RIETI Discussion Paper Series 26-E-026

Long-term Absenteeism: Effects of cognitive skills, non-cognitive skills, household structure and financial situation

ASAKAWA, Shinsuke
Saga University

ABE, Mayuko
Japan Center for Economic Research

OHTAKE, Fumio
RIETI

SANO, Shinpei
Kobe University

NAKATA, Kazuko
Setsunan University

RIETI
Research Institute of Economy, Trade & Industry, IAA
The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 26-E-026
April 2026

Long-Term Absenteeism: Effects of Cognitive Skills, Non-Cognitive Skills, Household Structure and Financial Situation†

Shinsuke Asakawa (Saga University), Mayuko Abe (Japan Center for Economic Research),
Fumio Ohtake (Osaka University/RIETI), Shinpei Sano (Kobe University),
Kazuko Nakata (Setsunan University)

Abstract

Using administrative data from Amagasaki City (2019–2023), this study identified the factors associated with long-term absenteeism among elementary and junior high school students. Ordinary least square regressions revealed that students with low mathematics scores and those from single-parent or welfare-recipient households faced a higher risk of long-term absenteeism. Regarding non-cognitive skills, lower levels of extraversion, agreeableness, conscientiousness and emotional stability, and higher openness correlated with increased absenteeism. Notably, the probability of long-term absence remains substantially higher in 2023 than in 2019, even after controlling these characteristics. Blinder–Oaxaca decomposition shows that the increase in absenteeism during the COVID-19 pandemic was not driven by changes in student attributes but by the amplified impact of academic achievement, non-cognitive skills, and family environment. For elementary school students, class size was also an influential factor. However, a significant portion of the increase remains unexplained by the observed variables, suggesting that uncaptured structural or environmental shifts likely played substantial roles.

Keywords: long-term absenteeism, cognitive skills, non-cognitive traits, administrative data
JEL Classification: I20, I24, I31

† This study is conducted as a part of the Project "Comprehensive Study on Establishing Effective EBPM" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). The draft of this paper was presented at the RIETI DP seminar for the paper. We would like to thank participants of the RIETI DP Seminar for their helpful comments. This study is also part of a project conducted by the Amagasaki City Institute for Learning and Development. We are deeply grateful to the members of the Institute for their invaluable advice and support throughout this research. We also thank Eiji Yamamura (Seinan Gakuin University), Tomohiko Inui (Gakushuin University), Jun Iritani (Kobe University), as well as participants at the 18th Annual Conference of the Association of Behavioral Economics and Finance and the 2025 Spring Meeting of the Japan Economic Association, for their insightful comments. This work was supported by JSPS KAKENHI Grant Numbers JP25H00388, JP23K25529 and Saga University.

1. Introduction

Long-term student absenteeism is globally recognized as a critical issue that impedes human capital formation. The literature demonstrates that absenteeism not only diminishes academic performance and lower educational attainment, but also adversely affects long-term labor market outcomes, such as wages and employment (Chang and Romero 2008; Gottfried 2010, 2011, 2014; Goodman 2014; Aucejo and Romano 2016; Cattan et al. 2023; Liu et al. 2021; Blanden et al. 2023). While the negative consequences of long-term absenteeism are well-documented, the specific factors that make certain students more susceptible to long-term absenteeism have not yet been fully elucidated in the economics literature. Identifying these determinants has become an urgent priority, especially because the number of students experiencing long-term absenteeism has significantly increased since the COVID-19 pandemic.

Although their scope remains limited, several economic empirical studies have investigated the determinants of long-term absenteeism. Extant studies highlight factors such as class size, relative socioeconomic status within the classroom, fluctuations in household income, and level of social trust (Nakamuro 2017; Tanaka and Morozumi 2021; Inoue and Tanaka 2024; Gennetian et al. 2018; Yamamura 2011). Furthermore, pedagogy, psychology, sociology, and public health research suggests that the risk of long-term absenteeism increases when students face socioeconomic disadvantages, hold negative attitudes toward school, or have limited engagement between parents and schools (Sosu et al. 2021; Gubbels et al. 2019).[1] In sum, long-term absenteeism is a complex phenomenon driven by the interplay of family and school-level factors, especially socioeconomic status.

Nevertheless, several important issues remain unaddressed. First, few studies have hitherto considered the student attributes that play a central role in the education production function, especially cognitive and non-cognitive skills and household structures (Rumberger 1995; Balkis et al 2016; Soma et al 2025). As these skills are empirically linked to long-term labor market outcomes (Deming 2017; Lee and Ohtake 2018; Edin et al. 2022), their influence on absenteeism cannot be overlooked. Second, few studies have examined students' specific attributes based on their reasons for absenteeism, regardless of their varied backgrounds.[2] If student backgrounds differ depending on the cause of their absence, interventions must also differ. Third, despite the post-pandemic surge in long-term absenteeism, few studies have investigated how the correlation between student attributes and

[1] Gubbels et al. (2019) also find that the risk of dropout, which is a consequence of long-term absenteeism, is associated with a low IQ or learning difficulties and low academic achievement.

[2] According to PISA 2022, among students who reported having been absent from school for more than three consecutive months, the most frequently cited reason was "I was sick" (70.5%). The next most common reasons were "I did not feel safe at school" and "I was bored" (both 18.5%). These were followed by "School was closed because of a natural disaster" (17.6%) and "I had to take care of a family member" (16.5%) (OECD 2023).

absence has evolved. With absenteeism rates doubling in developed countries such as the US and the UK following the pandemic, it is crucial to identify structural shifts in the drivers of absenteeism.[3]

This study addresses these three critical issues by using administrative microdata from 2018 to 2023 from Amagasaki City, a Japanese municipality. First, we analyze the attributes (cognitive/non-cognitive skills and household structure) associated with a higher propensity for long-term absenteeism. Specifically, we employ an ordinary least squares (OLS) model and use a dummy variable that indicates long-term absence as a dependent variable. Explanatory variables include prior-year test scores in Japanese and mathematics, prior-year non-cognitive evaluations, gender, household attributes (e.g., public assistance or school expense subsidies), and family environment (e.g., single-parent status), lifestyle habit (time use) and physical characteristics. Furthermore, following Chetty et al. (2011), who demonstrate the importance of classroom environments, we included class size and other classroom environment variables as explanatory variables. Second, we categorize the reasons for long-term absence into three types, namely school refusal, family/personal circumstances, and sickness, and analyze the student attributes and family backgrounds associated with each type. Third, we examine how the correlation between student/household characteristics and long-term absenteeism changed after the COVID-19 pandemic based on Blinder–Oaxaca decomposition (Blinder 1973; Oaxaca 1973). This method allows us to clarify whether the post-pandemic increase in absenteeism has been driven by changes in student/household compositions (the endowment effect) or by changes in the magnitude of the impact these attributes have on absentee behavior (the coefficient effect).

Focusing on Japan as a case study for long-term absenteeism offers two primary contributions to the literature. First, despite historically low absenteeism rates in Japan, the number of absent students has increased since the beginning of the COVID-19 pandemic. Recent international research, primarily in Western countries, defines "chronic absenteeism" as missing 10% or more of the school year (approximately 18 days or more) and has empirically examined its negative impacts (Chang and Romero 2008; Gottfried 2014). In contrast, "long-term absenteeism" in the Japanese educational system is defined as missing 30 days or more per year. This suggests that our analysis focuses on a more severe state of absenteeism compared to international standards. Even under this stricter definition, the percentage of students absent for 30 days or more in Japan doubled from 2.47% in FY2018 to 5.52% in FY2024, continuing to reach record highs. The fact that a similar doubling

[3] For example, in the US, the share of students who missed 10% or more of school days rose from around 15% before the pandemic to nearly 30% afterward, while in the UK, it increased from approximately 10% to 20% (U.S. Department of Education 2025; Malkus 2025; Department for Education England 2025). Similar post-pandemic increases in student absenteeism and declines in attendance rates have been reported in Australia, Canada, and the Netherlands (Australian Curriculum, Assessment and Reporting Authority 2025; CBC News 2025; NL Times 2025).

trend—comparable to the 10% threshold in the U.S. (U.S. Department of Education 2025)—is observed even under Japan's stricter 30-day definition underscores the severity and universality of the pandemic's impact. Notably, this surge occurred despite Japan having relatively limited restrictions on in-person classes compared to other OECD countries (OECD 2023). [4] Under these circumstances, identifying the student attributes associated with the highest risks has become a critical challenge for educational policy both in Japan and abroad.

Second, highly linked administrative panel data are available in certain Japanese municipalities. [5] The data used in this study encompass not only long-term absence records but also individual-level longitudinal information, including standardized test scores, non-cognitive skill evaluations, household structures, public assistance, school expense subsidy status, and Basic Resident Register information. Furthermore, the Japanese public-school system is characterized by extremely rigorous attendance management and precludes punitive measures such as 'suspension,' which are common in the United States. Consequently, Japanese absenteeism data contains minimal noise from such external factors, providing an internationally exceptional sample for capturing 'pure attendance behavior' rooted in student attributes and family environments. These individual-level panel data, based on a census of all students, overcome the limitations of previous studies (e.g., Higeta and Suetomi 2013; Aoki et al. 2020, 2021) that relied on aggregated data by region or school. This enables a precise examination of the relationship between diverse student characteristics and absenteeism behavior.

