Absence and Completion among students in Vocational Education

Preliminary Draft. Please do not distribute without the authors' permission.

Fane N. Groes*[1], Edith Madsen[1], and Tróndur M. Sandoy[2]

[1]*Department of Economics, Copenhagen Business School, Denmark*
[2]*Department of Economics, The University of the Faroe Islands, Faroe Islands*

This Version: June 6, 2023

Abstract

We analyze the effect of school absence on program completion among a group of students in Vocational Education and Training (VET) in Denmark. According to human capital theory, being present in class and participating in class activities is an important determinant of human capital formation and therefore the causal effect of absence on educational performance is of interest. To analyze this effect we use data on daily student attendance from the administrative systems of VET schools in combination with register data on completion and student background characteristics. There is a very strong correlation between absence and completion. In order to identify the causal effect of absence on completion, we use local weather conditions such as precipitation and wind as instruments for student absences. We further introduce a new instrument for absences that uses variation over time in absences for the individual student to support our results. We find that absences during the first two weeks of a 20-week vocational school introductory program has large and significant causal effects on the probability of graduation from the program.

Keywords: Economics of Education, Vocational Education and Training, Absence, Instrumental Variables.
JEL classification: I20, I21, I28, C26.

1 Introduction

In this paper, we estimate the causal effect of school absence on the probability of program completion for students in vocational education in Denmark. The impact of absence on educational achievement is of obvious interest since most educational programs are based on the

*Corresponding author: Copenhagen Business School, Department of Economics, Porcelænshaven 16a, 2000 Copenhagen, Denmark. E-mail: fg.eco@cbs.dk

assumption that coming to and participating in class is an important input in the skill formation process of the individual student. This together with the fact that student absence in many education programs is substantial underlines the importance of providing evidence of the impact of students being absent from class. Moreover, the existing research on the topic is for students in elementary or academic upper secondary education and in general, there is limited research on student outcomes in vocational education programs.

In our analysis, we consider the basic introductory program in specific vocational programs. This is the focus of our analysis, as there is a very large dropout from this part of the vocational education, see, e.g., Confederation of Danish Employers (2023) and Groes et al. (2021). The basic course is school-based, runs over 20 weeks and it is compulsory for all students. This part of vocational education is taking place at the vocational schools and the classes consists of both practical and theoretical parts. Our outcome is whether the students complete this part of the vocational education. We use data on daily student attendance from the administrative systems at vocational schools in combination with administrative register data from Statistics Denmark on education spells and student background characteristics. Using two different types of instrumental variable estimation strategies, we show that, for students in vocational education, the causal effect of school absence on the probability of program completion is large and significant.

Estimating the causal effect of absences is potentially plagued by endogeneity of absence in that the same unobserved factors that affect individual student absence may also be the ones that affect the student's school outcomes. Examples of such factors are the student's ability level and motivation for learning. Analysis of the causal effect of absence on educational outcomes is challenging especially because of data availability on educational outcomes in combination with student absence and variation in student absence that is uncorrelated with unobserved student factors.

To overcome the potential selection problem that arises from students self-selecting into being absent from school, which might also correlate with completion, we propose two different types of instruments for school absences. Our preferred strategy is a classical instrumental variable, where we use local weather conditions as instruments. The second is a new instrumental variable approach where we use that student absences are measured repeatedly over time giving a panel dataset of this variable.

In our first strategy we use that the number of days with precipitation and high wind leads to more absences among a group of vocational students. Using our weather variables as instruments for absence, we find a causal effect on the compliers of 1.4-1.9, such that a 10-percentage point increase in absence during the first two weeks of school causes between 14-19 percentage point decreases in the probability of completing the 20 weeks introductory program. When analyzing heterogeneous effects, we find that absence among students who do not live with their parents and students with no previous labor market attachment or school enrollment has the highest correlation with our weather instruments. This supports a hypothesis where students with less support for getting out of the door in the morning and students who are not used to having a scheduled start time for the day are the students driving our first stage results.

The second instrument is a panel data instrument, inspired by Arellano and Bover (1995), where we make use of the fact that even though our outcome measure, completion, is cross-sectional, we have repeated observations over time in absences. The idea is that under some assumptions, we can remove the individual-specific fixed effect from student absences and use this as an instrument for the average individual absence during the first two weeks of school.

Our analysis adds to a small but growing literature estimating the causal effect of school absence on student's educational outcomes. Using panel data which has within student variation over time or between subjects, Cattan et al. (2023) and Liu et al. (2019) find negative effects of absence on short-run academic outcomes and longer-run educational and socioeconomic outcomes for students in elementary, middle, and high school.

Aucejo and Romano (2016) use variation in flu exposure across counties and time to instrument for absences and find that a reduction in absence increases math and reading scores among students in North Carolina public schools.

We contribute to this literature by analyzing the effect of absences in VET programs on the probability of completion. This is an education area characterized by high dropout and with students coming from low socioeconomic backgrounds. Accordingly this area has a potentially high policy interest as students who do not complete vocational school have a high probability of ending up as unskilled, see Groes et al. (2021).

Closest to our first instrumental variable approach is Goodman (2014), which considers elementary and secondary school students and instruments absence with snowfall. Goodman (2014) separates between days with heavy snowfall that causes school closures and days with moderate snowfall that affect individuals' absences. He finds that absences and not school closures affect test scores because school closures affect everyone at the school, and the teachers can change their teaching accordingly. Goodman (2014) argues that heavy snowfall that causes school closures violates the exclusion restriction as an instrument for student absence because the snowfall affects both absence and the learning process through school closure.[1] In our data, we do not have any days with school closures; thus, our weather instruments affect only individual student absences and not school closures.

Another threat to this identification strategy is if the weather event that causes students to be absent from class also causes them to accumulate less human capital (lower productivity) if coming to class. Several studies have shown that weather affects productivity. Dell et al. (2014) and Park et al. (2020) present evidence that high temperatures affect labor productivity and disrupts learning time and Heissel and Norris (2018) show that hours of sunlight in the morning before school increases student test scores. Mellon (2020) surveys the literature using weather as an instrument and encourages a thorough discussion of why the exclusion restriction holds when using weather as an instrument. In our paper, if the weather directly affects student productivity, this would violate our exclusion restriction. We argue that since we use variation in precipitation and wind, which we conjecture only affect how costly it is to get to school, they do not matter for indoor productivity once students are at the school. We believe our results are strengthened by the fact that we use weather over short periods and only throughout Denmark, which does not have large and permanent spatial differences in weather. Furthermore, we include school location and month fixed effects in our analysis, such that we use weather variation within the season (students start at different points in time within a given month and across years) and school location. Since Denmark is a country that often experiences precipitation and high wind,

[1] Kristensen et al. (2020) and Gottfried (2009, 2010) use background characteristics to control for selection into absence or instruments absence with distance to school. Both of these approaches possibly suffer from potential bias in estimating the causal effect of absence on school outcomes.

all schools are built such that students and teachers can comfortably undertake education during such weather conditions. However, extreme rain may affect how wet students and teachers get during their transport to school, which can last for a while during the day and potentially affect productivity. It could also be the case that heavy rain gives delays in the traffic. In that case it could be that teachers are late to class, which in turn affects teacher productivity. There are no extreme weather occurrences in our sample period and we therefore believe that our exclusion restrictions hold, such that weather does not affect student productivity directly during our sample period.[2] Further, we believe that our exclusion restrictions are more likely to hold compared to studies using health shocks as instruments for absences. This is because learning is less likely to be affected when the weather is bad compared to when experiencing a health shock.

The second instrument we propose is inspired by the panel data literature. We are interested in causally estimating the effect of average absence during the first two weeks of school on the probability of completing an introductory course after 20 weeks. Using the assumption of mean-stationarity from Arellano and Bover (1995), we assume that we can divide individual weekly absences into an additive individual-specific fixed effect and a time-varying random component that are independent of each other. Under this assumption, we can remove the individual-specific fixed effect by taking the weekly difference, leaving only the time-varying random part of the weekly absences. If we assume no correlation between the de-meaned part of absences and the regression error in the completion regression, we can use it as an instrument for the average absence. Because we only have repeated observations on the explanatory variables, the assumption of mean-stationarity is stronger than the one required in a classical panel data setting with repeated observations on both the explanatory and the dependent variable.

To our knowledge, we are the first to propose an instrument that requires panel data on the explanatory variable but only cross-sectional data on the dependent variable. Besides identifying the causal effect of absence on completion in this paper, we believe this IV approach can be useful in other settings where researchers have repeated observations of the explanatory variable but only one observation of the outcome of interest. Examples could be, measuring the effect of hours spent in daycare on long-run school outcomes (grades, level of education etc.) or the effect of worker absence on the probability of job promotion. It is important to emphasize that the method only controls for endogeneity that arises because of an individual-specific fixed effect that is constant over time.

Using the panel data instrument we find that the effect of average absence during the first two weeks on completion is negative such that increases in absences decreases the probability of completion. The panel instrument is currently a work in progress. As we have high share of students with zero weekly absence during the first two weeks, the first-difference across weeks does not remove the fixed effect for these students. We are waiting for better data, which will

[2] Sarsons (2015) shows that there is also threats to identification when using weather as in instrument for the effect of income shocks on conflicts in developing countries. In our setting, we assume that weather does not affect income or any other background characteristics. Auffhammer et al. (2013) survey potential pitfalls when using station level weather data. One concern is that the weather is extrapolated using a grid, which together with the underlying data process of weather causes spatial multi-collinearity that can lead to increased standard errors on the weather. In our data, the weather at the different school locations is indeed highly correlated on a given day. However, we obtain variation in weather across schools for the same starting month because students start at different dates across educations and schools.

help us overcome this issue.

To sum up our findings, using precipitation and wind as instruments for absence, we find a large and significant negative effect of absence during the first two weeks on the probability of completion from the second basic course. We support our results with a new panel instrument that is still work in progress. The large effects of absence on program completion from using the weather instruments, suggest high pay-offs for policy interventions encouraging vocational education students to attend classes during the first weeks of school. We leave this for future research.

The paper mainly contributes to three strands of the literature in addition to the bodies of work we have discussed above. First, it contributes to the literature on what causes absence from school. Currie et al. (2009) shows that days with high Carbon Monoxide (CO) increase school absences for public school students in Texas and Zimmer (2019) finds that visiting a doctor increase the absences among children ages 6 to 13. There is also a small literature on how teacher quality affects student absenteeism where Gershenson (2016), Tran and Gershenson (2018), and Liu and Loeb (2019) show that teachers significantly affect student absences. Our results include all types of absences, both from sickness and absences that do not have a reason attached to them. The first stage estimation using our weather instruments is similar in nature to estimating the effect that Carbon Monoxide has on absences. The absences from doctor visits and teacher quality should not affect the validity of using weather as an instrument as long as neither doctor visits nor teacher quality is related to daily changes in the weather. Our panel instrument uses individual variation over time, so this instrument could potentially use variation from doctor visits, but the fixed effect of teacher quality will be removed when de-meaning student absences. Finally, experimental evidence has found that changing parental beliefs can reduce students' absences in the early grades. Rogers and Feller (2016) and Robinson et al. (2018) show that a parent-focused intervention on the beliefs about the importance of school attendance and their children's placement in the absentee distribution significantly decreases chronic absenteeism. Due to the large effects, we find of student absences during the first two weeks of enrollment in vocational school, we believe that the vocational education program in Denmark is another obvious place to introduce interventions that reduce student absence.

Second, a complementing strand of the literature analyzes the effect of school days or instruction time on educational outcomes. Analyzing the effect of school days on student performance, Marcotte (2007), Hansen (2011), and Marcotte and Hemelt (2008) use snow days to instrument for school closures and find that number of school days before the exam positively affect the student test scores. Using variation in exam dates and intelligence tests given in the military, Fitzpatrick et al. (2011) and Carlsson et al. (2015) show that more school days before the exam increase students' test scores. Groppo and Kraehnert (2017) use difference-in-differences to show that severe winters in Mongolia affect the medium- and long-run education outcomes, while Craig and Martin (2019) find that eliminating student suspension increases student test scores. A different way of increasing students' time in school is by expanding the instruction time during the day. Using reforms that change the school day length, Dominguez. Patricio and Ruffini (2018) show there is a positive effect on educational attainment by increasing the school day in Chilean elementary and secondary schools and Lavy (2020) shows that increased time at school with more school resources lead to increase student achievement. Finally, Lavy (2015), Rivkin and Schiman (2015), and Bingley et al. (2018) utilize within student differences in taught hours across subjects to find a positive effect on test scores and that these positive effects vary by the classroom environment. Our identification strategy is also closely related to this last strand of literature. However, instead of using within student variation and weather as exogenous changes in school length, we use the variation to predict student absence, holding the school length and instruction time constant. Goodman (2014) and Aucejo and Romano (2016) both show that the effect of reducing absence is larger than a comparable effect of increasing instruction time. With this in mind, the fact that we find particularly large effects of absences suggests we would not find as large effects if we instead increased the instruction time in the vocational education program.

Finally, this paper also contributes to the literature on student dropout from vocational schools. Denmark, like Germany, Switzerland, and Austria, has what Eichhorst et al. (2015) refers to as a dual system for the VET, which is characterized by a high degree of formalization, vocational schools that provide the school-based part of the dual apprenticeship, and accreditation of the firms that students train in during their apprenticeship. Vocational schools in Denmark have a large dropout rate and are constantly under pressure to increase student completion.

At the same time, the students at vocational schools have relatively high absences at an average rate of 7.5 percent of days during the first two weeks of school increasing to around 12 percent at the end of week three. By providing causal evidence of student absence on the probability of completion, our result hopefully contributes to a better understanding of how to help students in vocational education.

By analyzing the causal effect of absence on the completion probability, we contribute to the understanding of the vocational education system, where the economic literature mainly has concentrated on the effect of obtaining a vocational education on labor market outcomes (see Hanushek (2012), Hanushek et al. (2017), Hampf and Woessmann (2017), Bertrand et al. (2021), and Silliman and Virtanen (2022)). Stratton et al. (2017) analyze the probability of completion from vocational school, taking selection on grades from mandatory school into account. They find that prior math scores are particularly important for completion, which we also find in our results and therefore include as one of the background characteristics in our estimation.

The remainder of the paper is organized as follows: Section 2 describes the Danish VET system. Section 3 describes the data we use along with our sample selection criteria. Section 4 shows descriptive statistics. Section 5 describes our empirical strategy. Section 6 describes our results. Finally, section 8 concludes.

2 Institutional setting

In Denmark, the present vocational education and training system (VET) was introduced in 2014 and implemented in August 2015. The VET system consists of more than 100 types of vocational programs; a few examples of specific educations are carpenter, electrician, and hairdresser.

The Danish VET consists of a basic program and the main program. The basic program takes one year and is divided into two separate courses. Each takes 20 weeks (excluding holidays) and consists of theoretical and practical classes where all teaching is exclusively at the vocational school, which is in contrast to the main program where the majority of the time is spent on internships away from the school. The main program will typically take 3-3.5 years and alternate between school courses and apprenticeships, where the latter takes place in a specific company or organization. There is an exam at the end of both basic courses and the end of the main program.

The first basic course (GF1)[3] is very broad and mainly meant to help the students choose a specific education (for example, carpenter), which they will follow from the second basic course and on through the main program. In general, class attendance is compulsory for VET students, and the teachers and student counselors have to monitor student absences and follow up on students being absent from classes. However, there are no strict rules regarding the maximum absence allowed for a VET student.

