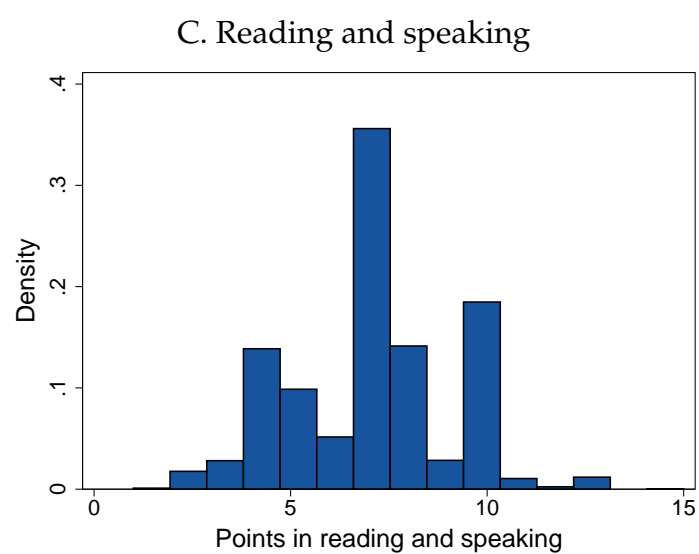
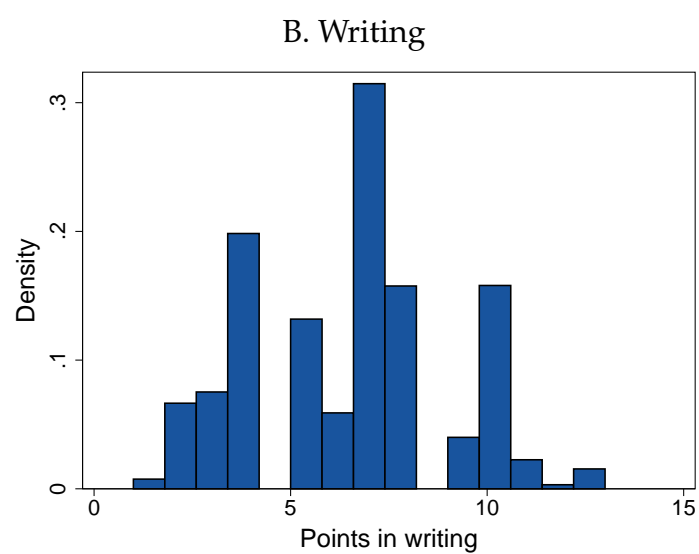
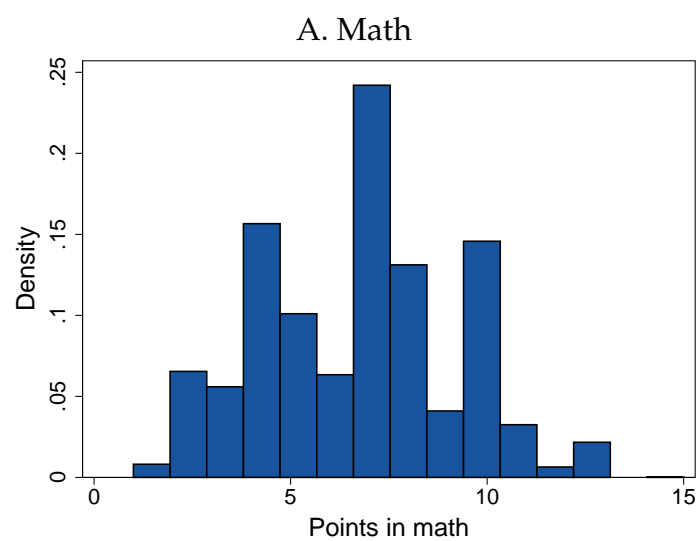