The empirical results indicate that students from single-parent households, families receiving public assistance, and those with lower mathematics scores pose a significantly higher risk of long-term absenteeism. Regarding non-cognitive skills, a higher risk of absenteeism was observed among students with lower levels of extraversion, agreeableness, conscientiousness, and emotional stability,

---

[4] According to PISA 2022, only 15.5% of the students in Japan reported that their school building was closed for more than three months due to COVID-19, which is the fourth-lowest figure among OECD countries and far below the OECD average of 50.3% (OECD 2023).

[5] In Japan, only a limited number of local governments maintain individual-level administrative data in a format accessible to researchers and, as a result, most existing studies on long-term absenteeism rely on school- or municipality-level aggregated data. For example, Higeta and Suetomi (2013) use municipality-level panel data to show that in municipalities where a larger number of elementary school districts are integrated into a single junior high school district, students experience more significant environmental changes during the transition from elementary to junior high school, which indicates a higher likelihood of long-term absenteeism. Using school-level microdata from the national Survey on Problematic Student Behavior, Aoki et al. (2020, 2021) document that absenteeism accumulates over grades in junior high school and suggest that reductions in the number of classes at grade transitions may contribute to increases in absenteeism. Studies using individual-level data are relatively rare, but a few exceptions exist. For instance, Hosaka (1996) analyzes a three-year follow-up survey of long-term absentees from all public elementary and junior high schools in a single city and finds that long-term absenteeism more likely to persist at the junior high school level than at the elementary school level. Using nine years of individual panel data from all public schools in a mid-sized city, Ito et al. (2017) show that increases in class size are associated with reduced teacher support and peer assistance. Focusing on family background, Sudo (2024) and Kajiwara (2021) demonstrate that household characteristics—particularly single-parent households and parental educational attainment—are more strongly associated with absenteeism and that children from low-income households are more likely to experience long-term absenteeism.

as well as higher levels of openness to experience. Notably, even after controlling these attributes, the probability of long-term absenteeism was substantially higher in 2023 than in 2019. Furthermore, the characteristics of students and their family environments varied depending on the reasons for their absence. Students absent due to school refusal or sickness tended to have specific personality traits, while students absent due to sickness also faced additional economic and domestic challenges. By contrast, for students absent due to family/personal circumstances, academic and personality traits showed no significant impact, suggesting that their absenteeism is primarily driven by economic factors.

The results of the Blinder–Oaxaca decomposition suggest that the post-pandemic surge in long-term absenteeism is not attributable to changes in student composition (the endowment effect) but rather to the amplified impact of existing factors, such as academic performance, non-cognitive skills, and family environment, on absentee behavior (the coefficient effect). For elementary school students, the increasing influence of class size is also a contributing factor. However, a large portion of the increase is captured by the constant term, which cannot be fully explained by the observed variables, suggesting that unobserved structural or environmental shifts play a pivotal role.

The remainder of this paper is organized as follows: Section 2 reviews the institutional background, including Japan's policy for addressing school refusal. Section 3 describes the dataset. Section 4 outlines the identification strategies used in the study. Section 5 presents the primary estimation results, robustness checks, reason-specific analyses, and results of the Blinder–Oaxaca decomposition. Finally, Section 6 concludes the study.

2. Institutional Background

2.1. Definition of Long-Term Absenteeism

The definition of long-term absenteeism varies significantly across countries and educational systems. For instance, the U.S. Department of Education (2025) defines chronic absenteeism as missing 10% or more of the school days in an academic year—equivalent to approximately 18 days—regardless of the reason or whether the absence is excused or unexcused. By contrast, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan defines long-term absenteeism as missing a total of 30 days or more within a single academic year, regardless of whether the absences are continuous or intermittent.

According to official Japanese statistics, long-term absences are subdivided according to their primary underlying reasons. The most prominent category is school refusal (*futoko*), which refers to situations where a student does not or cannot attend school due to psychological, emotional, physical, or social factors, specifically excluding those cases primarily driven by illness or economic hardship. Other distinct categories include absences due to sickness, those resulting from economic reasons related to household financial difficulties, and other circumstances, which primarily encompass various family or personal matters.

2.2. Attendance within the Japanese Educational System

Japan has implemented a nationwide compulsory education system consisting of six years of elementary school and three years of junior high school education. During this period, grade retention is extremely rare; students typically progress to the next grade and complete their compulsory education regardless of their attendance records. Furthermore, Japanese schools generally maintain approximately 200 instructional days per year, a relatively high figure compared to many other OECD member nations (OECD 2016).

In Japan, school attendance is strictly regulated through legally coded procedures. Under the Order for Enforcement of the School Education Act (Articles 19–21), school principals are mandated to monitor the attendance status of all enrolled students continuously. If a student is absent for seven consecutive days without a justifiable reason or if their attendance is deemed significantly poor, the principal is required to notify the local education board. Upon receiving such a notification, the board is legally obligated to contact the students' parents or guardians to encourage them to attend school. These provisions underscore that daily attendance records are not merely a matter of school convention but are managed within a robust legal framework that requires coordination between schools, guardians, and local authorities.

In practice, homeroom teachers monitor daily attendance. Specifically, they record the attendance status of each student in an official attendance register (*shussekibo*) for every school day, while also identifying and documenting the reasons for absences based on communication with the students and their guardians. At the end of each academic year, these cumulative records are compiled at the school level and reported to the MEXT via the local boards of education. Consequently, the attendance data used in this study were derived from administrative records generated through routine and continuous procedures at the school level. Unlike retrospective surveys, these data are highly reliable and accurate.

2.3 Transformation of Attendance Management during the Pandemic

In response to the spread of COVID-19, Japanese schools implemented emergency measures between 2020 and 2021, including temporary nationwide closures, staggered school attendance, and the introduction of online instruction. Consequently, the MEXT and local governments have adopted flexible applications to classify attendance. Specifically, administrative adjustments were made such that absences were not officially recorded as "absent" in the cumulative school ledger (*shido yoroku*) under certain conditions. These included absences due to viral infection, parental decisions to stay home to mitigate infection risks, and vaccinations. Such cases were instead categorized as "suspension of attendance" (*shusseki teishi*), thereby excluding them from the official count of days absent.[6]

Although in Japan in-person classes resumed relatively early compared to many other advanced economies, the prevalence of long-term absenteeism has remained exceptionally high since the pandemic. A critical point is that the fundamental administrative definition of long-term absenteeism—defined as 30 or more days of absence per year—remained unchanged throughout the pre- and post-pandemic periods. By utilizing indicators based on this consistent definition, this study can examine trends in long-term absenteeism and the underlying structural changes.

2.4 Characteristics and Socioeconomic Background of Amagasaki City

Amagasaki City, located in Hyogo Prefecture adjacent to Osaka City, is a representative urban municipality, with a population of approximately 460,000, being characterized by a mix of industrial, commercial, and residential areas (Amagasaki City 2025). One feature of the city is its relatively disadvantaged socioeconomic profile compared to the national average. Specifically, the public assistance rate in Amagasaki is approximately 4%, which significantly exceeds the national average of approximately 1.5%, indicating a high concentration of economically distressed households. Furthermore, approximately 25% of the students in compulsory education receive school expense subsidies designated for low-income households, which is markedly higher than the national average of approximately 15%. Additionally, the results from the National Assessment of Academic Ability conducted by the MEXT indicate that the city's average scores tend to fall below the national average and that the city faces educational challenges. Consequently, data from Amagasaki City provide a highly informative sample for examining how the risks of long-term absenteeism manifest in urban areas facing social and economic challenges.

Notably, the long-term absenteeism rates in Amagasaki City (1.34% for elementary and 2.40% for junior high schools between 2019 and 2023) remain at or below the national average (approximately 2.47% in 2018), showing broad consistency with national trends. The fact that Amagasaki maintains

---

[6] Regarding statistical data on long-term absenteeism, previous studies highlight issues arising from transitions in the definition of long-term absence and the classification criteria for reasons of absence. Hosaka (2009, 2024) and Yamamoto (2008) demonstrate that periodic changes in criteria and terminology undermine data continuity. Particularly during the recent COVID-19 pandemic, changes in statistical items have made rigorous comparisons across regions and time periods difficult (Hosaka and Shige 2023). Nevertheless, despite these limitations, the rate of long-term absenteeism remains a crucial indicator for understanding the phenomenon of school refusal (Yamamoto 2008).

a representative level of absenteeism despite its socioeconomically disadvantaged conditions suggests that it serves as an appropriate standard model for elucidating the mechanisms of long-term absenteeism in urban Japan. Given these factors, the findings of this study focusing on Amagasaki City hold significant implications for developing countermeasures against long-term absenteeism across Japan.

3. Data and Variables

3.1 Data

This study employs a longitudinal panel dataset constructed by integrating six types of administrative microdata provided by Amagasaki City, Hyogo Prefecture, Japan. First, we utilize the Amakko Survey, which includes the results of academic achievement tests administered annually to students from the first grade of elementary school to the second grade of junior high school. This dataset also contains student attributes such as individual IDs, academic years, and school/class information, alongside comprehensive survey responses covering lifestyle habits, family environments, learning attitudes, and school conditions. Second, we incorporate the Long-term Absenteeism Records, which document students being absent for 30 or more days per year. These records include individual IDs and school/class information, as well as the primary reason for each absence. Third, the Basic Resident Register provides fundamental attributes such as individual and household IDs, year and month of birth, gender, household relationships, and school districts. Fourth, School Expense Subsidies data, which record the status of public assistance and quasi-public assistance (subsidies for low-income households) at the individual level, is utilized as a proxy for household socioeconomic status. Fifth, the Student Enrollment Records provide the number of students and classes for each grade within each school; we use this information to calculate class size (the average number of students per class) at the school-grade level. Finally, we include height and weight data obtained from annual health checkups conducted every April. These physical measurements play a dual role in this analysis: first, to control the increasing trend of childhood obesity following pandemic-related lifestyle changes and, second, to account for potential psychological stress or bullying associated with physical characteristics, as suggested in the literature (Janssen et al. 2004).