This project is concerned with students enrolled in the second basic course (GF2). Two possible channels allow students to enter the second basic course. The first channel is through the first basic course, reserved for students who attend VET within two years after completing the mandatory 9th or voluntary 10th grade. The second channel is for students who completed their lower secondary education more than two years earlier. These students start directly on the second part of the basic course, i.e., they skip the first basic course. This means they choose a specific education at the beginning instead of after half a year. Students over 25 years old attend the special VET for adults, which can be with or without the second basic course and possibly with a shorter main program.[4] For older students, the vocational school will decide whether the person needs the second basic course or whether he/she can start directly in the main program. The special VET for adults takes up a large share of the students in the VET system. Having many older students enrolled is very different from the upper secondary education (high school) that prepares students for higher education programs. Here, most students are under 20 years old, as most recently completed lower secondary education.

The different types of students and their different pathways through the VET system imply that many vocational schools have two starting dates for the second basic course within a year: January and August. For example, students who recently completed lower secondary school will typically start the first basic course in August and therefore start the second basic course in January the year after. The other students begin immediately with the second basic course and, therefore, can begin their education in either January or August. Altogether, this means that the mix of students in the second basic course can differ substantially depending on the time of the year.

Another feature of the Danish VET system is that it offers a VET education combining general upper secondary education and vocational education and training (EUX). EUX qualifies students for jobs as skilled workers and gives direct access to higher education within a wide range of programs. In our analysis, we exclude students who combine upper secondary education with VET.[5]

[3]We use GF1 and first basic course interchangeably for the remainder of the paper.

[4]The composition of 25+ study depends highly on their prior education and work experience. An example of this could be a person who has worked several years as an unskilled carpenter and wants to have formal education as a carpenter.

[5]A full overview of the different pathways through the Danish VET system is found in appendix figure 8

3 Data Sources and Sample Selection

For our analysis, we use three different data sources. We have collected data on absence from eight vocational schools that we merge with the Danish register data to get student completion and background characteristics. We also merge our data with daily meteorological observations and a measure of distance from the school.

3.1 VET School data

We collected our primary data on absence at eight primarily large vocational schools in Denmark that offer technical education for this project.[6]

The data covers all individual spells at the schools for 2015-2019, including education, institution, and start date.[7]

The data also contains information on the daily student class schedule, the number of scheduled hours, how many hours the student attended, and the reason for not attending to some extent. Schools classify student absence as either ”excused” (for example illness) or ”not excused” (reason unknown). In this project, we do not distinguish between the types of absence but instead consider total absence. We do this primarily because many of the student absences have missing information on the cause, but also because the type of absence is likely to be subject to miss-classification error. For example, a student might report being ill while the reason for absence is something else. In addition, our starting point is that absence in itself, no matter the cause disrupts the instruction and training of the student.

We consider absence within the first two weeks of classes in the second part of the basic course. We do this for two reasons. Firstly, to avoid sample size issues. Students drop out continuously, so we will not have information on some students for the last weeks. Secondly, it is to avoid having to deal with shocks to the teaching/learning process, which could correlate with both the dropout decision and absence. We elaborate further on the second reason in section 5. Using the collected data, we construct our measures of interest: *individual percentage of absence during the first two weeks of the second basic course*.

3.2 Register data

We combine the school data with the register data from Statistics Denmark. We match the school data by individual id, education, institution, and starting date to the Danish Student Register (KOTRE), where we observe, by dates and school, each educational enrolment spell and any credentials obtained from the students' educations. We can match 95.6 percent of the student spells from the school data to the register data. For the majority of students, we match their spell by the exact matriculation date. For the remaining students, we create matches by allowing the matriculation dates in the school data to differ by up to 21 days compared with

[6]The schools were selected if they had a carpenter education to ensure we had one large education represented at all the schools, to ensure that the number of observations was high.

[7]The data also includes the end date and reason for the end of a spell. We choose not to use this information from the school data because the end of a spell date is unreliable since schools sometimes overwrite the data if a student starts a new education. Further, the variable which contains the reason for ending a spell has many missing observations.

the register data, provided the education and institution are the same and a sufficient number of observations in our data share this matriculation date (we have chosen 15). For the analysis, we use the exact first day as the start date, which is the information we have from the school data.

Completion information is, as mentioned earlier, only complete for some spells in the school data, so we use the information in KOTRE to define our outcome, namely completing the second basic course within seven months. In addition, we use the enrolment status and the matriculation date from the register data to define the timing of the completion.

For background characteristics of the students in the second basic course, we combine data from five different administrative registers. First, from the Danish Student Register, we observe, by dates, each educational spell the students have ever enrolled in and any credentials obtained from these. We use this data to construct the highest completed degree at the time of the first enrollment in the second basic course and an indicator for students coming from the first basic course as well as previous unsuccessful attempts at the second basic course. For individuals who completed 9th grade after 2002, we observe the grades used to qualify for vocational and high school. Among all the grades from the 9th grade, we chose the Danish and math grade received at the end-of-year exam, which are national exams given to all 9th-grade students. Second, from the demographics register (BEF), we extract background characteristics on age, sex, and immigration status. Third, we use the Integrated Database for Labor Market Research (IDA) to observe parents' labor market status. Fourth, we obtain information on the parents' highest completed education level from the education register (UDDA). Lastly, we obtain information on students' primary employment status during the year before enrolment in the second basic course from the AKM register.

3.3 Weather and distance data

We combine our data with weather data from the Danish Meteorological Institute's (DMI) online service. DMI collects daily information on, e.g., wind, precipitation, and temperatures from sensors dispersed over Denmark, so we have access to information on the weather conditions at the municipality of the schools.[8]

For our analysis, we use aggregated measures of our weather observations. However, since we observe both the weather and the students' classes daily, we can use the students' schedules during the first two weeks of school and, for each student, merge the weather data by date and municipality for the days that the students have scheduled classes. Since we can match our weather measures to the school days by date, students with different start dates also have different weather observations during their first two weeks of classes, even if they go to the same school. Further, students who, e.g., only have nine days of scheduled classes are only matched to the weather during those nine days.

We create two primary weather measures for our analysis. We use daily millimeters (mm) of precipitation and high wind speeds, which we average individually over the first two weeks of scheduled school days. The chosen measures are mm of daily precipitation and the highest

[8]To generate municipality-level information, DMI uses information from the weather stations, which are dispersed across the country. Based on the measures obtained from the weather stations, DMI first generates a fine weather grid for the whole country and aggregates it to the municipality level. Thejill et al. (2020) gives a more detailed description of the methods used to generate the grid and municipality aggregation.

average wind speeds in meters per second (ms) in a 10 min. interval.

Lastly, we combine our data with the distances between the parish in which the student lives and the parish the school is located in. We define distance as the bird's flight distance from one parish's centroid to another parish's centroid. We include the distance as a measure of the distance the student has to travel to school, and since most students live relatively close to their respective schools, some of the usual concerns with bird's flight distance, such as crossing the sea, are not a problem. It is worth noting that since we use parish-level measures, we lose the within-parish-level variation in distance, although this is not a big concern as parishes are relatively small areas, and there is still much between parishes variation.

3.4 Sample Selection

For our analysis, we restrict our sample to the period after a reform to the VET area in Denmark, which was implemented on August 1, 2015. The reform introduced a new structure, where the first part of the VET program is now divided into a first and a second basic course. Furthermore, to follow the students up to seven months after their enrollment in the second basic course, we select the last month of enrolment as August 2018 and focus on the two regular starts of the education during August and January.[9] Finally, since our collected sample of schools primarily includes technical schools, we restrict our sample to only include the seven largest VET programs *carpenter, joiner, electrician, smith, plumber, house painter,* and *mason*.

The VET students in the second basic course have a high dropout rate, and the schools have students dropping out every week. Therefore, we select students that stay enrolled in the course during the first six weeks and have at least 15 scheduled hours during each of the first two weeks to avoid including students who have already dropped out in our measure of student absence. This leaves us with a final analysis sample of 5782 observations or $85,915$ daily x student observations.

4 Descriptives

Table 1 shows means and standard deviations for the students' characteristics and our measures of interest, which are absence and completion. Column (1) presents the full sample and Columns (2) and (3) present the sample split by whether or not the students attended the first basic course prior to their enrollment in the second basic course. The chosen VET educations are very male-dominated with on average 89 percent of students being male. Most of the students in our analysis sample are natives (90 percent). The average 9th-grade math and Danish grades are 4.4 and 4.7, which are lower than the grades in the population, where the average overall grade is around seven (7.3 in 2018 (The Danish Ministry of Education, 2018)). Further, around half of students (59%) live with their parents, although this varies a lot by whether the students have attended the first basic course or not. Students have, on average, ten days of school during the first two weeks, corresponding to full school weeks, excluding weekends. Further, on average, students have 5 hours of classes, excluding breaks, each day of which they are absent for around half an hour. During the first two weeks, 43 percent of students have zero hours of absence.

[9]This is an end of sample restriction made at the time we started the project.

Seventy percent of the students in our sample completed within seven months, and this pattern is similar for students with and without the first basic course. Thirty-two percent of the students in our analysis sample have attended the first basic course.

Comparing the students with a prior first basic course in Column (2) to students who start directly at the second basic course in Column (3), we see that they are more likely to be males (96 compared to 86 percent) and on average younger (17 years compared to 22 years). Students with a prior first basic course are also more likely to be natives (94 percent) compared to students without (89 percent), where the difference mainly is in the share of 1st generation immigrants (1 compared to 7 percent) while the share of second-generation immigrants is similar.

Beyond the individual student characteristics, we also include parents' education and employment status, presented in Table 2. Most noticeably, 46 percent of students have a parent where Vocational training is their highest education. In addition, a little over 70 percent of students have mothers and fathers who are employed, around 7 percent have mothers or fathers registered as unemployed, 17 percent have mothers or fathers who are out of the labor force, and 4-5 percent have mothers or fathers with missing observations.

4.1 Pattern in Absences

We next turn to describe the patterns of absence in our data. Table 3 shows the weekly patterns of absences for the first three weeks of school. In panel A, we see that the percentage of absent students increases over the three weeks, with 27, 41, and 48 percent absent for at least one hour during weeks 1, 2, and 3. The average hours of absence also increase over the first three weeks, both unconditionally and conditionally on having some absence, while the scheduled number of hours per week remains the same. When we analyze the effect of absence on completion, we use absence in percent of scheduled hours, which is presented in Panel B. Also here, we see that the students increase their absence in percent over the first three weeks, both conditionally and unconditionally on having some absence during a given week, although the increase from week 2 to 3, conditionally on having positive hours, is smaller than from week 1 to 2. In panel C, we see how the individuals' absences correlate over weeks. The first row in Panel C presents the weekly probability of having some absence for students who had some absence in week 1. For these students, the probability of absence during week 1 is 100 percent by definition. For weeks 2 and 3, they are 43 percent and 38 percent. Comparing these percentages to the overall sample in row 1 of panel A, we can see that when we condition on students with some absence in week 1, the share of students with some absence in week 2 is similar to the unconditional share, while the share of students with some absence in week 3 is lower than the unconditional share. This pattern is different for students with some absence during weeks 2 and 3, presented in rows 2 and 3 in Panel C. For both these types of students, having some absence in either week two or week three is associated with a higher share of students with some absence than the unconditional share in the other two weeks. These statistics suggest that the absences during week 1 are more random than the absences during week two and week three, which seems to be more individual-specific.

We use average absence during the first two weeks as our variable of interest when we analyze the effect of absence on completion. We saw from Table 1 that 43 percent of students have zero absences during the first two weeks. In order to investigate the heterogeneity in

Table 1: Descriptive student characteristics

	Full Mean (SD)	First Basic Course Mean (SD)	No First Basic Course Mean (SD)
Male	0.890	0.959	0.858
	(0.312)	(0.198)	(0.349)
Age	20.635	16.967	22.337
	(5.261)	(0.740)	(5.580)
log(distance (km))	2.408	2.505	2.363
	(1.104)	(0.909)	(1.181)
Native	0.901	0.942	0.883
	(0.298)	(0.233)	(0.322)
Immigrant	0.050	0.013	0.068
	(0.219)	(0.111)	(0.252)
2nd gen. immigrant	0.048	0.045	0.050
	(0.214)	(0.208)	(0.217)
Lives with parents (< 25)	0.590	0.955	0.421
	(0.492)	(0.207)	(0.494)
Unemployed	0.016	0.000	0.024
	(0.126)	(0.000)	(0.152)
In school	0.617	1.000	0.440
	(0.486)	(0.000)	(0.496)
Other	0.123	0.000	0.180
	(0.328)	(0.000)	(0.384)
Employed	0.203	0.000	0.298
	(0.402)	(0.000)	(0.457)
Benefits recipient	0.040	0.000	0.059
	(0.197)	(0.000)	(0.236)
9th grade danish score	4.358	4.245	4.419
	(2.724)	(2.145)	(2.990)
9th grade math score	4.760	4.701	4.793
	(3.081)	(2.603)	(3.314)
Daily hours (2 weeks)	5.189	5.198	5.184
	(0.465)	(0.466)	(0.464)
Daily hours absence (2 weeks)	0.390	0.354	0.406
	(0.652)	(0.570)	(0.686)
Days school (2 weeks)	9.924	9.972	9.901
	(0.338)	(0.208)	(0.381)
Absence> 0 (2 weeks)	0.527	0.529	0.526
	(0.499)	(0.499)	(0.499)
GF1	0.317	1.000	0.000
	(0.465)	(0.000)	(0.000)
Graduated within 7 months	0.695	0.712	0.687
	(0.460)	(0.453)	(0.464)
Observations	5782	1833	3949

Note: The tabel reports mean characteristics and standard deviations in parentheses for the full sample and the sample split by whether the students have a prior spell in the first basic course (second column), or not (third column). We do not observe 9th grade Danish and Math grades for all students, the reasons for this are that the grades were not recorded in the register before 2002 and also some students have not attended 9th grade in Denmark.

Table 2: Descriptive parental characteristics

	Full Mean (SD)	First Basic Course Mean (SD)	No First Basic Course Mean (SD)
Highest completed parental education			
Missing	0.053	0.017	0.069
	(0.223)	(0.131)	(0.253)
Primary school	0.249	0.259	0.244
	(0.432)	(0.438)	(0.430)
Highschool	0.036	0.027	0.040
	(0.185)	(0.163)	(0.195)
Vocational school	0.465	0.556	0.422
	(0.499)	(0.497)	(0.494)
Short track HE	0.051	0.056	0.049
	(0.220)	(0.229)	(0.216)
Prof. BA	0.097	0.063	0.112
	(0.295)	(0.244)	(0.315)
BA	0.006	0.003	0.007
	(0.075)	(0.057)	(0.082)
MA/PhD	0.045	0.017	0.058
	(0.208)	(0.131)	(0.234)
Fathers employment status			
Employed	0.729	0.806	0.693
	(0.444)	(0.395)	(0.461)
Unemployed	0.065	0.063	0.065
	(0.246)	(0.244)	(0.247)
Out of the labor force	0.171	0.124	0.192
	(0.376)	(0.330)	(0.394)
Missing	0.035	0.006	0.049
	(0.185)	(0.077)	(0.216)
Mothers employment status			
Employed	0.707	0.785	0.671
	(0.455)	(0.411)	(0.470)
Unemployed	0.075	0.065	0.080
	(0.264)	(0.247)	(0.271)
Out of the labor force	0.162	0.136	0.175
	(0.369)	(0.343)	(0.380)
Missing	0.055	0.014	0.074
	(0.229)	(0.118)	(0.263)
Observations	5782	1833	3949

Note: The tabel reports mean characteristics and standard deviations in parentheses for the full sample and the sample split by whether the students have a prior spell in the first basic course (second column), or not (third column).