Figure A4: Distribution of scored points by subject

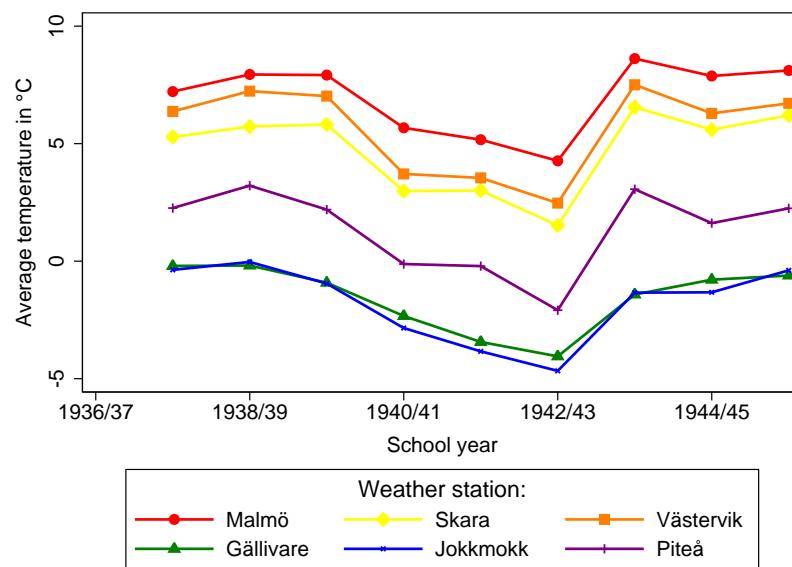


Figure A5: Average temperature over the time for selected weather stations