To analyze the relationship between long-term absenteeism and student attributes including academic achievement, non-cognitive skills, and family characteristics, we integrate six datasets using student IDs and academic years as unique keys. The final analytical sample consists of students who participated in the Amakko Survey. Specifically, the analysis compares the students who participated in the survey in the previous year and became long-term absentees in the current year with those who participated in the survey in the previous year and did not become long-term absentees in the current year.

While the Amakko Survey covers the period from 2018 to 2023, the scope of our analysis is constrained by the nature of the available variables. Questions regarding noncognitive skills were primarily administered to students in the third grade and above. However, because the evaluation scale for third graders is a two-point scale, whereas it is a four-point scale for fourth graders and above, we exclude students in the third grade or below to ensure consistency. Furthermore, to investigate the longitudinal association between prior-year performance and current-year absenteeism, we used test scores (Japanese and mathematics) and noncognitive evaluations from the preceding year as explanatory variables. Consequently, the final analytical sample is limited to students ranging from the fifth grade of elementary school to the second grade of junior high school between 2019 and 2023.

Regarding inconsistencies in school information due to varying survey timings across datasets, we prioritize school-administered records, specifically the Amakko Survey, height and weight data, and Long-term Absenteeism Records. For observations with remaining missing values, we supplement the information using the Basic Resident Register, which serves as the primary administrative record.[7] Similarly, data concerning class assignments are primarily based on the Amakko Survey, supplemented by the Long-term Absenteeism Records for missing observations.[8]

It is necessary to address the characteristics of long-term absentees in our analytical sample and the limitations of the sample selection. As previously mentioned, to dynamically analyze the impact of academic achievement, non-cognitive skills, and family attributes on long-term absenteeism, a student had to be classified as a "long-term absentee" in a given year and must have participated in the Amakko Survey in the preceding year to provide baseline measurements. Consequently, not all long-term absentees recorded in the administrative data could be included in the analysis. Specifically, although the administrative data yielded 7,353 long-term absentees (person-years), only 1,699 (approximately 23%) matched the prior-year survey results. This discrepancy is likely because many long-term absentees were absent on the day the survey was administered. Therefore, our primary analytical sample is restricted to students who were able to attend school and respond to the survey in the previous year, potentially introducing a selection bias toward a group with relatively less severe

[7] The Amakko Survey contains a significant number of missing school codes (29,995 observations). Furthermore, the school districts indicated in the provided data often reflect the student's current district rather than their district at the time of the survey; for example, current junior high school students were listed as belonging to their junior high district even for their elementary school years (33,690 observations). To address these issues, we primarily use height and weight data for imputation, supplemented by 184 observations from the Long-term Absenteeism Records and 247 observations from the Basic Resident Register.

[8] Class-level data are primarily derived from the Amakko Survey. For instances where class information was absent in the Amakko Survey but available in the Long-term Absenteeism Records (5,672 observations), the latter was utilized.

absenteeism.

To evaluate the effect of this bias, we adopt two analytical approaches. First, we compare the 1,699 students who participated in the prior-year survey and became long-term absentees in the current year with those who participated in the prior-year survey and did not become long-term absentees in the current year Second, as a complementary analysis, we utilize a comprehensive dataset covering all 7,353 long-term absentees, regardless of whether they had participated in the prior-year survey. For this broader sample, the explanatory variables were limited to administrative data that did not rely on the survey such as public assistance, gender, single-parent household status, and sibling composition. By comparing the results of this complementary analysis (Section 5.4, Table 4) with our main findings, we clarify the positioning of our primary sample within the entire population of long-term absentees and verify the validity of our estimation results.

3.2 Definition of Variables

This section details the definitions and construction of the variables used in the analysis. The dependent variable, the long-term absenteeism dummy, is based on the Long-term Absenteeism Records provided by Amagasaki City. Although the official definition by MEXT identifies long-term absentees as those missing 30 or more days within an academic year, nationwide school closures implemented in early March 2020 due to COVID-19 make it difficult to maintain a consistent comparison across the pre- and post-pandemic periods using this standard definition. To ensure an equitable evaluation across periods, we define the dummy variable as 1 if a student was absent for 30 days or more during the 11-month period from April to February (excluding March), and 0 otherwise.[9]

The primary explanatory variables, academic achievement and non-cognitive skills, are derived from the Amakko Survey. For academic achievement, we use Japanese and mathematics test scores. For noncognitive skills, we employ scores for each of the Big Five personality traits: extraversion, agreeableness,[10] conscientiousness, emotional stability (neuroticism), and openness to experience.[11] To facilitate the interpretation of the estimated coefficients, these variables are first standardized as T-

[9] As a supplementary validation, we confirm that the qualitative conclusions of our estimation results remain unchanged even when using full-year attendance data, including the month of March.

[10] One potential factor contributing to long-term absenteeism is absence of friendships. To control for the influence of this factor, the analysis employs the response to the question item “I have few friends I can rely on” from the Amakko Survey as an indicator of agreeableness (cooperativeness) and incorporates it as a control variable.

[11] The Amakko Survey adopted the Japanese version of the Ten Item Personality Inventory (TIPI-J), standardized by Oshio et al. (2012), to measure children’s personality traits (the Big Five). This scale was developed based on the original short-form TIPI created by Gosling et al. (2003) and offers the advantage of efficiently measuring five factors—Extraversion, Agreeableness, Conscientiousness, Emotional Stability (Neuroticism), and Openness to Experience—using only ten items. Regarding its validity and reliability, Oshio et al. (2012) and Oshio et al. (2014) confirmed sufficient correlations between the TIPI-J and existing long-form scales, such as the Big Five Inventory. Based on these prior studies, the author concluded that the TIPI-J is a valid indicator of non-cognitive skills for large-scale surveys.

scores (with a mean of 50 and standard deviation of 10) for each academic year and grade. They are then divided by 10 and included in the model as lagged variables from the previous year.

Student and household attributes are constructed by integrating multiple administrative datasets. From the School Expense Subsidies data, we create dummy variables indicating whether a student belongs to a household receiving either public assistance or quasi-public assistance. Gender (female dummy) was identified using both the Amakko Survey and Long-term Absenteeism Records. Additionally, a single-parent household dummy variable is derived from the Basic Resident Register. To account for the school environment, class size is calculated for each grade within each school using the enrollment and class counts provided in the Student Enrollment Records.

Furthermore, for the robustness checks presented in Section 5.3, we use additional variables related to family structure and lifestyle habits. Using the Basic Resident Register, we focus on students identified as the child, grandchild, or stepchild of the household head (comprising 99.79% of the total sample) to construct a first-born dummy, the age of the youngest child, the number of siblings, and a dummy variable indicating whether any siblings were also long-term absentees. Regarding lifestyle indicators, the Amakko Survey provide prior-year data on the number of books available at home, daily television viewing time, study time, and gaming time. From the height and weight data, we construct a variable representing the student's obesity status in the preceding year.[12] Following Chetty et al. (2011), classroom environment variables are also included as explanatory variables. Sample sizes vary slightly across these variables owing to non-responses to certain survey items. A comprehensive list of all variables used in this analysis, including their names, data sources, and detailed definitions, is provided in the Appendix.

4. Estimation Strategy

We use OLS to empirically examine the relationship between long-term absenteeism and individual characteristics, based on equation (1):[13]

[12] Obesity status is determined using the Rohrer Index—a standard physical measure for school-aged children calculated as 10 × (weight in kg)/(height in m)^3. We define dummy variables for being overweight (Rohrer Index between 145 and 160) and obese (Rohrer Index of 160 or higher).

[13] The primary objective of this study is not to predict the probability of long-term absenteeism, but to elucidate the marginal effects of various factors—such as students' personal characteristics, family environments, and lifestyle habits—on long-term absenteeism. Therefore, Ordinary Least Squares (OLS), functioning as a Linear Probability Model (LPM), was adopted as the estimation method, as it allows the estimated coefficients to be interpreted directly as average marginal effects. Although OLS estimation for binary outcomes has the limitation that predicted values may fall outside the [0, 1] interval, the LPM is known to provide sufficiently robust and easily interpretable results when the main interest lies in the marginal effects of multiple explanatory variables, as is the case in this study. In addition, the key independent variables exhibit limited within-individual variation over the five-year period, making fixed-effects or first-difference estimates inefficient and statistically insignificant.

$$Outcome_{it} = \alpha X_{1,it} + \beta X_{2,i} + \gamma C_{it} + \mu_s + \mu_g + \mu_t + \varepsilon_{it}, \quad (1)$$

where $Outcome_{it}$ is a dummy variable indicating whether student i is classified as a long-term absentee in year t; $X_{1,it}$ includes the lagged standardized IRT-based academic scores (Japanese language and mathematics) and lagged standardized Big Five non-cognitive skill measures, as well as dummy variables for quasi-public assistance, public assistance, and single-parent households; $X_{2,i}$ is a gender dummy; C_{it} denotes class size in year t; $\mu_s, \mu_g, and\ \mu_t$ represent school, grade, and year fixed effects, respectively; and ε_{it} is an error term. We conduct the analysis using three samples: the pooled sample of students from grades 5 through 8, the elementary school sample (grades 5 and 6) and the junior high school sample (grades 7 and 8).