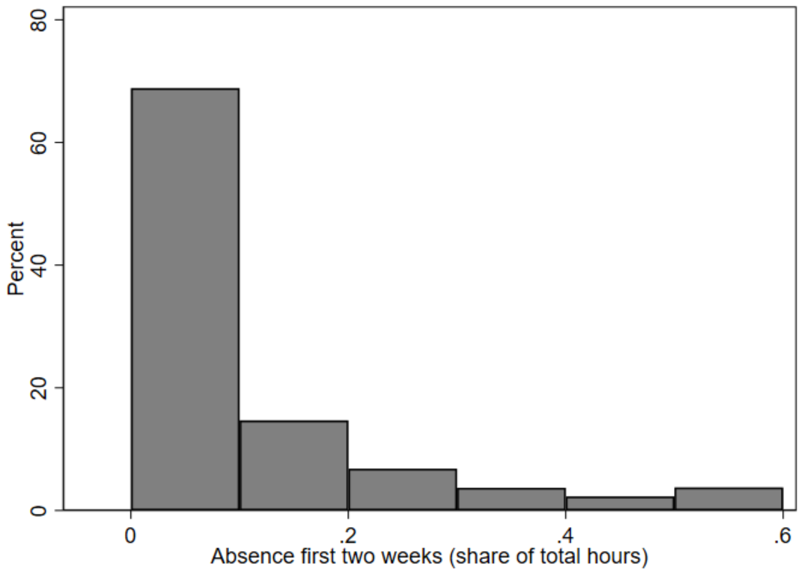
Table 3: Descriptive patterns in student absence

Panel A	Week 1	Week 2	Week 3
	Mean(SD)	Mean(SD)	Mean(SD)
Hours absence> 0	0.268	0.413	0.484
	(0.443)	(0.492)	(0.500)
Hours absence	1.335	2.531	2.979
	(3.268)	(4.650)	(4.783)
Hours absence (cond. > 0)	4.983	6.130	6.156
	(4.658)	(5.505)	(5.264)
Hours	25.705	25.790	25.741
	(2.603)	(2.842)	(3.033)
Panel B	Week 1	Week 2	Week 3
	Mean (SD)	Mean (SD)	Mean (SD)
Pct. absence	0.052	0.098	0.116
	(0.127)	(0.181)	(0.187)
Conditional pct. absence	0.194	0.238	0.240
	(0.181)	(0.214)	(0.206)
Panel C: Conditoinal on hours absence> 0	Share	Share	Share
	Week 1	Week 2	Week 3
Week 1	1.000	0.428	0.388
Week 2	0.659	1.000	0.588
Week 3	0.702	0.689	1.000
Observations with positive absence	1549	2387	2798

Note: Panel A reports means and standard deviations in parentheses for different measures of absence and hours for the first three weeks of the second basic course. Panel B reports means and standard deviations in parentheses for weekly pct. absence and weekly pct. absence conditional on having positive absence. Panel C reports the conditional fractions of observations with positive weekly absence split by the three first weeks. The rows indicate which week is conditioned on and the columns indicate which week the fraction is for. Panel C also reports the number of observations with positive absence in a given week. The table is based on the analysis sample containing 5782 observations.

absence across students, we consider the distribution of average absence within the first two weeks of class. In Figure 1 we show the distribution of average individual absences during the first two weeks in bins of 10 percentage point absences. Figure 1 shows some heterogeneity in student absence. Approximately 65 percent of students are absent for 0 to 10 percent of the hours during the first two weeks, 18 percent of students are absent for 10 to 20 percent of the hours, 8 percent of students are absent for between 20 to 30 percent of the hours, and few students are absent for than 30 percent the hours. Figures 2 and 3 show the distribution of hours by education and by whether the student had prior enrollment in the first basic course. Both Figures show the same patterns as Figure 1, such that there does not seem to be much difference in the absence distribution within the first two weeks across education or prior enrollment in GF1.

Figure 1: Distribution of hours absent as a share of total hours for the first two weeks

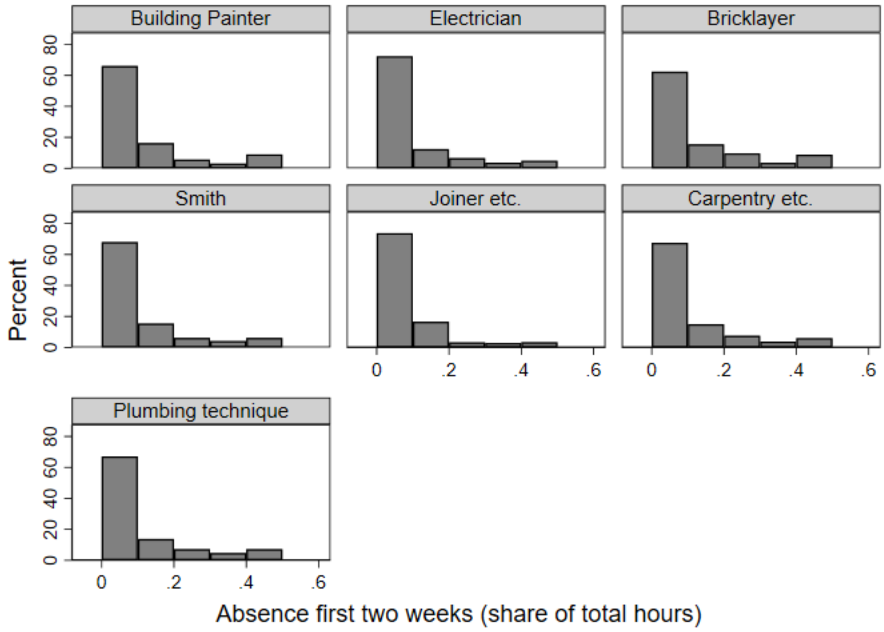


Note: The figure has the share of students on the second axis and the percent hours absent on the first axis. The first bare only includes students with zero absence.

We use the individual weekly variation in absences for the panel data instrument. Table 3 describes the average weekly variation, and Figure 4 shows the daily absences for all students and students with some absence during the day. In order to avoid selection issues, the sample consists of students that have scheduled class every day Monday through Friday during the first four weeks (20 weekdays). Figure 4 shows that the unconditional daily absence increases over the first 20 days, although at a decreasing rate. On the first day of school, around 4 percent of students are absent, and on day 20, this has increased to around 17 percent. The percent of daily hours absent for students with some absence during the day stays somewhat constant at around 65 percent of daily absence, meaning the average student who is absent is so for around 4 out of 6 hours during the day.

The patterns in daily absence are much the same across all educations as illustrated in Figure 5. However, when we look at the absence by whether the student has a prior enrollment in the first basic course or not, Figures 7(a) and 7(b) show that the older students starting directly on

Figure 2: Distribution of hours absent as a share of total hours for the first two weeks by education

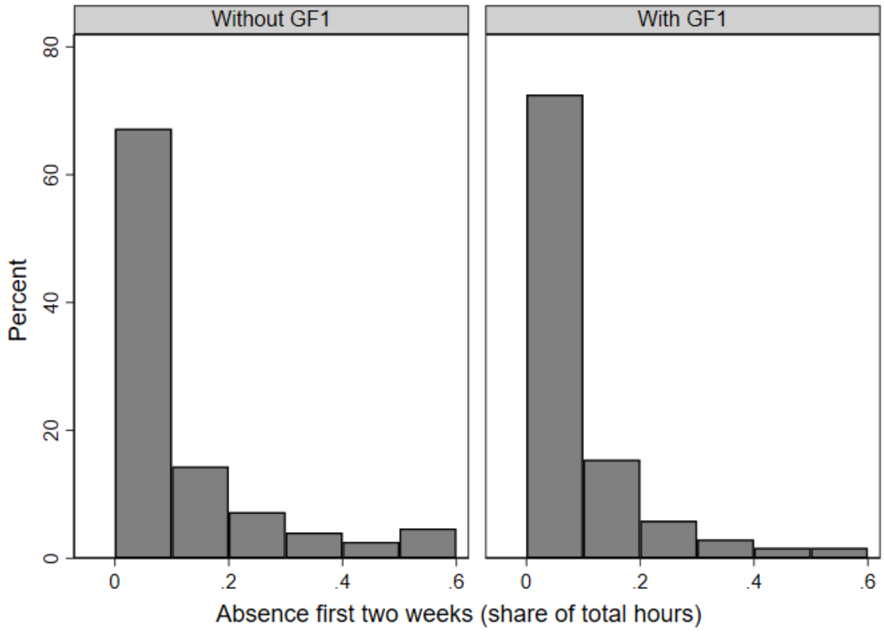


Note: The figure consists of 7 panels, one for each of the educations in our analysis sample. The panels all have the share of students on the second axis and the percent of hours absent on the first axis. The first bar in each panel only contains students with zero absence.

the second basic course grow to have a higher percent daily absences and more absence on the days where they have some absence. So on days the older students are absent, they tend to be absent for more hours than the younger students illustrating that just using an indicator for daily absence may not capture this type of heterogeneity in hours of daily absence. The last curiosity is that the percent of daily absence conditional on having some absence varies over the days. The daily absences are higher around days 1, 5, 10, 15, and 20. To the extent that students start on a Monday, this shows that students who have some absence during the day have more hours absent on Mondays and Fridays.

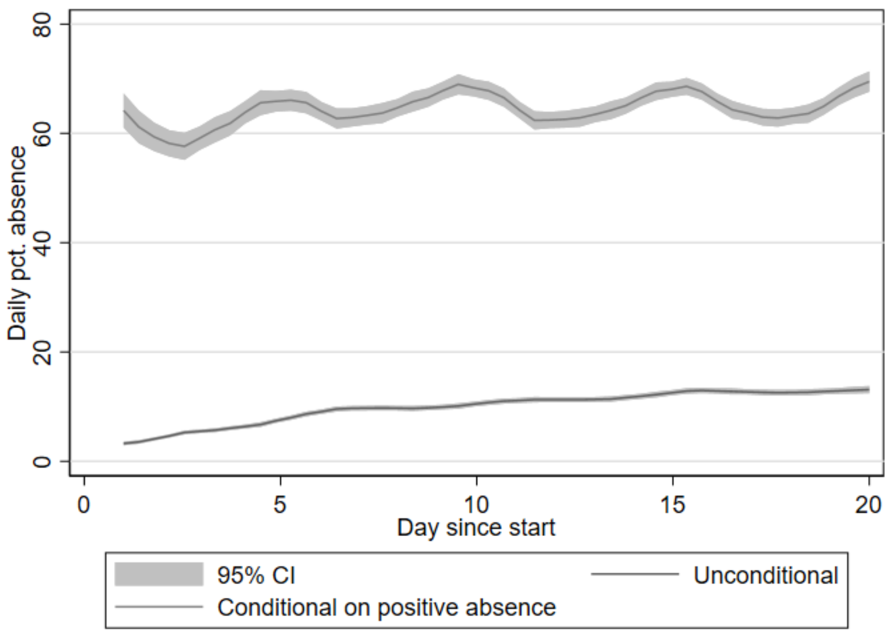
Finally, Figure 7 shows the relationship between the average absence during the first two weeks of school and the completion of the second basic course. We see that the relationship is very strong. Approximately 75 percent of the students with no or a very low level of absence complete the education, whereas the number is approximately 55 percent for those with an absence level of 20 percent (corresponding to two full days of absence during the first two weeks if the hours are equally distributed across weekdays). Part of this relationship can most likely be explained by the fact that students with high ability, effort, motivation, etc., are more likely to perform well in their courses and, in the end, complete the second basic course and are more likely to attend class. Therefore, we expect only part of the relationship to be a causal effect of absence on completion.

Figure 3: Distribution of hours absent as a share of total hours for the first two weeks by whether the students come from the first basic course or not



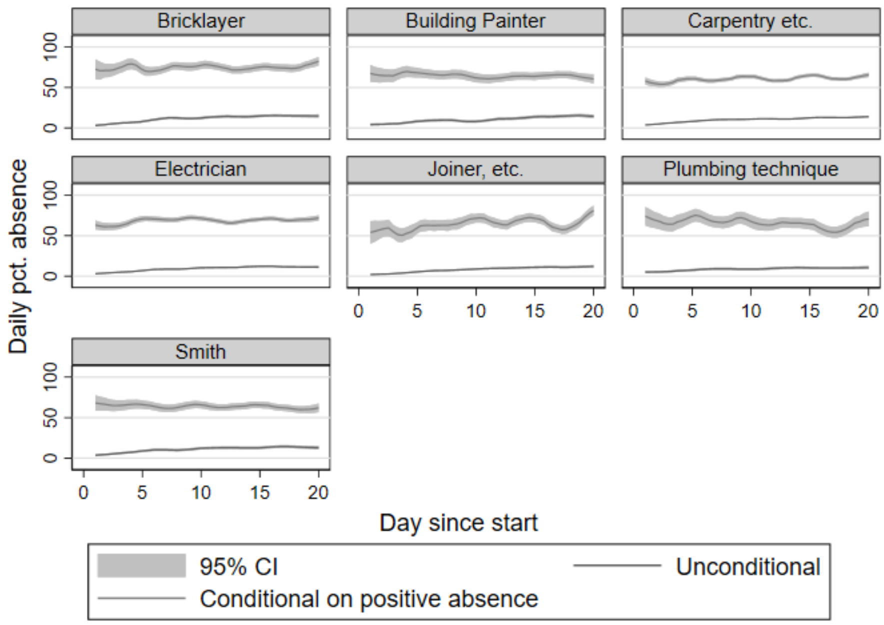
Note: The figure consists of two panels, the first is for students who start directly on the second basic course and the second is for students from the first basic course. The panels both have the share of students on the second axis and the percent of hours absent on the first axis. The first bar in each panel only contains students with zero absence.

Figure 4: Daily percent absence



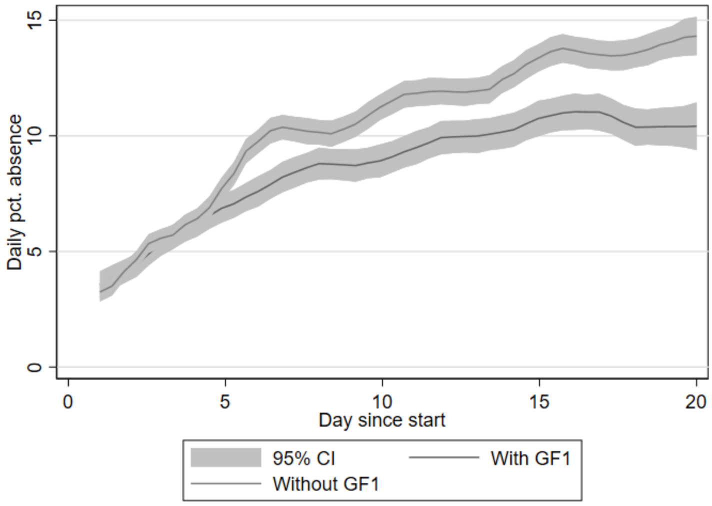
Note: The figure has the daily percent absence on the second axis and days since the matriculation date on the first axis.

Figure 5: Daily percent absence by education

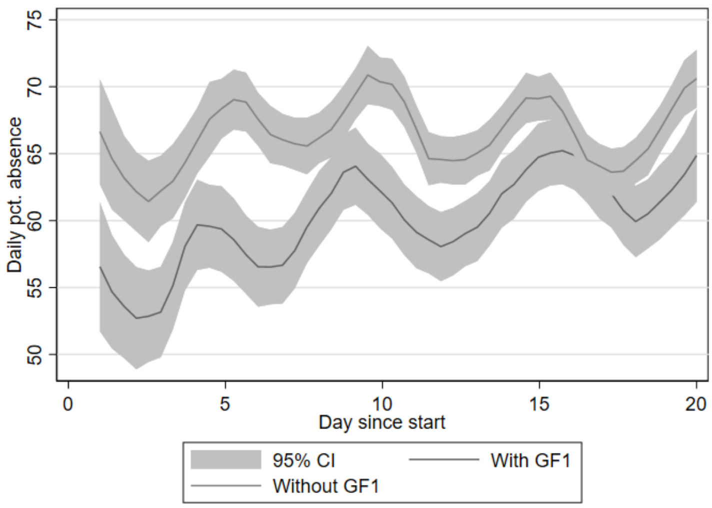


Note: The figure consists of 7 panels, one for each of the educations in our analysis sample. The panels have daily percent absence on the second axis and days since the matriculation date on the first axis.