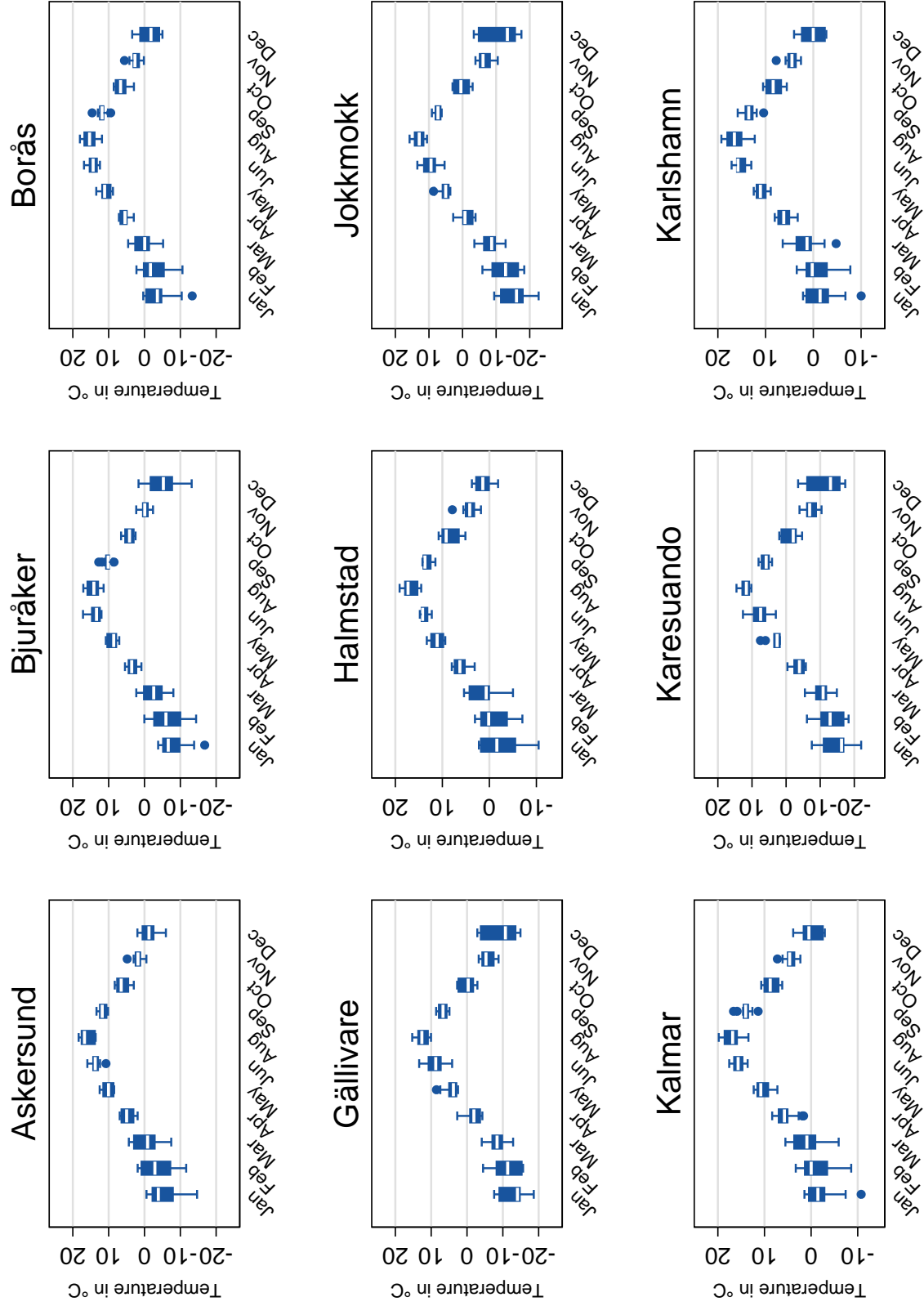


Figure A6: Monthly box plots of temperatures by region

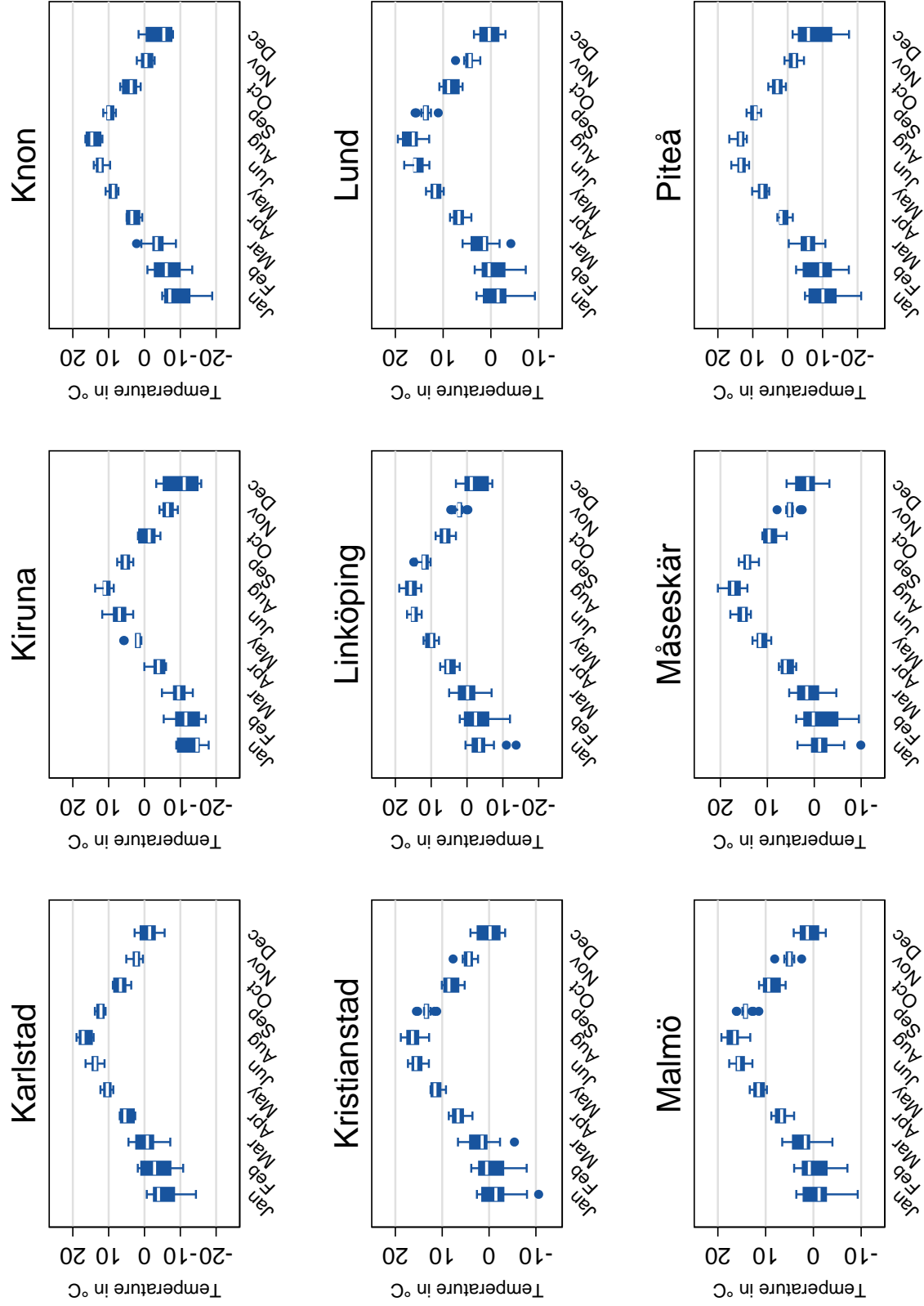