5. Results and Discussion

5.1. Descriptive Statistics

Table 1 presents the summary statistics for each variable categorized by school level (elementary and junior high school) and long-term absenteeism status. The rate of long-term absenteeism was 1.34% (426 of 31,904) among elementary school students and 2.40% (657 of 27,406) among junior high school students. This indicates that the probability of absenteeism tends to be higher at the junior high school level, which is consistent with the literature (e.g., Aoki et al. 2020).

A comparison of the attributes of long-term absentees and non-absentees revealed several key trends. First, regarding academic achievement and the Big Five non-cognitive skills, the long-term absentee group generally exhibited lower mean scores than the non-absentee group across all indicators. Second, in terms of gender, although the mean absenteeism probability for female students was slightly higher than that for male students in this sample, the difference was not statistically significant. By contrast, marked differences were observed in household economic conditions and structures. Specifically, students from households receiving quasi-public or public assistance as well as those from single-parent households had a significantly higher probability of long-term absenteeism than their peers from more advantaged backgrounds. Furthermore, the data show that having a sibling who is also a long-term absentee is positively correlated with a student's likelihood of being absent, suggesting a strong correlation with absentee behavior within the household.

5.2. Baseline Results

Table 2 (Columns 1–3) and Figure 1, which visualize the estimated coefficients, reveal the following relationships between academic achievement and long-term absenteeism.

First, regarding Japanese language scores, a statistically significant negative association was observed only in the junior high school sample, whereas no significant relationship was found for the overall or elementary school samples. In contrast, arithmetic and mathematics scores showed a significant negative correlation across all samples, indicating a strong link between low mathematical achievement and long-term absenteeism. As this analysis relies on Ordinary Least Squares (OLS), we cannot entirely rule out the possibility of reverse causality—where long-term absenteeism further exacerbates academic decline. However, our model incorporates lagged test scores from the previous period as explanatory variables, thereby accounting for temporal precedence to some extent. In addition, mathematics is a cumulative subject, it is plausible that prior academic struggles diminish subsequent learning motivation and trigger school avoidance. Indeed, previous studies (e.g., Rumberger 1995; Gubbels 2019; Balkis et al. 2016; Soma et al. 2025) consistently identify low academic achievement as a direct predictor of long-term absenteeism and school dropout. Consequently, our results are consistent with these findings, suggesting that academic underperformance represents a key pathway leading to long-term absenteeism.

The relationship between non-cognitive skills and long-term absenteeism varied depending on the specific trait. Extraversion, agreeableness, conscientiousness, and emotional stability were negatively correlated with the probability of long-term absenteeism. This finding is consistent with Lounsbury et al. (2004), who analyzed the correlation between personality traits and the number of absent days. Furthermore, considering that Chetty et al. (2011) demonstrated that improvements in non-cognitive skills lead to higher college attendance and increased future earnings, our results suggest that long-term absenteeism may have a negative impact on students' long-term outcomes. Conversely, openness to experience, which reflects curiosity and intellectual inquisitiveness, positively correlated with the probability of long-term absenteeism among junior high school students, although conscientiousness and openness were not statistically significant among elementary school students. This positive relationship suggests that, in the Japanese educational system, which often prioritizes cooperation and conformity, students with high openness may not be adequately valued, potentially leading to a mismatch with the school environment.[14]

Regarding student and household attributes, being from a single-parent household or receiving public assistance was positively correlated with the probability of long-term absenteeism. Notably, the

[14] One possible reason for the discrepancy in openness between elementary and junior high school students is that the wording and content of the questions defining "openness" differ between the two groups (see the Appendix for details), potentially capturing different traits despite using the same index name. Furthermore, while many elementary schools in Amagasaki allow students to wear casual clothes, junior high schools generally mandate uniforms. Given that extant research suggests that uniforms can intensify peer pressure for conformity within schools (Lee et al. 2021, 2024), the junior high school environment may pose greater challenges for students with high openness.

magnitude of the impact on public assistance is largest. In terms of gender, the coefficient on the female dummy variable is positive and statistically significant only in the junior high school sample. These findings are consistent with the literature and reaffirm that the socioeconomic and environmental conditions surrounding a child play a decisive role in long-term absenteeism.

The results show that an increase in class size significantly increased the probability of long-term absenteeism only in the elementary school sample. By contrast, no statistically significant correlation was observed among junior high school students. While the research design of this study differs from that of Nakamuro (2017), who examines the relationship between school refusal and class size, our findings are similar. This suggests that implementing smaller class sizes, particularly during the early to middle stages of elementary education, may contribute to the prevention and reduction of long-term absenteeism.

Finally, we discuss the estimated coefficients on the dummy variables for each year. Taking the 2019 academic year as the baseline, the coefficients for 2020 are significantly negative across the sample. This result likely reflects structural factors, specifically the substantial reduction in total school days due to nationwide school closures implemented during the initial COVID-19 outbreak. For the 2021 academic year, the coefficient remained significantly negative for the overall sample, although no statistically significant differences were observed when the sample was divided by school level. Notably, the coefficients for the 2023 academic year were significantly positive for each sample. This indicates that, even after controlling for individual and household attributes, the probability of long-term absenteeism has increased in recent years. The factors driving this surge in absenteeism, which has become particularly pronounced in the post-pandemic period, are examined in greater detail in Section 5.6.

5.3. Robustness Checks

In this section, we conduct robustness checks of the previous estimation results by restricting the sample to new cases of long-term absenteeism and by adding explanatory variables for individual characteristics, household environment, and classroom environment.

While the analysis in Section 5.2 controls academic achievement and non-cognitive skills from the previous year, we cannot rule out the possibility that these variables reflect the *consequences* (or precursors) of an already emerging trend of absenteeism rather than being its *causes*. Therefore, to ensure the robustness of our findings, we conducted a sub-sample analysis limited to students who had never experienced long-term absenteeism until the previous year. The objective of this analysis is to identify factors influencing the "pure onset of absenteeism" by entirely excluding the effects of past attendance habits. The results of this sub-sample analysis were qualitatively consistent with our main findings. This strongly suggests that the factors identified in this study are not merely maintaining existing absenteeism but are independent determinants that trigger the onset of new cases.

Columns (1)–(3) of Table 3 present the estimation results incorporating sibling attributes and lifestyle variables. The sibling attributes include the age of the youngest child, a first-born dummy variable (for those with at least one sibling), number of siblings, and a dummy variable indicating whether any siblings are currently long-term absentees. For clarity, the sibling's long-term absenteeism dummy is multiplied by 100 and the coefficient should be interpreted as 0.01 times the reported value.

In the models with these additional variables, the coefficients on the primary explanatory variables, including prior-year mathematics scores and the Big Five indices, generally tended to decrease in magnitude compared with the baseline model. For specific variables, a higher age of the youngest child and a larger number of siblings were both negatively correlated with the probability of long-term absenteeism. Most notably, the presence of a sibling who was a long-term absentee exhibited a strong positive correlation with students' own probability of being absent. This result suggests that, even after controlling for academic achievement and personality traits, unobserved home environment factors exert a strong influence or, alternatively, that there may be a spillover effect of absentee behavior within the household. Regarding time-use variables, we found a negative correlation between the prior-year television viewing time and the probability of long-term absenteeism, whereas gaming time was positively correlated with the likelihood of absence. Notably, even after controlling for these lifestyle variables, the correlations between academic indicators, Big Five traits, family environment variables, and the probability of long-term absenteeism remained robust.

Columns (4)–(6) of Table 3 present the estimation results after adding the prior-year obesity dummy as a physical indicator. Except for the elementary school sample, being prone to obesity showed a statistically significant positive correlation with the probability of long-term absenteeism. This suggests that the psychological stress stemming from physical characteristics or changes in the overall health status may influence absentee behavior.

Furthermore, we incorporate the number of books at home in the preceding year as a proxy for a household's cultural background (Table 3, Columns (7)–(9)). The results revealed a positive correlation between the number of books and the probability of long-term absenteeism in the samples that excluded elementary school students. However, these findings require careful interpretation. While a larger collection of books is generally expected to have a positive impact on academic achievement, the positive correlation observed here may reflect reverse causality; that is, parents might have intentionally purchased more books to encourage home learning as students' absenteeism increased.

Finally, this study examines the impact of the proportion of long-term absentees within a class and the quality of the classroom environment on individual-level absenteeism (Table 3, Columns (10)–(12)). The analysis reveals that the ratio of long-term absentees in a class does not have a statistically significant relationship with an individual's probability of long-term absence in any of the samples (total, elementary, and junior high school). Furthermore, using data from the Amakko Survey, class-level indicators are used as proxies for the classroom environment. These include the mean scores and standard deviations of Japanese and mathematics tests, as well as measures of interpersonal relationships and social norms within the class (e.g., apologizing or expressing gratitude). [15] The results show that the coefficient for the class average in mathematics is positive and significant, suggesting a tendency for long-term absenteeism to increase in classes with higher average math performance. By contrast, no significant correlation can be observed between interpersonal relationships or social norms within the class and long-term absenteeism.

5.4. Validation of Selection Bias

Our analysis thus far has focused on students who were in a state of long-term absenteeism during the current academic year and had participated in the "Amakko Survey" in the previous year. However, as previously noted, many long-term absentees were likely absent on the day the survey was administered, resulting in a subestof the population being excluded from the analytical sample. Consequently, our sample is biased toward students with relatively less severe absenteeism. This suggests that the estimated coefficients may be underestimated. To examine the presence of selection bias resulting from this sampling process, this section presents a validation analysis using a comprehensive sample of all long-term absentees identified through administrative data.