Figure 6: Daily percentage of absence by prior GF1 or not



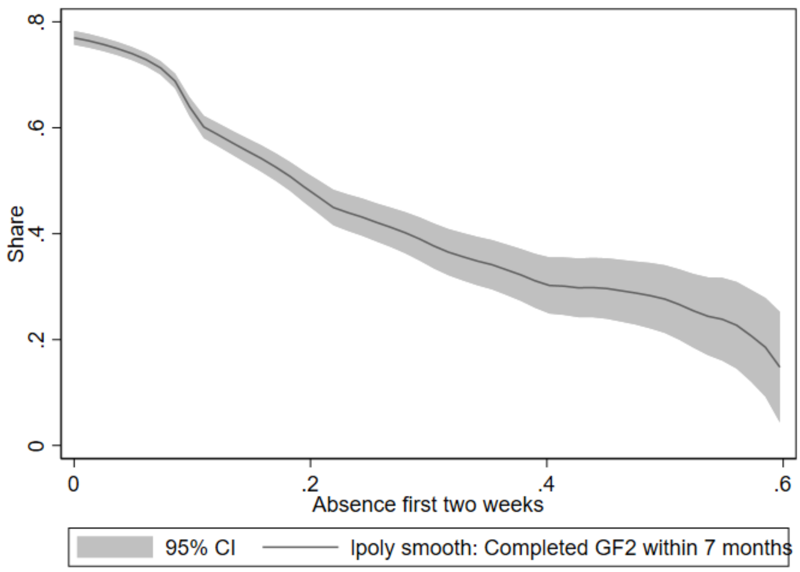
(a) Unconditional



(b) Conditional on positive daily absence

Note: The figure consists of two panels. Both panels have days since initial enrollment on the first axis and daily percent of absence on the second axis. Panel (a) displays the daily unconditional percent of absence for students with a prior first basic course (red line) and without a prior first basic course (blue line). Panel (b) displays the daily conditional percent absence for students with a prior first basic course (dark grey line) and without a prior first basic course (light gray line).

Figure 7: Relationship between total absence during the first two weeks and completion approximated by polynomial regression



Note: The figure has the share of students completed within seven months on the second axis and the percent of absence during the first two weeks on the first axis. The line is from the kernel weighted polynomial regression of absence on completion.

5 Empirical strategy

In this section, we describe our empirical strategy. According to human capital theory, being present in class and participating in class activities is an important determinant of human capital formation, and therefore, the causal effect of absence on educational performance is of interest. In the previous section, we saw a strong correlation between being absent from class within the first two weeks and the likelihood of completing the second basic course. We expect part of the correlation between absence and course completion to be confounded by underlying unobserved individual-specific factors such as student ability and motivation, as it is likely that high-ability and high-motivation students are more likely to come to class and to complete the course. Therefore, in order to identify the causal effect of absence from class on the likelihood of completing the second basic course, we consider two identification strategies:

a) Exogenous instruments based on meteorological measurements of weather conditions

b) A panel data instrument for absence

We explain the strategies in detail below.

5.1 IV methodology using weather

Our main identification strategy relies on instruments based on daily meteorological measurements of weather conditions at the vocational school locations. We use measures of precipitation and wind in the first two weeks of the course to match the period for which we measure absence. These precipitation and wind measures vary with year and month (January and August weeks) and with the location of the school as well as with the exact starting dates of the course (some start Monday in a given week, others start Wednesday, and some start the week after). In addition, we control for year, month, location, and education group fixed effects. Controlling for the combination of fixed effects rules out the case where weather variations affecting all students' absence in only one year or month (or location or education) are driving the results.

To identify the effects of absence on graduation, we make use of a standard IV setup, where we will use different specifications of weather during the first two weeks of the second basic course as instrumental variables; hence we will run a 2SLS regression model to estimate the effect of absence during the first two weeks of the second basic course on the probability of graduating with the following first- and second-stage equations

$$\bar{a}_i = \delta_0 + Z_i\delta_1 + X_i\delta_2 + \lambda_m + \lambda_y + \lambda_s + \lambda_e + v_i \tag{1}$$

$$y_i = \beta_0 + \beta_1\bar{a}_i + X_i\beta_2 + \gamma_m + \gamma_y + \gamma_s + \gamma_e + \epsilon_i, \tag{2}$$

where y_i is completion of the second basic course for individual i at the end of the course. The explanatory variable of interest is $\bar{a}_i$, which is the individual-level average absence within the first two weeks of the course. We are mainly interested in the parameter β_1, which is the marginal effect of absence on completion. X_i is a vector of exogenous covariates such as age, gender, immigrant status, parental education, and parental labor market attachment Finally, γ_m, γ_y, γ_s, and γ_e are respectively month, year, school, and education fixed effects.

Equation 2 is our equation of interest. We use this equation for the OLS regressions and the second stage with instrumented average absence. For the instrument, we use exogenous variation in the average weather, Z_i, during the first two weeks of school to predict average absence. More specifically, using weather during the first two weeks of school, we use the percentage of days with precipitation above 3 mm as our first instrument and the percentage of days with wind above 11 m/s for a 10 minute interval as our second instrument. We use the instruments separately and together, as well as interacted. We show that our result is robust to different definitions (cut-off values) of the weather variables. This is explained in more detail in section 6.2.

5.2 Panel data instrument

In addition to the main identification strategy in Section 5.1, we supplement our analysis with instruments based on the availability of repeated observations over time of individual absences (on a daily or weekly basis). The idea is that under specific assumptions, we can remove the individual-specific fixed effects from absence, and then under the assumption that there is no remaining correlation with the error term in the regression equation, this provides an instrument for a more aggregated measure of absence, e.g., the difference in weekly absence as an instrument for biweekly average absence. Altogether, the underlying assumptions are stronger than those required in the case with panel data observations of both the dependent and explanatory variables. This type of assumption is well-known from the dynamic panel data literature, where it is used to estimate the autoregressive parameter in a dynamic panel data model with individual-specific fixed effects. In that setting, the estimation is done by using lagged first-differences as instruments for the equation in levels. The assumption is known as mean-stationarity and introduced in Arellano and Bover (1995).

We consider the following regression model:

$$y_{iT} = \beta_0 + \beta_1 \bar{a}_i + X_i \beta_2 + \eta_i + u_{iT} \tag{3}$$

where y_i is completion of the second basic course for individual i at the end of the course after T time periods (20 weeks of school). Note that equation 3 is similar to equation 2 except for notation where we have omitted the fixed effects in the notation (they are included in the vector of exogenous covariates, X_i) and the error term is now split in two parts.

The regression error consists of the two terms η_i and u_{iT}. Note that we cannot distinguish the two error terms from each other in the cross-section setting. We assume that average absence $\bar{a}_i$ is independent of u_{iT} but can be correlated with η_i. The first error term η_i captures individual-specific fixed effects such as student ability and the part of student motivation that is constant over time. The term η_i affects both absence and the completion of the second basic course. The second error term u_{iT} captures all other unobserved parts of completion that, by assumption, are independent of average individual absence within the first two weeks. The term u_{iT} can contain both individual-specific time-constant effects and individual time-varying shocks that happen to the learning process during the 20 weeks of the course (for example, teacher quality and things going on in and outside the class). Part of the identifying assumption is that there is no reverse effect from learning process shocks to absence, i.e., $\bar{a}_i$ is independent of u_{iT}, and this is also the reason for considering absence during the first two weeks of class. We conjecture that after some time in class, we are more likely to find reverse causality in the sense that shocks to the teaching/learning process that impact the gains from the teaching in the class also has an impact on the likelihood of coming to class in later weeks. We find it plausible that this reverse effect is not present within the first two weeks of class, where students are finding out what the course is about and attendance is possibly less influenced by realizations of gains from attending class. Another part of the identifying assumption is that the effect of absence in weeks one and two is the same such that there is only one endogenous variable in the regression equation of interest.

Letting a_{it} be absence within a week, t, we assume

$$a_{it} = \widetilde{a}_{it} + \alpha_i \tag{4}$$

where $\widetilde{a}_{it}$ and α_i are independent of each other with mean zero and $\widetilde{a}_{it}$ is independent of the regression errors η_i and u_{iT}. The individual-specific effects α_i and η_i can depend on each other, and the dependency between absence and the regression error happens through this term. Altogether this implies that the first-differences $\Delta a_{i2} = a_{i2} - a_{i1} = \tilde{a}_{i2} - \tilde{a}_{i1}$ are independent of η_i such that $E\left[\Delta a_{i2}(\eta_i + u_{iT})\right] = 0$. In addition we have that $E\left[(a_{i1} + a_{i2})(a_{i2} - a_{i1})\right] = E(\widetilde{a}_{i2}^2) - E(\widetilde{a}_{i1}^2)$ such that Δa_{i2} is a valid instrument for $(a_{i1} + a_{i2})$ if the second order moments of $\widetilde{a}_{it}$ are not constant over time. We can test whether this holds in the first-stage regression. In addition, any function of Δa_{i2} can be used as instrument for $(a_{i1} + a_{i2})$, for example Δa_{i2}^2 (this requires assumptions on third order moments) or $|\Delta a_{i2}|$. Note that $\tilde{a}_{i1}$ and $\tilde{a}_{i2}$ can be dependent over time. Altogether the assumption on absence a_{it} means that a_{it} must be additive in the individual-specific effect α_i such that we can remove the term by subtracting the individual-specific mean. As mentioned above, this assumption is referred to as mean-stationarity in the panel data literature, see Arellano and Bover (1995). The assumption rules out that low- and high-ability students can have different trends in absence over time, and therefore it cannot be the case that low-ability students increase absence over time more than high-ability students but they can have permanent high level of absence. In the situation with more than two observations over time of absence that are exogenous except for the α_i-part, the instruments will be the individual-mean corrected absences (or first differences of absences) at the different periods. The approach requires that absence is a continuous variable such that it can be expressed in a linear additive form as in equation 4. We discuss this in more details in the next section.

6 Results

This section presents our results regarding the effect of absence during the first two weeks of school on the probability of completing the second basic course. We first present the baseline OLS estimates of equation 2. Second, we present the results from using our chosen instruments based on meteorological observations, presented in section 5.1, and include first-stage estimates as well. In all the estimations, we use the average percent of hours absent during the first two weeks of school as our variable of interest and an indicator of whether the student has completed the second basic course within 7 months after graduation as our outcome variable. We also show the results are robust to using percent days with some absence rather than percent hours absent during the first two weeks.

Third, we support our results from the main specifications with the panel data instrument from section 5.2. We see that the results are qualitatively similar to our main results, using meteorological weather observations as instruments. We include the first-stage estimates and robustness checks where we condition the sample on students with some absence.

6.1 OLS results

Table 4 shows the OLS results from estimating the effect of absence within the first two weeks on completion from equation 2. The table presents results for three different groups where panel A includes the full sample, panel B only includes students *with* a prior spell in the first basic course and panel C only includes students *without* a prior spell in the first basic course. The table's columns show how the coefficient on absence changes as we add controls and fixed effects sequentially. The relationship between absence during the first two weeks and completion probability is highly significant and stable around a coefficient of approx -1 for all three groups. In column (7), we see that including all the controls and month, year, education group, and school fixed effects, the OLS estimates for all students indicate that a 10 percentage point increase in hours absent (corresponding to one full day of absence) during the first two weeks is associated with a 9.1 percentage point lower probability of completing within seven months.

Table 4: The Effect of pct. Hours Absence during the first two weeks on Graduation: OLS results

	Graduated	Graduated	Graduated	Graduated	Graduated	Graduated	Graduated
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: All							
Pct. absence	-1.039***	-0.916***	-0.906***	-1.037***	-1.032***	-1.030***	-0.909***
	(0.046)	(0.047)	(0.048)	(0.047)	(0.048)	(0.049)	(0.047)
Observations	5782	5782	5782	5782	5782	5782	5782
R-squared	0.081	0.144	0.154	0.081	0.095	0.096	0.174
Panel B: With GF1							
Pct. absence	-1.030***	-0.902***	-0.908***	-1.033***	-1.061***	-1.065***	-0.948***
	(0.132)	(0.119)	(0.119)	(0.132)	(0.134)	(0.135)	(0.119)
Observations	1833	1833	1833	1833	1833	1833	1833
R-squared	0.062	0.133	0.146	0.063	0.087	0.088	0.183
Panel C: Without GF1							
Pct. absence	-1.039***	-0.909***	-0.898***	-1.036***	-1.001***	-0.997***	-0.883***
	(0.055)	(0.061)	(0.061)	(0.056)	(0.055)	(0.056)	(0.061)
Observations	3949	3949	3949	3949	3949	3949	3949
R-squared	0.088	0.159	0.169	0.091	0.104	0.107	0.189
Student controls	No	Yes	Yes	No	No	No	Yes
Parent controls	No	No	Yes	No	No	No	Yes
Year/Month FE	No	No	No	Yes	No	Yes	Yes
School/Education FE	No	No	No	No	Yes	Yes	Yes

Notes: The table shows the raw OLS estimates for pct. absence during the first two weeks on graduation (0/1). The bottom panel indicates if student controls (age, immigration status, log distance, male, 9th grade danish score, 9th grade math score, and labour market attachment prior to enrolment), parent controls (highest completed parental education, mothers and fathers labour market attachment), year and month fixed effects, and school and education fixed effects are included in the model in the respective column. Panel A uses the full sample, Panel B only uses students with a prior GF1, and Panel C only uses students with no prior GF1. Standard errors clustered on the school by start date level are reported in parentheses. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

In appendix table 12 we present the raw OLS results for percent days with some absence instead of percent hours of absence. The estimates are similar in magnitude and size as the ones for percent hours of absence in table 4 and they remain stable as we include more control variables.

The censoring of absence at zero is a concern for our panel data instrument because it introduces a non-linearity in absences. Therefore, we also present all results conditional on students with some absence in at least two weeks, leaving us with approximately 18 percent of the sample. Table 5 presents the OLS estimates for absence within the first two weeks as explanatory variable when we condition on only including students with some absence in both weeks. In table 5, we see that the association between completion and absence within the first two weeks is lower when we condition on some absence than the unconditional absence in table 4. As we saw in table 4, the point estimate in table 5 is only slightly smaller (towards zero) when we include control variables.