Figure A6: *continued*

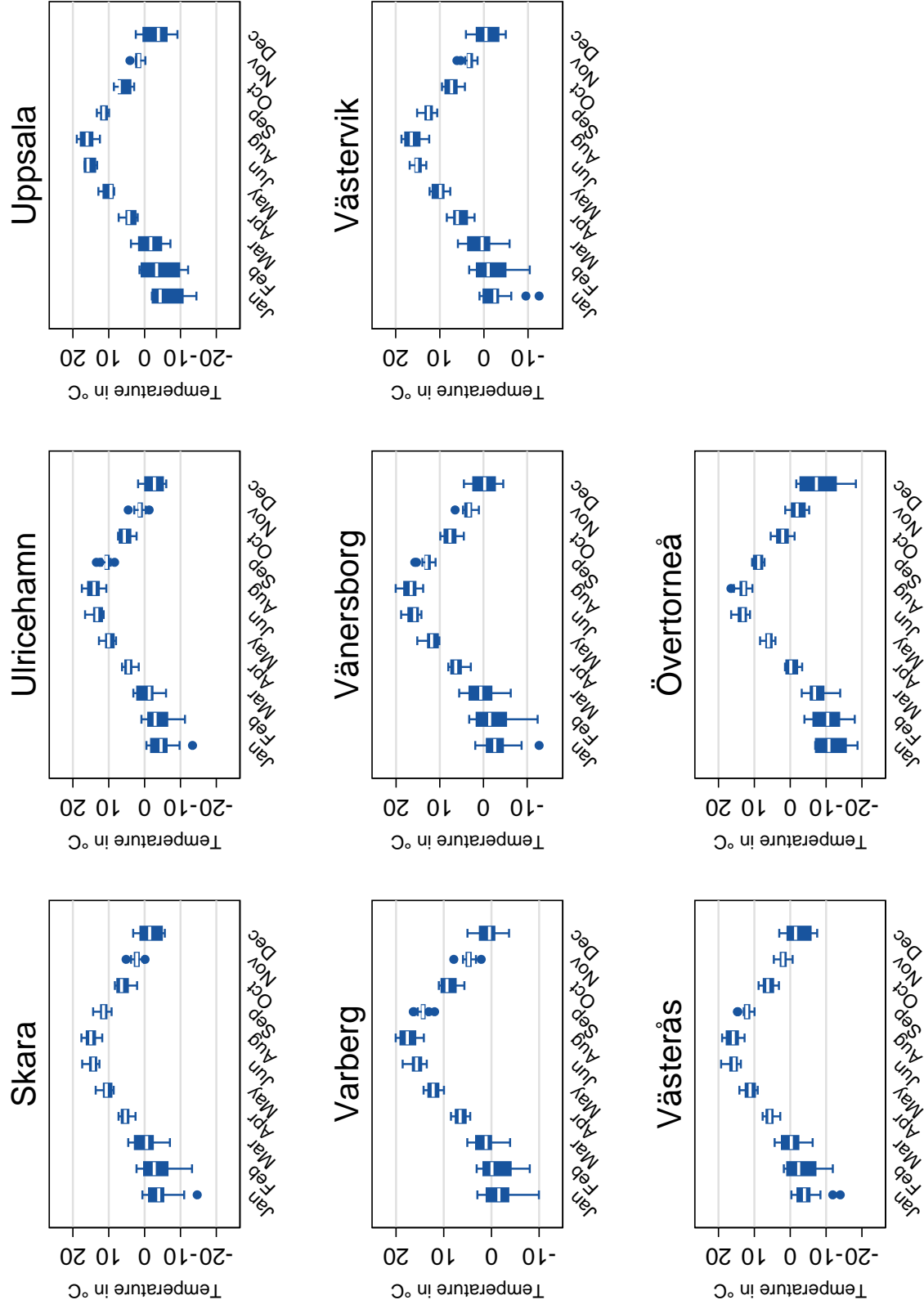


Figure A6: *continued*

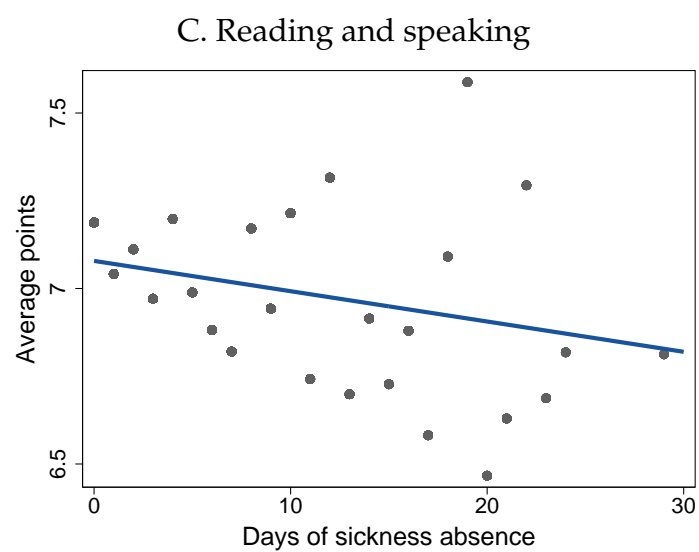