Specifically, we construct a dataset encompassing all students recorded as long-term absentees in a given year (totaling 7,353 person-years), regardless of whether they participated in the prior-year survey. Using this dataset, we conduct a regression analysis limited to explanatory variables available solely from administrative records, namely quasi-public assistance or public assistance status, gender, single-parent households, and sibling attributes, and compare these results with those from the main

[15] Specifically, we use indicators derived by averaging class-level responses to survey items regarding the classroom environment such as: "The class is an environment where students can sincerely say 'I'm sorry' and reconcile," "Students in this class express gratitude to one another," "Students do not say hurtful things or tease others," "Students can play or form groups with anyone," and "Students acknowledge each other's strengths and efforts." These subjective measures of classroom climate do not emerge as significant factors in explaining long-term absenteeism within the framework of our analysis. Due to space constraints, these specific results are not reported in the tables but are available upon request.

sample.

Columns (1)–(4) of Table 4 show the estimation results for the sample restricted to those who took the survey in the preceding year (including 1,699 long-term absentees), consistent with previous models. Columns (5) and (6) present the results for a comprehensive sample of 7,353 long-term absentees.

Comparing the estimates in Columns (1)–(4), the coefficients on variables such as public assistance and single-parent status increase in magnitude when academic achievement and non-cognitive skill variables are omitted. However, no inconsistencies in the signs of the coefficients exist, suggesting that the omission of these key variables does not fundamentally alter the results. Furthermore, a comparison of Column (3)–(6) reveals that, while some differences exist in the findings for class size and the proportion of long-term absentees in a class, the coefficients on quasi-public assistance, public assistance, single-parent households, and sibling attributes remain consistent in sign, despite variations in magnitude. These results suggest that, for the primary variables of interest, such as household attributes, the estimation does not suffer from sample selection bias.

5.5. Analysis by Reason for Absence

To better understand the diverse factors behind long-term absenteeism, we conduct an analysis based on the specific reasons for absence. The Long-term Absenteeism Records provide monthly absence counts for each student, along with the primary reason as determined by the school, classified into four categories: school refusal (*futoko*), family/personal circumstances (*kajitsugo*), others, and sickness (*byoketsu*). Because a single student may be absent for multiple reasons during the one-year measurement period, we aggregate the days of absence for each reason from April to February. The category with the highest proportion of days is defined as the student's primary reason for absence.[16] In the full sample, the reasons are as follows: school refusal (45.7%), sickness (38.7%), family/personal circumstances (14.6%), and others (1.05%).

The reasons for absence exhibit distinct characteristics depending on students' grades.[17] According to Figure 2, the proportion of absences due to school refusal tends to increase as students progress to higher grades. By contrast, the proportion of absences due to family or personal circumstances is higher in lower grades. Regarding sickness, although no remarkable shifts are observed across grades,

[16] In cases where multiple reasons account for the same proportion of absences (e.g., equal days for school refusal and family circumstances), they are categorized as follows: school refusal/family as school refusal, family/sickness as sickness, and school refusal-sickness as school refusal. However, such cases are extremely rare, totaling only 23 observations.

[17] No significant differences are observed in the composition of reasons for absence between male and female students.

the proportion is slightly higher among junior high school students than among elementary school students.

Furthermore, we examine year-to-year transitions in the reasons for absence among students whose long-term absenteeism persisted for two or more consecutive years using a transition matrix (Table 5). The analysis reveals that students absent due to illness or family reasons are more likely to transition to school refusal in the following year. This suggests that as absenteeism becomes prolonged, it tends to result in school refusal, irrespective of the original reason.

To further elucidate the heterogeneity underlying long-term absenteeism, we conduct separate estimations using three primary reasons—school refusal, family/personal circumstances, and sickness—as dependent variables. The category "others," which includes a very small number of observations, is excluded from this analysis. The sample comprises elementary and junior high school students. Based on the estimation results presented in Table 6, we identify the following distinctive findings for each reason.[18]

First, long-term absenteeism categorized as school refusal (*futoko*) exhibits a strong correlation with personality traits (non-cognitive skills). Specifically, extraversion, agreeableness, and openness are significantly associated with the probability of school refusal. By contrast, there is little to no observed correlation with economic factors, such as public assistance or single-parent status, or household environmental variables, such as the number of books at home. These findings suggest that school refusal is primarily driven by a mismatch between the student's characteristics and the school environment, rather than household economic conditions.

Second, regarding absenteeism due to family/personal circumstances (*kajitsugo*), the results stand in stark contrast to those for school refusal. Almost no correlation is observed with a student's academic achievement or personality traits; instead, we find an extremely strong correlation with economic factors, including public assistance and single-parent household status. The proportion of students from households on public assistance among those absent for family reasons reaches approximately 14%, which is strikingly high compared to school refusal (4.3%) and sickness (9.4%). These facts underscore the fact that absences due to family circumstances are rooted in household poverty and single-parent family structures rather than individual capabilities.

Third, for absences attributed to sickness, significant correlations are observed not only for academic and personality traits but also for sibling attributes and household environment variables. Specifically, the public assistance dummy for elementary students and the single-parent dummy for

[18] We conducted a similar analysis by splitting the sample according to the reason for the initial occurrence of long-term absenteeism. The results were substantially similar to those of the current analysis.

junior high students are significant, alongside the influence of sibling-related variables. Behind the formal reason for sickness, there is a high probability that multiple factors—including students' own health, economic hardship, and multi-layered domestic challenges—could be interacting. This highlights the difficulty of providing effective support for this group.

In summary, these results demonstrate that the primary determinants of long-term absenteeism differ depending on the underlying reason. While school refusal is heavily influenced by personality traits, absences due to family circumstances are likely to arise from economic constraints. Sickness-related absenteeism often results from accumulated hardships, where multiple disadvantages overlap.

5.6. Analysis of the Post-Pandemic Increase in Long-Term Absenteeism: Evidence from Blinder–Oaxaca Decomposition

Here, we conduct a detailed examination of the factors driving the surge in long-term absenteeism that has become prominent since the COVID-19 pandemic. According to the Survey on Issues Related to Student Guidance such as Problematic Behavior and School Refusal of Students by the MEXT, the national long-term absenteeism rate for elementary and junior high school students nearly doubled 1.88% in 2019 to 3.72% in 2023.[19] Considering this national trend of rising long-term absenteeism, this study utilizes panel data from a single municipality to analyze how the determinants of long-term absenteeism changed between the pre- and post-pandemic periods.

Specifically, we perform a Blinder–Oaxaca decomposition comparing two points in time: the pre-pandemic period (2019) and the post-pandemic period (2023). By employing this method, it is possible to differentiate and evaluate the increase in the long-term absenteeism rate into two distinct components: (1) changes in the distribution of student and household attributes (endowment effect or composition effect) and (2) changes in the magnitude of the impact these attributes have on the probability of long-term absenteeism (coefficient effect or structural effect). Through this approach, we elucidate the structural shifts in how the massive social shock of the pandemic altered the relationships between existing family environments, individual personality traits, and absentee behavior. The Blinder–Oaxaca decomposition partitions the difference (R) in the mean of the observed outcome (Y) into a component explained by differences in characteristics (X) and a component explained by differences in the contributions of those characteristics (β), thereby quantifying the contribution of each component. The estimation model is given by equation (2):

[19] Based on the survey materials for fiscal year 2023.

$$
\begin{aligned}
R &= E(Y_A) - E(Y_B) \\
&= E(X_A)'\beta_A - E(X_B)'\beta_B \\
&= [E(X_A) - E(X_B)]'\beta_B + E(X_B)'(\beta_A - \beta_B) + [E(X_A) - E(X_B)]'(\beta_A - \beta_B) \qquad (2)
\end{aligned}
$$

where A denotes the after period, corresponding to the post-pandemic year 2023, and B denotes the before period, corresponding to the pre-pandemic year 2019; Y indicates whether a student is a long-term absentee; X denotes observed characteristics; and β are the estimated coefficients. The explanatory variables used in the preceding sections provide the set of observed characteristics.[20]

Table 7 presents the results of the Blinder-Oaxaca decomposition, detailing the factors behind the changes in long-term absenteeism probabilities from 2019 to 2023. The upper panel shows the results for elementary school students, whereas the lower panel shows the results for junior high school students.

For elementary school students, the probability of long-term absenteeism increased by 1.295 percentage points, from 0.931% in 2019 to 2.226% in 2023. We decompose this total difference into three components: the endowment effect (attributable to changes in the distribution of student and household attributes), the coefficient effect (attributable to changes in the impact or weight of each attribute on the probability of absenteeism), and the interaction effect (the rest parts of the differences). The analysis revealed that these components contribute 26.39, 61.15, and 12.46%, respectively. This indicates that most of the increase is explained by the coefficient effect.

A similar trend was observed among junior high school students. The probability of long-term absenteeism increased by 1.514 percentage points, from 1.811% in 2019 to 3.325% in 2023. When decomposed, the contribution of the endowment effect is 19.07%, whereas the coefficient effect reaches 75.69% (with an interaction effect of 5.24%). A major characteristic of this analysis is that, for both the elementary and junior high school samples, the structural component (coefficient effect) accounts for an overwhelmingly high proportion of the increase in absenteeism. These findings strongly suggest that the post-pandemic surge in long-term absenteeism was not primarily driven by shifts in the composition of student or household attributes, such as a sudden increase in poverty or changes in personality traits. Rather, it reflects a structural transformation, where students with certain existing attributes have become more susceptible to long-term absenteeism or found it more difficult to adapt to school life under the same environmental conditions.