Table 5: The Effect of pct. Hours Absence during the first two weeks on Graduation, conditional on two weeks positive absence: OLS results

	Graduated	Graduated	Graduated	Graduated	Graduated	Graduated	Graduated
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: All							
Pct. absence	-0.642***	-0.635***	-0.605***	-0.629***	-0.624***	-0.614***	-0.587***
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Observations	1021	1021	1021	1021	1021	1021	1021
R-squared	0.050	0.160	0.192	0.055	0.069	0.073	0.224
Panel B: With GF1							
Pct. absence	-0.761***	-0.829***	-0.800***	-0.763***	-0.812***	-0.832***	-0.873***
	(0.19)	(0.18)	(0.19)	(0.19)	(0.19)	(0.20)	(0.19)
Observations	341	341	341	341	341	341	341
R-squared	0.047	0.141	0.166	0.051	0.081	0.087	0.234
Panel C: without GF1							
Pct. absence	-0.577***	-0.571***	-0.547***	-0.578***	-0.530***	-0.531***	-0.515***
	(0.09)	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)	(0.08)
Observations	680	680	680	680	680	680	680
R-squared	0.045	0.189	0.229	0.053	0.073	0.078	0.268
Student controls	No	Yes	Yes	No	No	No	Yes
Parent controls	No	No	Yes	No	No	No	Yes
Year/Month FE	No	No	No	Yes	No	Yes	Yes
School/Education FE	No	No	No	No	Yes	Yes	Yes

Notes: The table shows the raw OLS estimates for pct. absence during the first two weeks, conditional on some absence in both weeks, on graduation (0/1). The bottom panel indicates if student controls (age, immigration status, log distance, male, 9th grade danish score, 9th grade math score, and labour market attachment prior to enrolment), parent controls (highest completed parental education, mothers and fathers labour market attachment), year and month fixed effects, and school and education fixed effects are included in the models in the respective column. Panel A uses the full sample, Panel B only uses students with a prior GF1, and Panel C only uses students with no proir GF1. Robust standard errors are in parentheses. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

The fact that all the estimates are robust to the inclusion of the different control variables suggests that the students only, to a small extent, select into being absent based on observed characteristics. However, the OLS effect of absence on completion might still suffer from potential bias due to unobserved omitted factors, in particular unobserved individual heterogeneity that is constant over time, which is why we continue with our strategy for the panel data instrument.

6.2 Weather Instrument Results

6.2.1 First-stage Results, Weather Instrument

Table 6 shows the first-stage estimates for the different weather instruments in column (1) percent days with over 3 mm precipitation during the first two weeks, in column (2) percent days with average wind speeds above 11 ms for 10 minutes, in column (3) the two instruments jointly, and in column (4) the two instruments jointly along with their interaction. Panel A reports the first-stage estimates for the full analysis sample. We see that wind is separately significant on the 5% significance level, while precipitation separately and their interaction is significant on the 10% level for the full sample. We see the same pattern, although with larger and more significant coefficients and F-values, in panel C, which contains the estimates for the students who did not attend the first basic course before their enrollment in the second basic course. In panel B, which reports estimates for the students who attended the first basic course before their enrollment in the second basic course, we see insignificant coefficients for all columns. Further, all the columns have very low F-values indicating that our weather measures are not suitable candidates as instruments for the students who attended the first basic course prior to their enrollment in the second basic course. The fact that none of the coefficients are significant leads us to conclude that the weather instruments are only appropriate for the students without a prior enrollment in the first basic course, and we, therefore, primarily focus on this group in our main analyses.[10] The first-stage results for panel C all point in the expected direction, namely that students are more absent over the first two weeks when there are more days with precipitation or high wind speeds.

To show that these first stage estimates are not only occurring for our particular values of precipitation and wind, tables 14 and 16 respectively show the first-stage estimates for a range of precipitation and wind specifications for the students without a prior GF1 spell. Table 14 shows that for any cutoff of precipitation, the first stage estimates are significant at the 1 percent level however, the F-values for the first stage are higher for smaller values of the precipitation cutoff. The F-value is highest when we count days with above 1 mm of precipitation. For days with wind speed above a certain cutoff, table 16 also shows significant first stage estimates for a range of different wind speeds and with F-values from 10 to 19 for wind speeds from 10 ms to 13 ms.

To further explore the variation in precipitation and wind, figures 9 and 10 show the share of students exposed to the different cutoff values. Figure 9 shows that, respectively, 65 and 80 percent of students without and with GF1 experience at least one day with precipitation above 1 mm. For our chosen value of 3 mm these numbers are 70 and 50 percent of students without and with GF1. The share of students experiencing high winds are somewhat higher with around 97 and 98 percent of students without and with GF1 experiencing at least one day with wind above 11 ms.

[10] A potential reason why the older students react more to weather conditions is that they have been out of school for a number of years and many of them have also not held a job prior to enrolling in vocational school. Because they are less used to getting out the door in the morning, the marginal student may be more likely to be affected by the weather relative to the younger students who have not tried anything but attending school every day. This may also be true for the students who live without their parents relative to students not yet having moved away from their parents. We will analyze these heterogeneous effects in Section 6.2.4.

Table 6: First-stage estimates for the different specifications of weather instruments, percent hours absent used as outcome (2 weeks)

	Pct. hours absent	Pct. hours absent	Pct. hours absent	Pct. hours absent
	(1)	(2)	(3)	(4)
Panel A: All				
Pct. days with precipitation > 3mm	0.032		0.022	0.004
	(0.020)		(0.022)	(0.019)
Pct. days with wind>11ms		0.039**	0.034**	0.029*
		(0.015)	(0.017)	(0.017)
Precipitation X wind				0.076
				(0.057)
Observations	5782	5782	5782	5782
F	2.42	6.51	3.82	3.06
Panel B: With GF1				
Pct. days with precipitation > 3mm	-0.035		-0.033	-0.033
	(0.029)		(0.028)	(0.037)
Pct. days with wind>11ms		0.023	0.018	0.018
		(0.028)	(0.026)	(0.036)
Precipitation X wind				0.000
				(0.092)
Observations	1833	1833	1833	1833
F	1.49	0.70	0.88	0.68
Panel C: Without GF1				
Pct. days with precipitation>3mm	0.082***		0.058**	0.019
	(0.026)		(0.027)	(0.023)
Pct. days with wind > 11ms		0.09***	0.073***	0.063***
		(0.020)	(0.021)	(0.021)
Precipitation X wind				0.172*
				(0.090)
Observations	3949	3949	3949	3949
r2				
F	9.76	19.51	12.59	11.27

Notes: The table shows the first-stage estimates for the different specifications of the weather instruments on our endogenous measure of interest, pct. hourss absent during the first two weeks. All models contain additional controls (age, immigration status, log distance, and male), year and month fixed effects, and school and education fixed effects. Panel A uses the full sample, Panel B only uses students with a prior GF1, and Panel C only uses students with no prior GF1. F-tests for joint significance are reported for each panel. Standard errors clustered on the school by start date level are reported in parentheses. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

6.2.2 Main Results: Weather Instruments

Table 7 contains our main results, the IV estimates of the effect of being absent during the first two weeks on the probability of graduating based on our second-stage specification in equation 2 with the weather instruments. As mentioned in section 6.2.1, we only report results for students without a prior spell in the first basic course, as the first-stage check showed that our weather instruments are not suitable for the group with a prior spell in the first basic course. Therefore, we display the second-stage estimates for the joint analysis sample and students with a prior spell in the first basic course in appendix table 13.

Column (1) of table 7 reports the second-stage estimate for precipitation as an instrument for percent hours absent. We see that students are estimated to be 1.95 percentage points less likely to graduate when they are 1 percentage point more absent in hours; This roughly corresponds to a 20 percentage point decrease in the probability of graduating when they have one additional day of absence during the first two weeks of courses.[11] The pattern for columns (2)-(4) are similar, although the estimated coefficients are slightly smaller in absolute size, with 1.06 – 1.36 percentage points. The coefficient in column (2), with our measure of wind as an instrument, is not significant, although when coupled or interacted with precipitation in columns (3) and (4) the coefficient on absence is significant. Further, the first-stage F-values range from 9.76 – 19.51, and we can reject the null on the 1% significance level for the Anderson Rubin wald test for column (1) and at the 5% significance level for column (3) and (4) .

We can again explore how the estimates vary over the precipitation and wind cutoff values. In tables 15 and 17 we show the second-stage estimates for different precipitation and wind specifications and in table 18 we show a matrix of second-stage estimates for the different joint precipitation and wind specifications. Table 15 show that the second-stage estimates are significant for counting days with 1, 3, 4 and 5 mm precipitation and that our chosen specification of 3 mm has the highest significance level at 5 percent. The coefficients that are significant at the 10 percent level vary 1.69 to 2.19, such that our chosen specification with a 2SLS estimate of 1.95 is somewhere in the middle. For different specifications of wind, table 17 shows that other cutoffs than 11 ms, indeed have significant second-stage estimates, such that future research should perhaps focus on using a cutoff of 12 or 13 ms rather than the 11 that we chose in the main specification. Table 18 shows the ranges of second-stage estimates when both precipitation and wind are included. Also here we see that we can reject the null at a lower significance level for the Anderson Rubin wald test if we use a cutoff of 12 ms or 13 ms for 3 mm of precipitation. If we use days with 13 ms wind and days with 3 mm precipitation, the 2SLS coefficient is 1.69 and is significant at the 1 percent level with an F-value of 13.26.

6.2.3 Robustness Results: Weather Instrument

Our results are robust to the definition of absence. We run the same estimations with percent days with some absence rather than percent of hours absent. The idea being that days with precipitation and wind will affect either being late for school or being absent the entire day. Categorizing percent days with any absence during the first two weeks of school is a variable

[11] To roughly translate the interpretation of the coefficient from a 1 percentage point increase in absence to one more day of absence, we can multiply the estimated coefficients with 10, which table 4 shows is the average number of school days during the first two weeks.

Table 7: Effect of absence on probability of graduation: 2SLS with weather as IV for absence (2 weeks)

	Graduated (1)	Graduated (2)	Graduated (3)	Graduated (4)
Pct. hours absent	-1.952**	-1.059	-1.404**	-1.357**
	(0.928)	(0.817)	(0.636)	(0.611)
Observations	3949	3949	3949	3949
First-stage F	9.76	19.51	12.59	11.27
AR p-value	0.004	0.205	0.018	0.045
Precipitation > $3mm$	Yes	No	Yes	Yes
Wind > $11ms$	No	Yes	Yes	Yes
Precipitation×Wind	No	No	No	Yes

The table reports second-stage estimates. The dependent variable is a dummy indicating if the student has graduated within 7 months (0/1). The bottom panel contains indicators for which instruments have been used in the first-stage estimation. The instruments used are pct. of days with over 3 mm of precipitation within the first two weeks, pct. days with average windspeed above 11 ms for a 10 min. interval, and the interaction of these two instruments. All models further include the log distance from students parish to the schools parish, age dummies, a dummy for males, dummies for first and second generation immigrants, and year, month, school, and education fixed effects. Standard errors clustered on the school by start date level are reported in parentheses and the reported F-values are from the first-stage. AR p-values are from the Anderson-Rubin wald test. Significance levels are reported as *** p<0.01, ** p<0.05, * p<0.1.

from zero to 100, where being absent half an hour or the full day, will give the students 10 percentage point (1 day out of ten) more absence. This is the same way we define the instruments. Table 19 shows the first-stage estimates for students without GF1 when we use percent days with positive absence. We see that both days with precipitation and days with wind are significant at the 5 percent level although with lower F-statistics than when we use percent hours absence. Using both instruments does not give any additional significance. Together the results show that days with precipitation and wind better predict variation in percent hours of absence rather than days with any absence. This could be explained, for example, if students have a tendency to miss the first class on all days, independent of the weather but that bad weather makes the students miss the entire day. This would give a higher percentage absence on the days with bad weather and would therefore make the weather variable have higher correlation with percent hours of absence rather than days with some absence. Using absence in percent days, table 20 shows the second-stage estimates of students without a prior GF1 enrollment. The estimates are significant at the 10 percent level and have slightly higher point estimates than when we use percent hours absent. This is such that 10 percentage point more absence (e.g. 1 day more absent during the first 10 days of school) translate into 20 percentage point lower probability of completing the GF2 course.

6.2.4 Heterogeneous effects: Weather Instrument

The fact that the weather instrument only affects percent hours absent for the students without a prior GF1 enrollment can be due to many different reasons. As we saw in section 4 the two groups with and without a prior GF1 are different on many observable characteristics. Two of these characteristics are the fact that only 42.1% percent of the students without a GF1 do not live with their parents (compared to 95.5% of students with a GF1) and that one quarter have neither worked nor been enrolled in school the year prior to enrollment. In this section, we show how the results vary across these two characteristics for the students without a GF1. The idea being that we want to check if weather affects absence more for students who do not live with their parents (who can encourage them to go to school even if it rains) and students who are not used to having a time to start the day because they have not been in school or at work the year prior to enrollment. Besides differing between students with and without GF1, our hypothesis is that these two characteristics are also likely to correlate with students who are more likely to be affected by weather conditions when they decide to go to school in the morning. We will test this below. Finally, we show heterogeneous results across math grades from 9th grade to test if weather is a stronger instrument for students with low or high grades.

The results on the heterogeneous effects are presented in tables 21, 22, and 23. The first row in each table presents the second stage estimates, the second row the OLS estimates and row three and four the first stage estimates for the two instruments. When we divide students by living with or without parents, in table 21, we see that the OLS estimates do not differ much across the groups. However, for the precipitation instrument, the results are driven by students not living with their parents. For this group the first-stage coefficients are much larger than for the group of students living with their parents. Further, the F-statistic is 24 for the precipitation instrument, which is also much larger than for the group of students living with their parents who only have an F-statistic of 1. Together, the results from table 21 show that our hypothesis holds for students not living with their parents as weather is a stronger predictor for their absence, which affects completion.

For students with and without a prior stable attachment (prior enrollment in school or employed), shown in table 22, the OLS estimates are larger for the group of students without a prior stable attachment indicating that the correlation between absence and completion is stronger for the group of students with no stable attachment. The first stage estimates show that weather predicts absence for both groups but the correlation is stronger for the group of students with no stable attachment. In this sense our hypothesis seems to also hold for students without a prior attachment, although there is no significant effect on completion for students without a prior attachment, most likely due to the smaller sample size. The first-stage estimates translate into significant second-stage estimates for the group with a stable prior stable attachment, although with higher point estimates. We consider this as work in progress, and will look further into this in the future.

Finally, when dividing the sample by high and low math grades from 9th grade, table 23 shows that the OLS estimates are almost the same for high and low grade students. In the first-stage we see that the weather is more correlated with absence for the low grade students, but the second stage estimates are insignificant. For the high grade students, the second-stage estimates are significant and around twice as large in magnitude as our main results, but the F-statistics of these regressions are low and we therefore do not want to interpret further on these estimates.

6.3 Panel Data Instrument Results

6.3.1 First-stage Results, Panel Data Instrument

To support our main analysis using different weather measures as an instrument, we implement the same analysis using our panel data instruments.

The panel data instrument is the first difference in average absence between two weeks and functions of the first difference that are the difference squared, the first difference and the difference squared together, and the absolute difference.

Table 8 shows the first-stage estimates from the four different specifications of our instrument for average absence during the first two weeks. The table reports the results for the full sample in panel A, students with a prior first basic course in panel B, and students without a prior first basic course in panel C. All of the instrument specifications correlate significantly with average absence, and if we focus on panel A, the reported F-tests are all above 100. However, a significant driver of the high F-values is that more than 80 pct. of the observations have at least one week out of two with zero absence resulting in a perfect correlation between average absence and the absolute value of the first difference for these observations. The model does not take the censoring of absence into account.

Therefore, we also perform the first stage estimation on the sample conditional on some absence. In this regression, we include students with absences in the first and second weeks of school. The results are presented in table 9 and show similar results, with all first-stage estimates significant and relatively high F-values. As before, we know that the high F-values, especially for the instrument using the absolute value of the first difference, are partly driven by many observations, with absence in either week one or week two being close to zero. Further, although the conditioning on positive values of absence may eliminate some of the bias from using a censored variable, it introduces a negative correlation between $\widetilde{a}_{it}$ and η_i in equation 4 such that the two are no longer independent and the instrument is therefore not valid. Even though this is the case, we proceed with a description of the results and afterwards briefly discuss a potential avenue of future research to the setup of the panel data instrument.