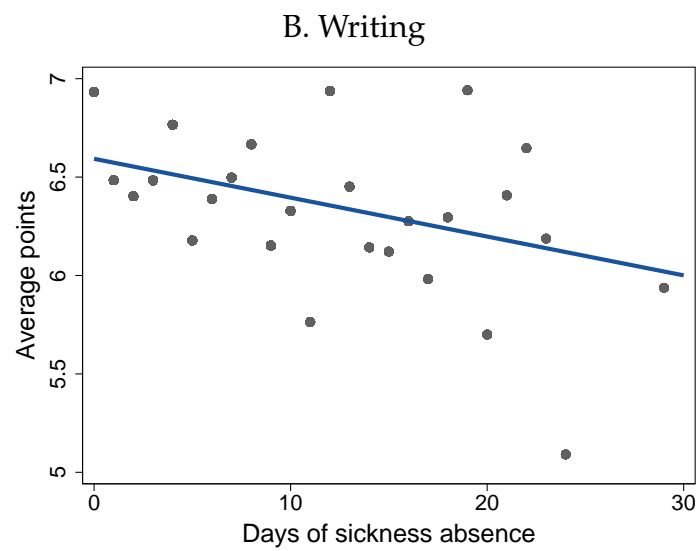
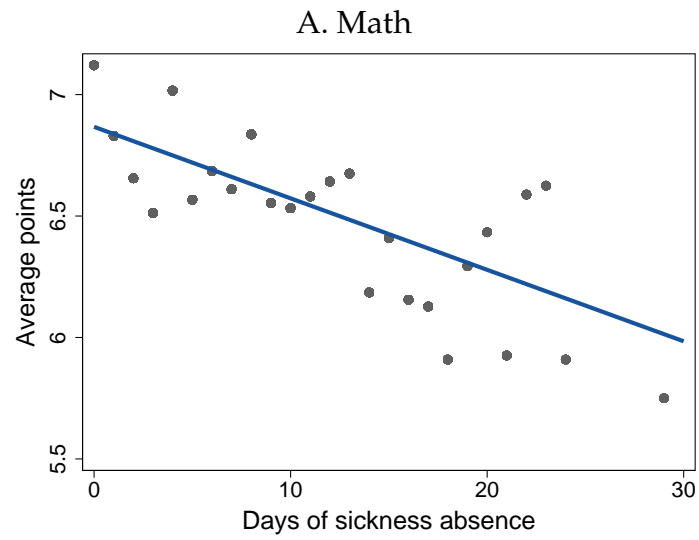


Figure A7: Relationship between days of sickness absence and points by subject