A breakdown of the coefficient effect listed in the share column reveals a marked increase in the

[20] Because the analysis focuses on changes across time points, year fixed effects are excluded from the set of explanatory variables.

influence (share) of specific variable groups on the rising probability of absenteeism. Specifically, the impact of academic scores on Japanese and mathematics has intensified significantly. Furthermore, regarding non-cognitive skills, the magnitude of the coefficients for conscientiousness in elementary school students and for extraversion, emotional stability, and openness in junior high school students increased following the pandemic. This suggests that students with lower academic performance or specific personality profiles are in a more vulnerable position, making them more likely to resort to long-term absenteeism after the environmental shifts brought about by COVID-19.

Changes are also evident in family and educational environments. In the elementary school sample, the influence of class size increased, suggesting that crowded learning environments may have exacerbated the risk of absenteeism in the post-pandemic era. Similarly, the influence of sibling attributes, such as the age of the youngest child and the total number of siblings, grew, indicating that multilayered household factors play a more significant role in absenteeism. By contrast, while economic attributes such as public assistance status and single-parent status remain powerful determinants of long-term absenteeism, the magnitude of their specific coefficients do not show significant changes between the pre- and post-pandemic periods.

However, within the structural component (the coefficient effect), the constant term accounts for the largest share. This implies that, even with an extensive set of independent variables incorporated into this model, the influence of broad structural shifts common to society as a whole, which remain uncaptured by the observed indicators, is exceptionally large. The unobserved environmental changes captured by this constant term primarily include the following three points

First, the transformation of the digital environment has profoundly impacted students' daily lives and learning. The widespread distribution of devices under the GIGA School Initiative has normalized home-based device usage. This shift likely triggered prolonged nighttime use, which disrupted daily routines through sleep deprivation and difficulty waking up. Such disruptions subsequently undermined student concentration and exacerbated learning gaps during classes. Furthermore, students may develop feelings of inferiority if they cannot master digital devices or if they fall behind in digitalized lessons. These factors can diminish self-efficacy and ultimately induce school avoidance. Consequently, such maladjustment and academic underperformance associated with device use may have hindered student participation in in-person school activities by increasing both psychological and physical resistance.

Second, the pandemic has fundamentally altered the perception of schooling. Experiences such as school closures and staggered attendance have weakened the social norm that "attending school every day is mandatory." This macro-level shift in mindset likely affected all households regardless of their specific attributes, creating a social environment where absenteeism is more socially acceptable.

Third, there has been a significant change in the home-based supervision environment. The widespread adoption of telecommuting has allowed more parents to stay home and care for their children. This change likely made it easier for parents to decide to keep their children at home if they could supervise them, even in situations where they might have previously forced them to attend school. Consequently, this has likely lowered the psychological barriers to choosing absenteeism.

From these perspectives, the large constant term does not indicate a flaw in our model. Rather, it suggests that the mindset and environment surrounding school attendance have undergone a fundamental transformation. This transformation is so profound that existing data alone cannot fully explain its extent.

It should be noted that the results in Table 7 are limited to long-term absentees who were able to respond to the questionnaire (n = 1,699), suggesting a potential selection bias toward students with relatively mild absenteeism. To address this concern, we re-conducted a two-period decomposition using the full sample (n = 7,353), including those who did not take the survey, with a limited set of explanatory variables (similar to the approach in Section 5.4).The results revealed that while the contribution of household environment variables increased to some extent in the full-sample analysis, the primary structure—where the constant term (structural component) provides the largest contribution—remained entirely consistent with the results from the restricted sample. In conclusion, these findings demonstrate that macro-level social transformations, which are difficult to quantify, are the primary driver of rising absenteeism rates regardless of severity. Thus, the findings of this study possess high generalizability and robustness for the entire population of long-term absentees.

In summary, the results of the decomposition analysis indicate that the surge in long-term absenteeism during the COVID-19 pandemic was not driven by changes in the distribution of students or household attributes (the endowment effect). Instead, it primarily stems from the amplified impact of academic achievement, non-cognitive skills, family environment, and class size on the probability of absenteeism (the coefficient effect). Nevertheless, beyond the shifts in these individual variables, the impact of unobservable environmental and social structural changes, represented by the large contribution of the constant term, cannot be ignored when explaining recent absentee behavior.

5.7. Discussion

The primary findings of this study can be summarized as follows. First, even when controlling for other factors, household environments—specifically, single-parent status and receipt of public assistance, as well as lower prior-year mathematics test scores—exhibited a significant positive correlation with the probability of long-term absenteeism. Notably, the magnitude of the impact of being from a household receiving public assistance was exceptionally large. Regarding non-cognitive skills (Big Five), students with lower levels of extraversion, agreeableness, conscientiousness, and emotional stability were more likely to experience long-term absenteeism. By contrast, higher levels of openness to experience were associated with an increased probability of absence.

Even after controlling these attributes, the probability of long-term absenteeism showed an increasing trend. An examination of detailed household attributes revealed that a younger age for the youngest child, fewer siblings, and the presence of a sibling who is also a long-term absentee significantly increased a student's likelihood of absence. Obesity was positively correlated with absenteeism. Moreover, the underlying drivers of long-term absenteeism differed based on the reasons for absence. The decomposition analysis comparing the pre- and post-pandemic periods further indicated that the primary driver of the recent surge in absenteeism was not a shift in the distribution of student attributes but rather a transformation in the magnitude of influence (the coefficient effect) that these attributes exert on absentee behavior.

According to the Survey on Issues Related to Student Guidance, such as Problematic Behavior and School Non-attendance of Students conducted by the MEXT, the increase in long-term absenteeism during the COVID-19 pandemic has been driven by both "sickness" and "school non-attendance." The rise in absenteeism due to sickness is primarily attributed to the onset of orthostatic dysregulation caused by disrupted daily routines, as well as the manifestation of developmental disorder characteristics triggered by environmental changes. Furthermore, the social acceptance of infection avoidance and minor physical symptoms (e.g., mild fever or cough) as legitimate reasons for absence likely decreased the psychological barriers to missing school. School non-attendance is increasingly driven by academic underachievement and disrupted daily routines. The disruption of learning continuity due to school closures and the disparity in non-cognitive skills under ICT-based environments may have played a significant role.

We examine the relationship between these background factors based on the results of the Blinder–Oaxaca decomposition analysis. One possible reason why the influence of academic performance and non-cognitive skills on long-term absenteeism has strengthened since the COVID-19 pandemic is the transformation of the social norms regarding school attendance. Previously, even when academic performance was poor, attendance was maintained by a strong normative sense that "attending school is a natural obligation." However, this norm was relaxed during the pandemic. Consequently, for students who feel a sense of inferiority or who experience difficulties in group life due to insufficient non-cognitive skills, the psychological barrier to choosing absence as an avoidance behavior may have been significantly lowered.

However, a positive correlation was estimated between “openness” and long-term absenteeism. Two hypotheses can be considered for this result. The first is that students with high levels of openness have high intellectual autonomy and might have proactively chosen learning methods or locations outside of school that offer higher educational returns. Indeed, Fukura and Nakachi (2022) examined the relationship between non-attendance and personality traits among university students. Their results show that students who possess high openness combined with a sense of superiority tend to have stronger emotions associated with school avoidance. In contrast, students who exhibit high openness but lack proactive behavior show a positive correlation with actual school avoidance. The second hypothesis is that long-term absenteeism occurs because students with high intellectual curiosity and individuality fail to adapt to the Japanese school system's emphasis on conformity. Such students may experience a breakdown in relationships with teachers, further distancing them from school.

To test the second hypothesis (the maladjustment hypothesis), responses to the items “I enjoy going to school” and “I feel that my teachers have recognized and valued me” were extracted from the Amakko Survey and used as dependent variables in a model estimated using the same specifications as those in equation (1) (Figures 3 and 4). The results show that the coefficient on openness is positive and statistically significant, indicating that students with higher openness tend to perceive school life more positively and feel recognized by their teachers. This contradicts the second hypothesis that high openness causes a mismatch with the school environment; instead, it supports the first hypothesis. However, the interpretation of our findings requires caution because the dependent variables—school enjoyment and teacher recognition—were measured in the same academic year as the long-term absenteeism. Specifically, we cannot entirely rule out the possibility of reverse causality (endogeneity). It remains unclear whether high openness leads to positive perceptions of school, or whether the absence of absenteeism itself fosters better school perceptions. Therefore, these results merely highlight one aspect of a correlation: that high openness does not necessarily lead to maladjustment at school. Identifying a direct causal relationship requires a more cautious interpretation.

Furthermore, the reason why the impact of class size has intensified only among elementary school students may be attributed to the increased workload of teachers. Since the COVID-19 pandemic, the tasks required of homeroom teachers have multiplied, including disinfection, temperature checks, supervision of silent dining (mokushoku), and the management of digital devices under the GIGA School Initiative. These burdens tend to be concentrated on elementary school teachers, who operate under a homeroom teacher system, in contrast to junior high schools, which utilize a subject-specific teacher system. Moreover, according to MEXT’s Survey on Students with Special Educational Needs in Regular Classrooms (2022), the percentage of students in regular classes who exhibit significant difficulties in learning or behavior has reached 10.4% in elementary schools and 5.6% in junior high schools. This increase requires additional support and is another factor depriving teachers of the capacity to attend to individual needs. Consequently, this may have exacerbated the influence of class size on student absenteeism.