Table 8: First-stage estimates for the different specifications of the panel data instrument, percent weekly hours absence for the first two weeks used as outcome

	Pct. hours absent (1)	Pct. hours absent (2)	Pct. hours absent (3)	Pct. hours absent (4)
Panel A: All				
Diff. pct. absence	0.232***		0.041***	
	(0.015)		(0.010)	
Diff. pct. absence sq.		0.766***	0.731***	
		(0.028)	(0.028)	
Abs. diff. pct. absence				0.577***
				(0.011)
Observations	5782	5782	5782	5782
F	255.63	740.30	374.71	2836.75
Panel B: With GF1				
Diff. pct. absence	0.201***		0.042**	
	(0.033)		(0.019)	
Diff. pct. absence sq.		0.755***	0.723***	
		(0.041)	(0.040)	
Abs. diff. pct. absence				0.564***
				(0.017)
Observations	1833	1833	1833	1833
F	37.44	337.72	175.97	1089.87
Panel C: Without GF1				
Diff. pct. absence	0.242***		0.041***	
	(0.016)		(0.012)	
Diff. pct. absence sq.		0.764***	0.728***	
		(0.034)	(0.034)	
Abs. diff. pct. absence				0.580***
				(0.013)
Observations	3949	3949	3949	3949
F	227.40	514.78	265.52	2076.57

Notes: The table shows the first-stage estimates for the different specifications of the panel instrument on our endogenous measure of interest, pct. hours absent during the first two weeks. All models contain additional controls (age, immigration status, log distance, and male), year and month fixed effects, and school and education fixed effects. Panel A uses the full sample, Panel B only uses students with a prior GF1, and Panel C only uses students with no proir GF1. F-tests for joint significance are reported for each panel. Standard errors clustered on the school by start date level are reported in parentheses. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

Table 9: First-stage estimates for the different specifications of the panel data instrument, percent weekly hours absence for the first two weeks used as outcome. Sample condtioned on two weeks positive absence

	Pct. hours absent	Pct. hours absent	Pct. hours absent	Pct. hours absent
	(1)	(2)	(3)	(4)
Panel A: All				
Diff. pct. absence	0.147***		0.0486***	
	(0.0257)		(0.0173)	
Diff. pct. absence sq.		0.613***	0.579***	
		(0.0379)	(0.0367)	
Abs. diff. pct. absence				0.475***
				(0.0240)
Observations	1021	1021	1021	1021
F	32.86	261.4	130.4	391.1
Panel B: GF1				
Diff. pct. absence	0.210***		0.0643**	
	(0.0495)		(0.0316)	
Diff. pct. absence sq.		0.758***	0.697***	
		(0.0791)	(0.0797)	
Abs. diff. pct. absence				0.537***
				(0.0438)
Observations	341	341	341	341
F	18.03	91.76	47.79	150.6
Panel C: without GF1				
Diff. pct. absence	0.117***		0.0354*	
	(0.0295)		(0.0209)	
Diff. pct. absence sq.		0.560***	0.538***	
		(0.0440)	(0.0427)	
Abs. diff. pct. absence				0.444***
				(0.0308)
Observations	680	680	680	680
F	15.62	162.0	81.19	207.9

Notes: The table shows the first-stage estimates for the different specifications of the panel instrument on our endogenous measure of interest, pct. hours absent during the first two weeks, conditional on some absence in both weeks. All models contain additional controls (age, immigration status, log distance, and male), year and month fixed effects, and school and education fixed effects. Panel A uses the full sample, Panel B only uses students with a prior GF1, and Panel C only uses students with no proir GF1. F-tests for joint significance are reported for each panel. Standard errors clustered on the school by start date level are reported in parentheses. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

6.3.2 Results: Panel instrument

As mentioned, to support our main results based on the weather measurement instruments, we present the second stage of the IV-estimation, using the panel data instrument for the first two weeks of absence, unconditionally and conditionally on positive absence during both of the two weeks. The results come with the caveat from section 6.3.1 that the high F-values are to a large extent driven by a high number of observations with zero absence in either of the the two weeks. Table 10 contains the second stage estimates of the effect of being absent during the first two weeks on the probability of completion based on the four functional forms of $\Delta a_{i2} = \tilde{a}_{i2} - \tilde{a}_{i1}$ from equation 4. All columns have completion within seven months as the outcome variable and contain our endogenous variable of interest, absence, measured as percent hours of absence during the first two weeks. We include all controls and month, year, education group, and school fixed effects in all estimations. The bottom panel indicates, for each column, which specification of the first difference in percent hours absent during the first two weeks has been used as the instrument in the associated first-stage regression presented in section 6.3.1. We further report first-stage F test values as a measure of the strength of our instruments and p-values for the Anderson-Rubin Wald test, which is robust to weak instruments (Anderson and Rubin, 1949).

In panel A of table 10, which reports results for the full sample, we find that an increase of a 10 percentage point in hours of absence results in a decrease of $7 - 10.07$ percentage points in the probability of completing, and this roughly corresponds to being absent for one additional day during the first two weeks of school, decreases the probability of completing with $7 - 10.07$ percentage points. These second-stage estimates are close to the OLS estimate of -1.03, including all control variables and fixed effects; this is most likely caused by the high mechanical correlation between average absence and the first differences in absence when absence in one week is zero. Moving on, we see that panel B, conditional on students with a prior spell in the first basic course, and panel C, conditional on students with no prior spell in the first basic course, show similar results. For panel B, we find that an additional day of absence reduces the probability of completing with $5 - 10$ percentage points, and for panel C, we find that an additional day of absence reduces the probability of completing with $7.6 - 10$ percentage points. Again all estimates in panels B and C are highly significant, with high F values and low Anderson-Rubin p-values.

We also perform the second-stage regressions for the absence during the first two weeks, conditional on some absence in both weeks. We present the results in Table 11. Unfortunately, the large drop in the number of observations results in a large loss of power, such that none of the estimated coefficients are significant, and their magnitude is much smaller.

7 Discussion

As mentioned above, the panel data instrument suffers from being censored in week 1 and week 2 and therefore gives improbable high correlation between the average absence during week 1 and 2 and the first difference in absence from the two weeks. We try to abate this issue by only analyzing first difference in absence for students with positive absence in both weeks however, this instead introduces a negative correlation between $\widetilde{a}_{it}$ and η_i, such that the two

Table 10: Effect of absence during the first two weeks on probability of graduation: 2SLS with the panel instrument as IV for absence

	Graduated (1)	Graduated (2)	Graduated (3)	Graduated (4)
Panel C: All				
Pct. hours absent	-1.068***	-0.701***	-0.720***	-0.861***
	(0.172)	(0.071)	(0.073)	(0.059)
Observations	5782	5782	5782	5782
First-stage F	255.63	740.30	374.71	2836.75
AR p-value	0.000	0.000	0.000	0.000
Panel B: With GF1				
Pct. hours absent	-1.022***	-0.495***	-0.521***	-0.746***
	(0.311)	(0.161)	(0.162)	(0.132)
Observations	1833	1833	1833	1833
First-stage F	37.44	337.72	175.97	1089.87
AR p-value	0.007	0.008	0.005	0.000
Panel A: Without GF1				
Pct. hours absent	-1.035***	-0.757***	-0.772***	-0.889***
	(0.186)	(0.088)	(0.090)	(0.080)
Observations	3949	3949	3949	3949
First-stage F	227.40	514.78	265.52	2076.57
AR p-value	0.000	0.000	0.000	0.000
$\Delta absence$	Yes	No	Yes	No
$(\Delta absence)^2$	No	Yes	Yes	No
$\lvert\Delta absence\rvert$	No	No	No	Yes

Notes: The dependent variable is a dummy indicating if the student has graduated within 7 months (0/1). The bottom panel contains indicators for which instruments have been used in the first-stage estimation. The instruments used are the first difference in weekly absence, the first difference in weekly absence squared, and the absolute difference in weekly absence. All models further include the log distance from students parish to the schools parish, age dummies, a dummy for males, dummies for first and second generation immigrants, and year, month, school, and education fixed effects. Standard errors clustered on the school by start date level are reported in parentheses and the reported F-values are from the first-stage. AR p-values are from the Anderson-Rubin wald test. Significance levels are reported as *** p<0.01, ** p<0.05, * p<0.1.

Table 11: Effect of absence during the first two weeks on probability of graduation. Sample condtioned on two weeks positive absence: 2SLS with the panel instrument as IV for absence

	Graduated (1)	Graduated (2)	Graduated (3)	Graduated (4)
Panel A: All				
Pct. hours absent	-0.400	-0.174	-0.192	-0.310*
	(0.370)	(0.174)	(0.172)	(0.160)
Observations	1021	1021	1021	1021
First-stage F	32.86	261.4	130.4	391.1
AR p-value	0.298	0.337	0.475	0.0631
Panel B: GF1				
Pct. hours absent	-0.711	-0.262	-0.308	-0.409
	(0.565)	(0.325)	(0.321)	(0.310)
Observations	341	341	341	341
First-stage F	18.03	91.76	47.79	150.6
AR p-value	0.237	0.466	0.489	0.226
Panel C: Without GF1				
Pct. hours absent	-0.274	-0.177	-0.182	-0.321*
	(0.530)	(0.210)	(0.209)	(0.193)
Observations	680	680	680	680
First-stage F	15.62	162.0	81.19	207.9
AR p-value	0.627	0.427	0.704	0.118
$\Delta absence$	Yes	No	Yes	No
$(\Delta absence)^2$	No	Yes	Yes	No
$\|\Delta absence\|$	No	No	No	Yes

The dependent variable is a dummy indicating if the student has graduated within 7 months (0/1). The bottom panel contains indicators for which instruments have been used in the first-stage estimation. The instruments used are the first difference in weekly absence, the first difference in weekly absence squared, and the absolute difference in weekly absence. All models further include the log distance from students parish to the schools parish, age dummies, a dummy for males, dummies for first and second generation immigrants, and year, month, school, and education fixed effects. Standard errors clustered on the school by start date level are reported in parentheses and the reported F-values are from the first-stage. AR p-values are from the Anderson-Rubin wald test. Significance levels are reported as *** p<0.01, ** p<0.05, * p<0.1.

are no longer independent and the instrument is therefore not valid. In future work, we plan to include absence during the last year of compulsory school (grade 9). This way, we can create an instrument that is the average absence during the first two weeks subtracted by the average absence during the 9th grade. We would need to condition on positive absence during the first two weeks of vocational school, but would most likely not need to do any conditioning on 9th grade absence, if all students have at least one hour of absence during the year. This way we would avoid the censoring and break the negative correlation between between $\widetilde{a}_{it}$ and η_i.

Another planned extension is to wait for Statistics Denmark to release new population data on vocational school absence. We believe this data will be released some time in 2023. With population data, instead of data from 8 schools, we will have more variation in starting dates and more geographical variation that can hopefully strengthen our weather instrument and give more robust second stage estimate from this instruments.

8 Conclusion

To conclude, we find that absence during the first weeks of the second basic course can account for many of the students who drop out. Our preferred estimate entails that one day of additional absence for students in the second basic course with no prior spell in the first basic course are 19.5 percentage points less likely to graduate the second basic course within seven months. We find no significant results for the full sample or the sample of students with a prior spell in the first basic course. Our estimated coefficient is larger in magnitude than the OLS estimate.

We find that weather is stronger correlated with the groups of students who we expect to be more on the margin between attending school or not when the weather is bad, namely the students who do not live with their parents and student with no previous labor market attachment or school enrollment. This indicates that these are the group of students driving our first stage results and the identification of our results.

We check the robustness of our results with a second approach to identify the causal effect of absence on the probability of graduating, namely a new panel instrument. We find estimates which are close to the OLS estimates and lower than the results using weather measurements as instruments. These results are driven by the fact that our panel data instruments are generated from weekly observations of absence, which, to a high degree, are bottom censored. In future research, we aim to explore other suitable absence measures for students who do not suffer from as severe bottom censoring.

Acknowledgements

We thank the Rockwool Foundations Intervention Unit and Hans Henrik Sievertsen for helpful comments and suggestions. This paper has also benefited from seminar participants at CBS. The authors thank the Rockwool Foundation for financial support.

References

Anderson, T. W. and Rubin, H. (1949). Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations. *Annals of Mathematical Statistics*, 20:46–63.

Arellano, M. and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68:29–51.

Aucejo, E. M. and Romano, T. F. (2016). Assessing the effect of school days and absences on test score performance. *Economics of Education Review*, 55:70–87.

Auffhammer, M., Hsiang, S. M., Schlenker, W., and Sobel, A. (2013). Using Weather Data and Climate Model Output in Economic Analyses of Climate Change. *NBER Working Paper Series*, No. 19087.

Bertrand, M., Mogstad, M., and Mountjoy, J. (2021). Improving educational pathways to social mobility: evidence from norway's reform 94. *Journal of Labor Economics*, 39(4):965–1010.

Bingley, P., Heinesen, E., Krassel, K. F., and Kristensen, N. (2018). The Timing of Instruction Time: Accumulated Hours, Timing and Pupil Achievement. *IZA Discussion Papers*, No. 11807.

Carlsson, M., Dahl, G. B., Öckert, B., and Rooth, D.-O. (2015). The Effect of Schooling on Cognitive Skills. *The Review of Economics and Statistics*, 97(3):533–547.

Cattan, S., Kamhöfer, D. A., Karlsson, M., and Nilsson, T. (2023). The Long-term Effects of Student Absence : Evidence from Sweden. *The Economic Journal*, 133:888–903.

Confederation of Danish Employers (2023). Frafald på erhvervsuddannelserne sker oftest på grundforløbet. https://da.dk/politik-og-analyser/uddannelse-og-kompetencer/2023/frafald-paa-erhvervsuddannelserne-sker-oftest-paa-grundforloebet/?utm$_{m}edium = emailutm_{c}ampaign = DA Accessed : 2023-04-18$.

Craig, A. C. and Martin, D. C. (2019). Discipline Reform, School Culture, and Student Achievement Ashley. Working Paper.

Currie, J., Hanushek, E. A., Kahn, E. M., Neidell, M., and Rivkin, S. G. (2009). Does Pollution Increase School Absences? *The Review of Economics and Statistics*, 91(4):682–694.

Dell, M., Jones, B. F., and Olken, B. A. (2014). What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature*, 52(3):740–798.

Dominguez. Patricio and Ruffini, K. (2018). Long-term gains from longer school days. *IDB Working Paper Series*, No. 1120.

Eichhorst, W., Rodríguez-Planas, N., Schmidl, R., and Zimermann, K. F. (2015). A road map to vocational education and training in industrialized countries. *Industrial and Labor Relations Review*, 68(2):314–337.

Fitzpatrick, M. D., 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.

Gershenson, S. (2016). Linking teacher quality, student attendance, and student achievement. *Education Finance and Policy*, 11(2):125–149.

Goodman, J. S. (2014). Flaking out: Student absences and snow days as disruptions of instruction time. *NBER Working Paper Series*, No. 20221.

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

Gottfried, M. A. (2010). Evaluating the relationship between student attendance and achievement in urban elementary and middle schools: An instrumental variables approach. *American Educational Research Journal*, 47(2):434–465.

Groes, F. N., Madsen, E., and Sandoy, T. M. (2021). Completion from vocational educations: A register based analysis. Technical report, Copenhagen Business School.

Groppo, V. and Kraehnert, K. (2017). The impact of extreme weather events on education. *Journal of Population Economics*, 30(2):433–472.

Hampf, F. and Woessmann, L. (2017). Vocational vs. general education and employment over the life cycle: New evidence from PIAAC. *CESifo Economic Studies*, 63(3):255–269.

Hansen, B. (2011). School Year Length and Student Performance: Quasi-Experimental Evidence. *SSRN Electronic Journal*.

Hanushek, E. A. (2012). Dual Education: Europe's Secret Recipe? *CESifo Forum*, 13(3):29–34.