The fact that the determinants of long-term absenteeism differ depending on the underlying reasons has significant policy implications. The estimation results by reason in Table 6 demonstrate that absenteeism due to family/personal circumstances is strongly influenced by the household's economic factors rather than the student's individual traits. Moreover, the transition probabilities (transition matrix) in Table 5 indicate a high probability that students who were absent due to family circumstances in the previous year will transition to school refusal (*futoko*) in the following year. Although absenteeism for family reasons accounts for only 14.6% of all cases, Figure 2 shows that it is particularly prevalent in the lower grades of elementary school (Grades 1–3). Altogether, these findings highlight that providing economic support to parents from an earlier stage, specifically since the lower elementary grades, serves as a critical intervention to prevent subsequent school non-attendance. Finally, considering that family environmental factors were significant for absenteeism due to illness and that the impact of family background was even more robust in the full sample analysis (Table 4, Columns (5) and (6)), the negative impact of home environment on absenteeism should be regarded as more severe than previously analyzed.

6. Conclusions

This study utilized a unique panel dataset, created by merging administrative microdata in Amagasaki City, Hyogo Prefecture, from 2018 to 2023 and analyzed the determinants of long-term absence and structural changes since the COVID-19 pandemic. The starting point was to elucidate how students' personal attributes, family economic conditions, and external shocks, such as the COVID-19 pandemic, interacted to influence the recent alarming surge in long-term absence rates.

The analysis revealed that academic achievement and non-cognitive skills are distinctly associated with long-term absenteeism. In particular, low scores in mathematics showed a stronger correlation with absence risk compared to Japanese language scores, suggesting that addressing learning delays in cumulative subjects is key to preventing absence. Regarding non-cognitive skills, while low levels of extraversion, agreeableness, conscientiousness, emotional stability encouraged absence, this study presented a paradox: students with high openness are exposed to higher absence risks. Furthermore, by categorizing long-term absence by reason, this study highlighted that, while school refusal (*futoko*) is primarily driven by individual personality traits, absence due to family/personal circumstances is triggered by economic hardship (e.g., welfare receipt or single-parent status) The fact that the absence of family reasons in early elementary grades serves as a leading indicator of subsequent school refusal suggests that early economic intervention is an extremely effective countermeasure against non-attendance.

The primary contribution of this paper lies in its use of the Blinder–Oaxaca decomposition to identify that the post-pandemic surge in long-term absence was not caused by changes in the distribution of student attributes, but rather by a structural change characterized by an increased impact of specific attributes on absence. This implies that the drastic environmental changes brought about by the pandemic have exacerbated the vulnerabilities of students struggling with academic underachievement, certain personality traits, and challenging family environments, thereby increasing their adaptation difficulties. Simultaneously, the significant contribution of the constant term to the model underscores the substantial impact of macro-social transformations beyond individual attributes.

The findings of this study offer four specific guidelines for restructuring current countermeasures against long-term absenteeism. First, it is essential to drastically accelerate the timing of interventions. As shown in Figure 2, the issue of long-term absenteeism is already prominent in the first grade of elementary school, which supports the importance of early intervention pointed out by Gottfried (2014) and Soma et al. (2025). Therefore, the current system—which relies heavily on responses in upper grades or junior high school—must be reformed. It is indispensable to establish an early warning system starting with school entry health checkups and attendance monitoring immediately after enrollment. Strengthening proactive ("push-type") support for lower-grade students is vital to prevent future academic underperformance and school dropout.

Second, it is crucial to adopt a learning-oriented approach, specifically through support tailored to arithmetic and mathematics. The strong negative correlation between mathematics scores and the risk of absenteeism suggests that, due to the subject's cumulative nature, even minor struggles can trigger a significant loss of motivation, potentially leading to school avoidance. Therefore, establishing systems to address learning gaps in mathematics is more than just a means of academic improvement. These initiatives—which could include after-school programs or ICT-based remedial tools—serve as extremely effective interventions for both preventing absenteeism and supporting students' return to school.

Third, we must develop targeted interventions based on the specific reasons and backgrounds behind each student's absenteeism. For students whose absences are primarily due to domestic or family circumstances, priority should be given to economic support, such as expanding and promoting school expense subsidy programs. In contrast, when absences are categorized as school refusal or due to illness, personality traits and academic struggles are often complexly intertwined. In such cases, it is necessary to provide a combined support package: psychological care provided by school counselors (SC) alongside the guaranteed availability of diverse learning environments, such as free schools.

Finally, long-term absenteeism is driven not only by individual student traits but also by complex, intertwined structural household challenges. These include economic hardship in single-parent households, the responsibilities of young carers, and irregular daily routines. Addressing these issues requires a framework that extends beyond the school's traditional boundaries. Specifically, it is necessary to establish a multidisciplinary outreach system, centered on school social workers (SSWs) and in collaboration with welfare and labor departments. Comprehensive and individualized support that includes not only lifestyle counseling but also employment assistance for parents provided by dedicated professionals is essential for these families. With these supports, we can substantially safeguard educational opportunities for children facing difficult circumstances.

Future research should clarify the factors contained within the constant term that were not captured by this model, namely organizational factors, such as the quality of interpersonal relationships within the classroom and school culture. To further verify the effectiveness of economic support in early grades, as suggested in this study, longitudinal follow-up studies including younger children and multilayered analyses incorporating class size and teacher characteristics are required. We hope that the findings of this study will contribute to the development of a more flexible and individually optimized educational support system.

Acknowledgements

This study is part of a project conducted by the Amagasaki City Institute for Learning and Development. We are deeply grateful to the members of the Institute for their invaluable advice and support throughout this research. We would also like to thank Eiji Yamamura (Seinan Gakuin University), Tomohiko Inui (Gakushuin University), Jun Iritani (Kobe University), as well as the participants at the 18th Annual Conference of the Association of Behavioral Economics and Finance and the 2025 Spring Meeting of the Japan Economic Association, for their insightful comments. This work was supported by JSPS KAKENHI Grant Numbers JP25H00388, JP23K25529 and Saga University.

Data Availability Statement

The data used in this study were obtained from Amagasaki City, Japan under a data-use agreement. The dataset contains sensitive personal information on welfare-receiving households and their children that cannot be made publicly available owing to legal and ethical restrictions under Japan's Act on the Protection of Personal Information. Data access requires formal approval from Amagasaki City and is subject to Japanese-language communication and institutional review procedures. Researchers interested in accessing data or learning about the application process should contact the corresponding author, who can facilitate communication with the municipal government. The statistical code used in the analysis is available from the corresponding author upon request.

References

Aoki Eiichi, Ken Yusa, and Taketoshi Goto (2020) “Secondary analysis of individual-level data from MEXT’s Survey on Issues Related to Student Guidance, such as Problematic Behavior and School Non-attendance”, *Annual Reports of the Graduate School of Education, Tohoku University* (in Japanese).

Aoki Eiichi, Ken Yusa, and Taketoshi Goto (2021) “Fluctuations in school non-attendance due to the expansion of class size upon promotion: Secondary analysis of individual-level data from MEXT’s Survey on Issues Related to Student Guidance and School Basic Survey”, *Annual Reports of the Graduate School of Education, Tohoku University* (in Japanese).

Amagasaki City (2025) “Population of Amagasaki City”, https://www.city.amagasaki.hyogo.jp/_res/projects/default_project/_page_/001/041/163/25.4.pdf (in Japanese).

Aucejo Esteban M., and Teresa Foy Romano (2016) “Assessing the effect of school days and absences on test score performance”, *Economics of Education Review*, 55, 70–87.

Australian Curriculum, “Assessment and Reporting Authority (2025) National Report on Schooling in Australia 2025”, https://www.acara.edu.au/reporting/national-report-on-schooling-in-australia/student-attendance (accessed on 26 December 2025).

Balkis Murat, Arslan Gökmen, and Duru Erdinç (2016) “The School Absenteeism among High School Students: Contributing Factor”, *Educational Sciences: Theory and Practice*, 16(6), 1819-1831.

Blinder Alan (1973) “Wage discrimination: Reduced form and structural estimates”, *The Journal of Human Resources,* 8(4), 436–455.

Blanden Jo, Matthias Doepke, and Jan Stuhler (2023) “Educational inequality”, *Handbook of the Economics of Education*, 6, 405–497.

Cattan Sarah, Daniel A. Kamhöfer, Martin Karlsson, and Therese Nilsson (2023) “The long-term effects of student absence: Evidence from Sweden”, *The Economic Journal*, 133(650), 888–903.

CBC News (2025) “Reported illnesses have tripled in some school districts. Experts say mental health is a factor”, https://www.cbc.ca/news/canada/school-absences-illness-mental-health-9.6988661 (accessed on 26 December 2025).

Chang Hedy N., and Mariajose Romero (2008) “The Critical Importance of Addressing Chronic Absence in the Early Grades”, National Center for Children in Poverty

Chetty Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan (2011) “How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star”, *The Quarterly Journal of Economics*, 126(4), 1593–1660.

Deming J. David (2017) “The growing importance of social skills in the labor market”, *The Quarterly Journal of Economics*, 132(4), 1593–1640.