Hanushek, E. A., Schwerdt, G., Woessmann, L., and Zhang, L. (2017). General education, vocational education, and labor-market outcomes over the lifecycle. *Journal of Human Resources*, 52(1):48–87.

Heissel, J. A. and Norris, S. (2018). Rise and shine: The effect of school start times on academic performance from childhood through puberty. *Journal of Human Resources*, 53(4):957–992.

Kristensen, N., Jensen, V. M., and Krassel, K. F. (2020). Panelanalyse af bekymrende skolefravær. Technical report, VIVE.

Lavy, V. (2015). Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries. *Economic Journal*, 125(588):F397–F424.

Lavy, V. (2020). Expanding School Resources and Increasing Time on Task: Effects on Students' Academic and Noncognitive Outcomes. *Journal of the European Economic Association*, 18(1):232–265.

Liu, J., Lee, M., and Gershenson, S. (2019). The Short-and Long-Run Impacts of Secondary School Absences. *IZA Discussion Papers*, No. 12613.

Liu, J. and Loeb, S. (2019). Engaging Teachers: Measuring the Impact of Teachers on Student Attendance in Secondary School. *Journal of Human Resources*, 56(2):343–379.

Marcotte, D. E. (2007). Schooling and test scores: A mother-natural experiment. *Economics of Education Review*, 26(5):629–640.

Marcotte, D. E. and Hemelt, S. W. (2008). Unscheduled School Closings and Student Performance. *Education Finance and Policy*, 3(3):316–338.

Mellon, J. (2020). Rain, Rain, Go Away: 137 Potential Exclusion-Restriction Violations for Studies Using Weather as an Instrumental Variable. *SSRN Electronic Journal*.

Park, R. J., Goodman, J., Hurwitz, M., and Smith, J. (2020). Heat and learning. *American Economic Journal: Economic Policy*, 12(2):306–339.

Rivkin, S. G. and Schiman, J. C. (2015). Instruction Time, Classroom Quality, and Academic Achievement. *Economic Journal*, 125(588):F425–F448.

Robinson, C. D., Lee, M. G., Dearing, E., and Rogers, T. (2018). Reducing Student Absenteeism in the Early Grades by Targeting Parental Beliefs. *American Educational Research Journal*, 55(6):1163–1192.

Rogers, T. and Feller, A. (2016). Discouraged by Peer Excellence: Exposure to Exemplary Peer Performance Causes Quitting. *Psychological Science*, 27(3):365–374.

Sarsons, H. (2015). Rainfall and conflict: A cautionary tale. *Journal of Development Economics*, 115:62–72.

Silliman, M. and Virtanen, H. (2022). Labor market returns to vocational secondary education. *American Economic Journal: Applied Economics*, 14(1):197–224.

Stratton, L. S., Gupta, N. D., Reimer, D., and Holm, A. (2017). Modeling Enrollment in and Completion of Vocational Education: The role of cognitive and non-cognitive skills by program type. Working Paper.

The Danish Ministry of Education (2018). Karakterer fra folkeskolens afgangseksamen. https://www.stil.dk/-/media/filer/uvm/stat/pdf18/181025-tabelnotat-karakterer-fra-folkeskolens-afgangseksamen-2017-2018.pdf. Accessed: 2023-04-17.

The Danish Ministry of Education (2019). Vocational education and training in Denmark. https://eng.uvm.dk/upper-secondary-education/vocational-education-and-training-in-denmark. Accessed: 2023-04-18.

Thejill, P., Boberg, F., Schmith, T., Christiansen, B., Christensen, O. B., Madsen, M. S., Su, J., Andree, E., Olsen, S., Langen, P. L., Madsen, K. S., and Pedersen, R. A. (2020). Methods used in the Danish Climate Atlas. Technical report, The Danish Meteorological Institute, Copenhagen.

Tran, L. and Gershenson, S. (2018). Experimental Estimates of the Student Attendance Production Function. *IZA Discussion Paper*, 11911.

Zimmer, D. M. (2019). Missing School to Visit the Doctor? Analysis Using a Copula-Based Endogenous Switching Regressions Model. Working Paper.

A Additional tables and figures

Table 12: The Effect of pct. days with some absence during the first two weeks on graduation: OLS results

	Graduated (1)	Graduated (2)	Graduated (3)	Graduated (4)	Graduated (5)	Graduated (6)	Graduated (7)
Panel A: All							
Pct. days absence	-0.997***	-0.885***	-0.875***	-0.996***	-0.981***	-0.979***	-0.867***
	(0.048)	(0.050)	(0.050)	(0.049)	(0.053)	(0.053)	(0.050)
Observations	5782	5782	5782	5782	5782	5782	5782
R-squared	0.059	0.129	0.139	0.060	0.071	0.072	0.157
Panel B: With GF1							
Pct. days absence	-1.004***	-0.882***	-0.881***	-1.011***	-1.010***	-1.020***	-0.914***
	(0.137)	(0.110)	(0.112)	(0.142)	(0.142)	(0.149)	(0.114)
Observations	1833	1833	1833	1833	1833	1833	1833
R-squared	0.042	0.119	0.131	0.044	0.064	0.065	0.167
Panel C: Without GF1							
Pct. days absence	-0.992***	-0.869***	-0.858***	-0.985***	-0.949***	-0.942***	-0.834***
	(0.058)	(0.066)	(0.066)	(0.059)	(0.061)	(0.061)	(0.067)
Observations	3949	3949	3949	3949	3949	3949	3949
R-squared	0.065	0.143	0.154	0.068	0.081	0.084	0.172
Student controls	No	Yes	Yes	No	No	No	Yes
Parent controls	No	No	Yes	No	No	No	Yes
Year/Month FE	No	No	No	Yes	No	Yes	Yes
School/Education FE	No	No	No	No	Yes	Yes	Yes

Notes: The table shows the raw OLS estimates for pct. days with some absence during the first two weeks on graduation (0/1). The bottom panel indicates if student controls (age, immigration status, log distance, male, 9th grade danish score, 9th grade math score, and labour market attachment prior to enrolment), parent controls (highest completed parental education, mothers and fathers labour market attachment), year and month fixed effects, and school and education fixed effects are included in the models in the respective column. Panel A uses the full sample, Panel B only uses students with a prior GF1, and Panel C only uses students with no prior GF1. Standard errors clustered on the school by start date level are reported in parentheses. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

Table 13: Effect of absence on probability of graduation: 2SLS with weather as IV for absence (2 weeks)

	Graduated (1)	Graduated (2)	Graduated (3)	Graduated (4)
Panel A: All				
Pct. hours absent	-3.644	-1.068	-1.898*	-1.841**
	(3.085)	(1.438)	(1.104)	(0.911)
Observations	5782	5782	5782	5782
First-stage F	2.42	6.51	3.82	3.06
AR p-value	0.050	0.493	0.109	0.159
Panel B: With GF1				
Pct. hours absent	-1.952**	-1.059	-1.404**	-1.357**
	(0.928)	(0.817)	(0.636)	(0.611)
Observations	3949	3949	3949	3949
First-stage F	9.76	19.51	12.59	11.27
AR p-value	0.004	0.205	0.018	0.045
Precipitation > $3mm$	Yes	No	Yes	Yes
Wind > $11ms$	No	Yes	Yes	Yes
Precipitation×Wind	No	No	No	Yes

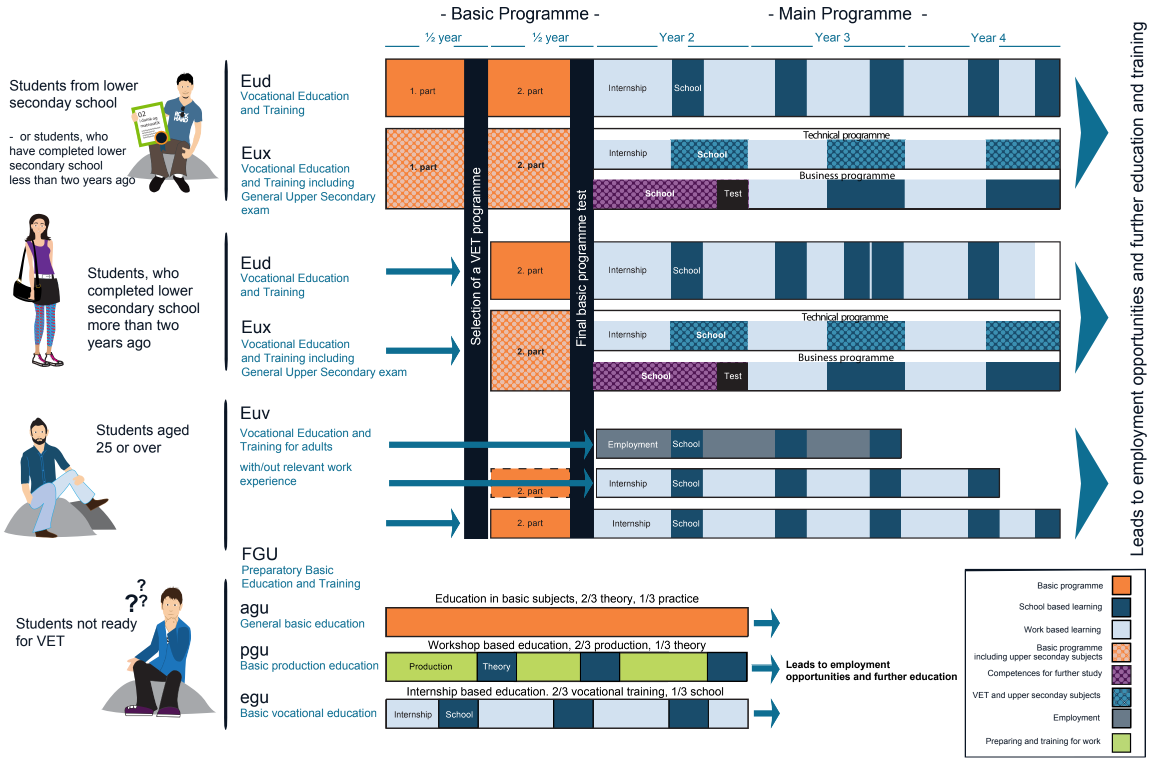
The dependent variable is a dummy indicating if the student has graduated within 7 months $(0/1)$. The bottom panel contains indicators for which instruments have been used in the first-stage estimation. The instruments used are pct. of days with over 3 mm of precipitation within the first two weeks, pct. days with average windspeed above 11 ms for a 10 min. interval, and the interaction of these two instruments. All models further include the log distance from students parish to the schools parish, age dummies, a dummy for males, dummies for first and second generation immigrants, and year, month, school, and education fixed effects. Standard errors clustered on the school by start date level are reported in parentheses and the reported F-values are from the first-stage. AR p-values are from the Anderson-Rubin wald test. Significance levels are reported as *** p<0.01, ** p<0.05, * p<0.1.

Table 14: Precipitation specification table, first-stage (2 weeks), without GF1

	1 mm	2 mm	3 mm	4 mm	5 mm	6 mm	7 mm	8 mm	9 mm
Precipitation	0.065***	0.075***	0.082***	0.075**	0.087**	0.047	0.108***	0.129***	0.101***
	(0.013)	(0.022)	(0.026)	(0.032)	(0.037)	(0.029)	(0.040)	(0.041)	(0.038)
F	25.00	11.73	9.76	5.44	5.53	2.54	7.41	9.89	7.18

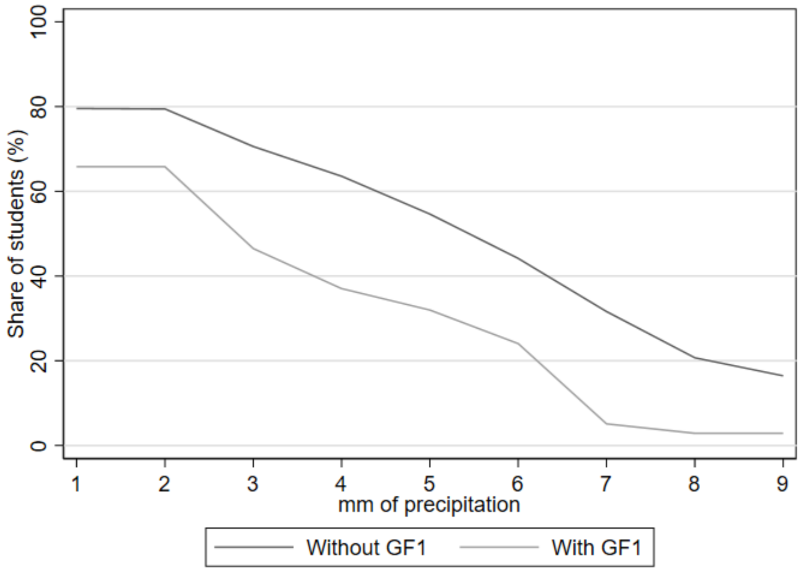
Notes: The reported estimates are the first-stage estimates from a regression with pct. hours of absence during the first two weeks with precipitation as instrument, using from 1 to 9 mm of precipitation as cutoffs for the instrument. Standard errors clustered on the school by start date level are reported in parentheses and F-tests for joint significance are reported as well. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

Figure 8: Pathways through the Danish VET system



Note: The figure is taken from the web-page of the Danish Ministry of Children and Education, where a description of the Danish VET system can also be found (The Danish Ministry of Education, 2019).

Figure 9: Share of students who have experienced at least one day with precipitation above a given number of mm



Note: The figure has the share of students who have experienced at least one day with a given number of mm or precipitation or more on the second axis and mm of precipitation on the first axis.

Table 15: Precipitation specification table, 2SLS (2 weeks), without GF1

	1 mm	2 mm	3 mm	4 mm	5 mm	6 mm	7 mm	8 mm	9 mm
Pct. hours absent	-1.687*	-0.942	-1.952**	-1.705*	-2.193*	-5.900	-1.615	-1.652	-1.718
	(0.906)	(0.809)	(0.928)	(1.009)	(1.243)	(4.443)	(1.228)	(1.106)	(1.449)
Observations	3949	3949	3949	3949	3949	3949	3949	3949	3949
F	25.00	11.73	9.76	5.44	5.53	2.54	7.41	9.89	7.18
AR p-value	0.039	0.226	0.004	0.026	0.048	0.019	0.138	0.139	0.225

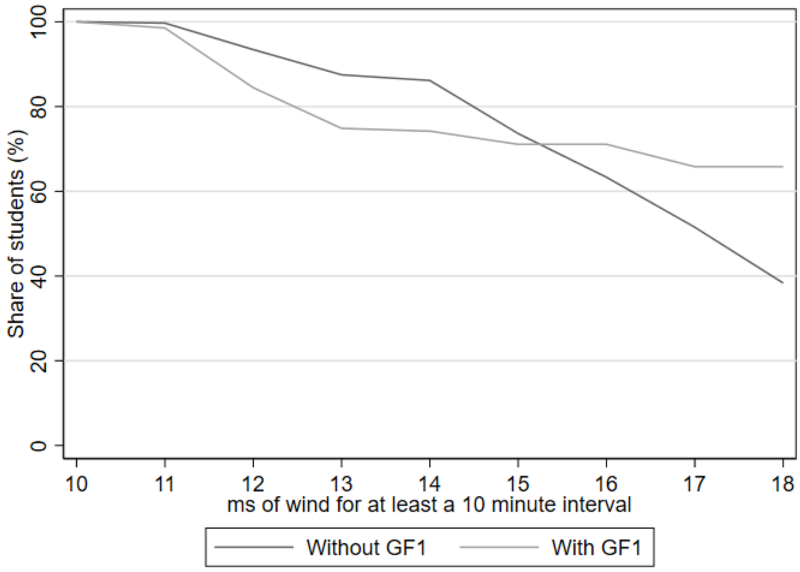
Notes: The reported estimates are the second-stage estimates from a regression with pct. hours of absence as endogenous variable while the dependent variable is a dummy indicating if the student has graduated within 7 months (0/1). The instrument used is precipitation, using from 1 to 9 mm of precipitation as cutoffs for the instrument. Standard errors clustered on the school by start date level are reported in parentheses and F-tests for joint significance from the first-stage are reported as well. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

Table 16: Wind specification table, first-stage (2 weeks), without GF1

	10 ms	11 ms	12 ms	13 ms	14 ms	15 ms	16 ms	17 ms	18 ms
Wind	0.071***	0.090***	0.082***	0.082***	0.084**	0.108***	0.119**	0.142	0.166*
	(0.022)	(0.020)	(0.021)	(0.019)	(0.034)	(0.028)	(0.052)	(0.090)	(0.096)
F	10.11	19.51	15.06	19.33	6.11	14.74	5.33	2.47	2.99

Notes: The reported estimates are the first-stage estimates from a regression with pct. hours of absence during the first two weeks with high wind as instrument, using from 10 to 18 ms of wind speed as cutoffs for the instrument. Standard errors clustered on the school by start date level are reported in parentheses and F-tests for joint significance are reported as well. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

Figure 10: Share of students who have experienced at least one day with wind above a given speed of ms



Note: The figure has the share of students who have experienced at least one day with wind speed of a given ms or more on the second axis and ms of wind on the first axis.

Table 17: Wind specification table, 2SLS (2 weeks), without GF1

	10 ms	11 ms	12 ms	13 ms	14 ms	15 ms	16 ms	17 ms	18 ms
Pct. hours absent	-1.828*	-1.059	-1.614*	-1.346*	0.316	-0.534	-1.612	0.960	1.755
	(0.945)	(0.817)	(0.850)	(0.749)	(1.039)	(0.862)	(1.269)	(1.747)	(1.404)
Observations	3949	3949	3949	3949	3949	3949	3949	3949	3949
F	10.11	19.51	15.06	19.33	6.11	14.74	5.33	2.47	2.99
AR p-value	0.025	0.205	0.048	0.075	0.758	0.549	0.221	0.601	0.193

Notes: The reported estimates are the second-stage estimates from a regression with pct. hours of absence as endogenous variable while the dependent variable is a dummy indicating if the student has graduated within 7 months (0/1). The instrument used is high wind speed, using from 10 to 18 ms of wind speed as cutoffs for the instrument. Standard errors clustered on the school by start date level are reported in parentheses and F-tests for joint significance from the first-stage are reported as well. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

Table 18: Interaction of precipitation and wind specification matrix table, 2SLS (2 weeks), without GF1

	1 mm	2 mm	3 mm	4 mm	5 mm	6 mm	7 mm	8 mm	9 mm
10 ms	-1.741**	-1.390*	-1.888**	-1.792**	-1.940**	-1.920**	-1.770*	-1.778**	-1.815*
	(0.826)	(0.760)	(0.806)	(0.853)	(0.879)	(0.958)	(0.922)	(0.885)	(0.942)
F	11.54	7.46	7.00	5.63	5.85	5.33	6.50	7.25	5.44
AR p-value	0.042	0.072	0.010	0.031	0.052	0.021	0.069	0.062	0.075
11 ms	-1.352*	-1.013	-1.404**	-1.210*	-1.340*	-1.104	-1.130	-1.138	-1.068
	(0.725)	(0.664)	(0.636)	(0.708)	(0.736)	(0.817)	(0.825)	(0.819)	(0.821)
F	17.20	12.54	12.59	10.00	10.51	9.95	10.97	11.12	9.90
AR p-value	0.113	0.303	0.018	0.085	0.110	0.062	0.309	0.291	0.391
12 ms	-1.655**	-1.309*	-1.767**	-1.641**	-1.794**	-1.799**	-1.614*	-1.623*	-1.626*
	(0.795)	(0.719)	(0.694)	(0.736)	(0.750)	(0.844)	(0.858)	(0.830)	(0.865)
F	15.58	11.03	11.53	8.27	8.90	7.93	9.16	9.35	7.89
AR p-value	0.083	0.129	0.008	0.043	0.044	0.024	0.134	0.120	0.139
13 ms	-1.605**	-1.101	-1.691***	-1.489**	-1.700**	-1.564**	-1.446*	-1.465**	-1.439*
	(0.786)	(0.671)	(0.654)	(0.676)	(0.731)	(0.777)	(0.788)	(0.712)	(0.746)
F	17.36	12.14	13.26	10.15	10.10	10.58	10.68	14.75	11.97
AR p-value	0.093	0.177	0.011	0.074	0.079	0.054	0.160	0.137	0.179
14 ms	-1.307*	-0.539	-1.113*	-0.625	-0.903	-0.375	-0.544	-0.573	-0.325
	(0.771)	(0.649)	(0.607)	(0.684)	(0.712)	(0.908)	(0.795)	(0.792)	(0.887)
F	13.10	7.26	7.39	4.18	5.13	5.20	6.89	9.73	6.22
AR p-value	0.046	0.411	0.014	0.063	0.116	0.050	0.267	0.267	0.384
15 ms	-1.397*	-0.778	-1.333**	-1.012	-1.235	-0.934	-0.962	-0.989	-0.878
	(0.780)	(0.665)	(0.665)	(0.694)	(0.782)	(0.899)	(0.768)	(0.743)	(0.800)
F	14.82	9.90	11.44	8.26	8.91	9.38	11.46	15.81	11.75
AR p-value	0.097	0.463	0.015	0.081	0.140	0.064	0.316	0.301	0.430
16 ms	-1.684*	-1.049	-1.867**	-1.670*	-1.991**	-2.436*	-1.614*	-1.636*	-1.662*
	(0.897)	(0.777)	(0.835)	(0.897)	(0.995)	(1.346)	(0.962)	(0.893)	(1.007)
F	12.97	7.15	7.52	4.52	5.46	5.51	8.06	11.50	7.72
AR p-value	0.115	0.329	0.013	0.073	0.125	0.066	0.231	0.174	0.242
17 ms	-1.739*	-0.897	-1.564**	-1.180	-1.647*	-2.007	-1.115	-1.032	-0.858
	(0.913)	(0.801)	(0.763)	(0.768)	(0.985)	(1.736)	(1.116)	(1.035)	(1.238)
F	13.08	6.84	7.01	5.39	5.56	4.17	5.42	11.63	7.50
AR p-value	0.027	0.301	0.017	0.079	0.114	0.052	0.235	0.270	0.385
18 ms	-1.786*	-0.937	-1.614**	-1.208	-1.702*	-1.604	-0.996	-1.049	-0.769
	(0.927)	(0.807)	(0.779)	(0.787)	(1.023)	(1.674)	(1.021)	(0.974)	(1.152)
F	13.82	7.46	7.41	5.71	5.59	3.59	5.35	8.01	6.29
AR p-value	0.017	0.135	0.014	0.056	0.063	0.034	0.137	0.125	0.180

Notes: The reported estimates are the first-stage estimates from a regression with pct. hours of absence during the first two weeks with the precipitation instrument interacted with the high wind instrument, using from 1 to 9 mm of precipitation (columns) and 10 to 18 ms of wind speed (rows) as cutoffs for the instrument. Standard errors clustered on the school by start date level are reported in parentheses and F-tests for joint significance are reported as well. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

Table 19: First-stage estimates for the different specifications of weather instruments, percent days absent used as outcome (2 weeks), without GF1

	Pct. days absent	Pct. days absent	Pct. days absent	Pct. days absent
	(1)	(2)	(3)	(4)
Pct. days with precipitation>3mm	0.073**		0.052	0.014
	(0.036)		(0.038)	(0.036)
Pct. days with wind>11ms		0.066***	0.046**	0.025
		(0.023)	(0.022)	(0.024)
Precipitation X wind				0.179
				(0.111)
Observations	3949	3949	3949	3949
F	4.05	8.29	5.60	4.51

Notes: The table shows the first-stage estimates for the different specifications of the weather instruments on our endogenous measure of interest, pct. days absent during the first three weeks. All models contain additional controls (age, immigration status, log distance, and male), year and month fixed effects, and school and education fixed effects. F-tests for joint significance are reported for each panel. Standard errors clustered on the school by start date level are reported in parentheses. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

Table 20: Effect of absence on probability of graduation: 2SLS with weather as IV for absence (2 weeks)

	Graduated	Graduated	Graduated	Graduated
	(1)	(2)	(3)	(4)
Pct. days absent	-2.195*	-1.481*	-1.848**	-1.809**
	(1.120)	(0.871)	(0.828)	(0.793)
Observations	3949	3949	3949	3949
F	4.05	8.29	5.60	4.51
AR p-value	0.001	0.079	0.003	0.004
Precipitation > $3mm$	Yes	No	Yes	Yes
Wind > $11ms$	No	Yes	Yes	Yes
Precipitation×Wind	No	No	No	Yes

The dependent variable is a dummy indicating if the student has graduated within 7 months (0/1) and the endogenous variable is pct. days absent during the first two weeks. The bottom panel contains indicators for which instruments have been used in the first-stage estimation. The instruments used are pct. of days with over 3 mm of precipitation within the first two weeks, pct. days with average windspeed above 11 ms for a 10 min. interval, and the interaction of these two instruments. All models further include the log distance from students parish to the schools parish, age dummies, a dummy for males, dummies for first and second generation immigrants, and year, month, school, and education fixed effects. Standard errors clustered on the school by start date level are reported in parentheses and the reported F-values are from the first-stage. AR p-values are from the Anderson-Rubin wald test. Significance levels are reported as *** p<0.01, ** p<0.05, * p<0.1.

Table 21: Heterogeneity in terms of living with parents: 2SLS with weather as IV for absence (2 weeks)

	Does not live with parents		Lives with parents	
	(1)	(2)	(3)	(4)
Second stage estimates				
Pct. hours absent	-1.765**	-1.832	-2.386	-0.061
	(0.850)	(1.173)	(3.405)	(1.178)
OLS estimates				
Pct. hours absent	-0.899***		-0.847***	
	(0.077)		(0.089)	
First stage estimates				
Pct. days with precipitation>3mm	0.093***		0.021	
	(0.019)		(0.019)	
Pct. days with wind>11ms		0.094***		0.073***
		(0.031)		(0.027)
Observations	2287	2287	1662	1662
F	24.11	9.03	1.16	7.25
AR p-value	0.023	0.110	0.437	0.960
Precipitation $> 3mm$	Yes	No	Yes	No
Wind $> 11ms$	No	Yes	No	Yes

All columns only include students without the first basic course. The first two columns only include students who do not live with their parents, while the third and fourth columns include students who live with their parents. The top panel is the second-stage estimates from a regression where the dependent variable is a dummy indicating if the student has graduated within 7 months (0/1) and the endogenous variable is pct. hours absent during the first two weeks using the instruments marked in the bottom panel. The second panel shows the OLS estimates from running the same regression without instruments. The third panel reports the first-stage estimates using pct. hours absent as dependent variable and the instruments marked in the bottom panel as explanatory varibles. The instruments used are pct. of days with over 3 mm of precipitation within the first two weeks, pct. days with average windspeed above 11 ms for a 10 min. interval, and the interaction of these two instruments. All models further include the log distance from students parish to the schools parish, age dummies, a dummy for males, dummies for first and second generation immigrants, and year, month, school, and education fixed effects. Standard errors clustered on the school by start date level are reported in parentheses and the reported F-values are from the first-stage. AR p-values are from the Anderson-Rubin wald test. Significance levels are reported as *** p<0.01, ** p<0.05, * p<0.1.

Table 22: Heterogeneity in terms of a prior stable attachment: 2SLS with weather as IV for absence (2 weeks)

	Work or in school		No stable attachment	
	(1)	(2)	(3)	(4)
Second stage estimates				
Pct. hours absent	-2.253**	-1.351	-1.334	-0.758
	(0.915)	(0.830)	(1.112)	(1.282)
OLS estimates				
Pct. hours absent	-0.760***	-0.760***	-1.118***	-1.118***
	(0.075)	(0.075)	(0.077)	(0.077)
First stage estimates				
Pct. days with precipitation>3mm	0.057***		0.096***	
	(0.011)		(0.024)	
Pct. days with wind>11ms		0.081***		0.115***
		(0.019)		(0.044)
Observations	2912	2912	1037	1037
F	27.10	18.89	15.56	6.71
AR p-value	0.005	0.117	0.245	0.584
Precipitation > $3mm$	Yes	No	Yes	No
Wind > $11ms$	No	Yes	No	Yes

All columns only include students without the first basic course. The first two columns only include students who where working or in school before their enrolment, while the third and fourth columns include students without a stable attachment before their enrolment. The top panel is the second-stage estimates from a regression where the dependent variable is a dummy indicating if the student has graduated within 7 months (0/1) and the endogenous variable is pct. hours absent during the first two weeks using the instruments marked in the bottom panel. The second panel shows the OLS estimates from running the same regression without instruments. The third panel reports the first-stage estimates using pct. hours absent as dependent variable and the instruments marked in the bottom panel as explanatory varibles. The instruments used are pct. of days with over 3 mm of precipitation within the first two weeks, pct. days with average windspeed above 11 ms for a 10 min. interval, and the interaction of these two instruments. All models further include the log distance from students parish to the schools parish, age dummies, a dummy for males, dummies for first and second generation immigrants, and year, month, school, and education fixed effects. Standard errors clustered on the school by start date level are reported in parentheses and the reported F-values are from the first-stage. AR p-values are from the Anderson-Rubin wald test. Significance levels are reported as *** p<0.01, ** p<0.05, * p<0.1.

Table 23: Heterogeneity in terms of grades: 2SLS with weather as IV for absence (2 weeks)

	High grade		Low grade	
	(1)	(2)	(3)	(4)
Second stage estimates				
Pct. hours absent	-3.163**	-4.241**	-0.918	1.715
	(1.504)	(2.091)	(1.218)	(1.436)
OLS estimates				
Pct. hours absent	-0.855***		-0.838***	
	(0.073)		(0.089)	
First stage estimates				
Pct. days with precipitation>3mm	0.049**		0.069***	
	(0.019)		(0.027)	
Pct. days with wind>11ms		0.055**		0.117***
		(0.025)		(0.035)
Observations	1942	1942	1297	1297
F	6.33	4.87	6.66	11.09
AR p-value	0.007	0.007	0.423	0.164
Precipitation > $3mm$	Yes	No	Yes	No
Wind > $11ms$	No	Yes	No	Yes

All columns only include students without the first basic course. The first two columns only include students who had a 9th grade Math grade of 4 or below, while the third and fourth columns include students who had a higher 9th grade Math grade. The top panel is the second-stage estimates from a regression where the dependent variable is a dummy indicating if the student has graduated within 7 months (0/1) and the endogenous variable is pct. hours absent during the first two weeks using the instruments marked in the bottom panel. The second panel shows the OLS estimates from running the same regression without instruments. The third panel reports the first-stage estimates using pct. hours absent as dependent variable and the instruments marked in the bottom panel as explanatory varibles. The instruments used are pct. of days with over 3 mm of precipitation within the first two weeks, pct. days with average windspeed above 11 ms for a 10 min. interval, and the interaction of these two instruments. All models further include the log distance from students parish to the schools parish, age dummies, a dummy for males, dummies for first and second generation immigrants, and year, month, school, and education fixed effects. Standard errors clustered on the school by start date level are reported in parentheses and the reported F-values are from the first-stage. AR p-values are from the Anderson-Rubin wald test. Significance levels are reported as *** p<0.01, ** p<0.05, * p<0.1.