Department for Education England (2025) “Autumn and spring term 2024/25 Pupil absence in schools in England”, https://explore-education-statistics.service.gov.uk/find-statistics/pupil-absence-in-schools-in-england/2024-25-autumn-and-spring-term (accessed on 26 December 2025).

Edin Per-Anders, Peter Fredriksson, Martin Nybom, and Björn Öckert (2022) “The rising return to noncognitive skill”, *American Economic Journal: Applied Economics*, 14 (2), 78–100.

Fukura Suguru and Nobuo Nakachi (2022) “Relationships between personality traits, general self-efficacy, and tendencies toward school refusal: A study of university students”, *Tezukayama University Bulletin of Psychological Science*, (5), 45-52.

Gennetian A. Lisa, Christopher Rodrigues, Heather D. Hill, and Pamela A. Morris (2018) “Income level and volatility by children’s race and Hispanic ethnicity”, *Journal of Marriage and Family*, 81(1), 204–229.

Goodman Joshua (2014) “Flaking out: Student absences and snow days as disruptions of instructional time”, *Working paper no. w20221, National Bureau of Economic Research.*

Gosling Samuel D, Peter J. Rentfrow, and William B. Swann Jr. (2003) “A very brief measure of the Big-Five personality domains”, *Journal of Research in Personality*, 37, 504–528.

Gottfried Michael 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.

Gottfried Michael A. (2011) “The Detrimental Effects of Missing School: Evidence from Urban Siblings”, *American Educational Research Journal,* 117 (2), 147–182.

Gottfried Michael A. (2014) “Chronic Absenteeism and Its Effects on Students’Academic and Socioemotional Outcomes”, *Journal of Education for Students Placed at Risk (JESPAR)*, 19(2), 53-75.

Gubbels Jeanne, Claudia E. van der Put, and Mark Assink (2019) “Risk factors for school absenteeism and dropout: A meta-analytic review”, *Journal of Youth and Adolescence*, 48(9), 1637–1667.

Higeta Takeshi and Kaori Suetomi (2013) “The impact of school education policies and faculty policies on long-term absenteeism among junior high school students”, *Bulletin of the National Institute for Educational Policy Research*, 142 (in Japanese).

Hosaka Toru (1996) “A follow-up study of long-term absenteeism and school non-attendance”, *The Japanese Journal of Educational Psychology* (in Japanese).

Hosaka Toru (2009) “Absence of “school-avoiding” students and teachers’ leave of absence”, *Gakuji*

Shuppan (in Japanese).

Hosaka Toru and Shige Ayumi (2023) "Has long-term absenteeism (school refusal) increased? An analysis of the "Survey on issues related to student guidance, such as problematic behavior and school refusal of students", *Chiba University Journal of Educational Practice* (in Japanese).

Hosaka Toru (2024) *School, Japanese society, and the act of "resting": From the "problem of school refusal" to "work-style reform"*, University of Tokyo Press (in Japanese).

Inoue Atsushi and Tanaka Ryuichi (2024) "The rank of socioeconomic status within a class and the incidence of school bullying and school absence", *Economics of Education Review*, 101, 102545.

Ito Hiroyuki, Hamada Megumi, Murayama Yasuro, Takayanagi Shinya, Nomura Kazuyo, Myogan Mitsunobu, and Tsujii Masatsugu (2017) "Causal relationships between class size, academic achievement, and emotional/behavioral problems: An investigation using a natural experiment design and multilevel modeling", *The Japanese Journal of Educational Psychology* (in Japanese).

Janssen Ian, Wendy M. Craig, William F. Boyce, and William Pickett (2004) "Associations between overweight and obesity with bullying behaviors in school-aged children", *Pediatrics,* 113, 1187–1194

Kajiwara Takehito (2021) "Why are children from poor families more likely to be absent from school? An empirical study using a school refusal generation model", *The Journal of Educational Sociology* (in Japanese).

Lee Sun Youn and Ohtake Fumio (2018) "Is being agreeable a key to success or failure in the labor market?", *Journal of the Japanese and International Economies*, 49, 8–27.

Lee Sun Youn, Ito Takahiro, Kubota Kohei and Ohtake Fumio (2021) "Reciprocal and prosocial tendencies cultivated by childhood school experiences: School uniforms and the related economic and political factors in Japan", *International Journal of Educational Development*, 83, 1–13.

Lee Sun Youn, Sasaki Shusaku, and Ohtake Fumio (2024) "Wearing school uniforms in childhood linked with wearing anti-COVID-19 masks in adulthood: An instrumental variable approach", *Education Economics*, 1–19.

Liu Jing, Monica Lee, and Seth Gershenson (2021) "The short- and long-run impacts of secondary school absences", *Journal of Public Economics*, 199.

Lounsbury John W., Robert P. Steel, James M. Loveland, and Lucy W. Gibson (2004) "An Investigation of Personality Traits in Relation to Adolescent School Absenteeism", *Journal of Youth and Adolescence*, 33 (5), 457–466.

Malkus Nat (2025) "Lingering absence in public schools tracking post-pandemic chronic absenteeism into 2024", *American Enterprise Institute*.

Nakamuro Makiko (2017) “The effect of class size reduction on bullying, violent behavior, and school refusal”, *RIETI Discussion Paper Series*, 17-J-014.

NL Times (2025) “Dutch school absenteeism hits record high as thousands miss classes or aren’t enrolled”, https://nltimes.nl/2025/12/20/dutch-school-absenteeism-hits-record-high-thousands-miss-classes-arent-enrolled (accessed on 26 December 2025).

Oaxaca Ronald (1973) “Male-female wage differentials in urban labor markets”, *International Economic Review*, 14(3), 693–709.

Organisation for Economic Co-operation and Development (OECD) (2023) *PISA 2022 Results Learning During – and From – Disruption. Volume II.*

Organisation for Economic Co-operation and Development (OECD) (2016) “How is learning time organised in primary and secondary education?”, *Education Indicators in Focus*, 38, OECD Publishing. https://doi.org/10.1787/5jm3tqsm1kq5-en (accessed on 26 December 2025).

Oshio Atsushi, Shingo Abe, Pino Cutrone (2012) “Development of the Japanese version of the Ten Item Personality Inventory (TIPI-J)”, *Japanese Journal of Personality*, 21 (1), 40–52

Oshio Atsushi, Shingo Abe, Pino Cutrone, and Samuel D. Gosling (2014) “Further Validity of the Japanese Version of the Ten Item Personality Inventory (TIPI-J)”, *Journal of Individual Differences*, 35(4), 236–244.

Rumberger Russell W.　(1995) “Dropping Out of Middle School: A Multilevel Analysis of Students and Schools”, *American Educational Research Journal*, 32(3), 583-625.

Soma Yuki, Yu Ogasawara and Hiromi Kobayashi (2025) “Early Predictors of School Absenteeism in First-Grade Children: A Multidimensional Longitudinal Study in Japan”, *Children*, *12*(9), 1265.

Sosu M. Edward, Dare Shadrach, Goodfellow Claire, and Klein Markus (2021) “Socioeconomic status and school absenteeism: A systematic review and narrative synthesis”, *Review of Education*, 9(3), 1–28.

Sudo Kosuke (2024) “The impact of family environment on children's experience of school non-attendance: An empirical study of school non-attendance as a ‘welfare’ issue”, *Research Bulletin of Meisei University* (in Japanese).

Tanaka Ryuichi and Atsuyoshi Morozumi (2021) “Additional teacher allocation and problematic behavior of students: A role of households’ economic environment”, *JCER Economic Journal*, 79, 93–126 (in Japanese).

U.S. Department of Education (2025) “Chronic absenteeism supporting student attendance and combatting chronic absenteeism in our nation’s schools”, https://www.ed.gov/teaching-and-administration/supporting-students/chronic-absenteeism (accessed on 26 December 2025).

Yamamoto Hiroki (2008) “Issues concerning official statistics on school refusal: Verification of regional disparities using the five-number summary and examination of alternatives”, *The Journal of Educational Sociology*, 83 (in Japanese).

Yamamura Eiji (2011) “The role of social trust in reducing long-term: School refusal and forming human capital in Japan”, *Economics of Education Review*, 30(2), 380–389.

Figures and Tables

Table 1. Summary Statistics (2019–2023: Full Sample, by School Level and Absenteeism Status)

Elementary school students	(1) Grades 5–6 Non-long-term Absentee			(2) Grades 5–6 Long-term Absentee		
	Obs.	Mean	Std. dev	Obs.	Mean	Std. dev
Long-term Absenteeism Dummy	31,478	0.000	0.000	426	1.000	0.000
Japanese (t-1)	31,010	5.007	0.996	359	4.580	1.169
Mathematics (t-1)	31,012	5.009	0.994	361	4.414	1.193
Extraversion (t-1)	30,766	5.002	1.000	353	4.815	0.970
Agreeableness (t-1)	30,766	5.005	0.998	353	4.686	1.068
Conscientiousness (t-1)	30,761	5.004	0.999	354	4.715	0.989
Emotional Stability (t-1)	30,762	5.003	1.000	353	4.791	0.926
Openness (t-1)	30,734	4.999	1.000	354	5.020	0.994
Quasi-public Assistance Status Dummy	31,478	0.156	0.363	426	0.256	0.437
Public Assistance Status Dummy	31,478	0.020	0.141	426	0.101	0.302
Female Dummy	31,478	0.496	0.500	426	0.465	0.499
Single-parent Household Dummy	31,478	0.153	0.360	426	0.362	0.481
Age of the Youngest Sibling	31,478	9.297	2.892	426	9.453	3.226
First-born Child Dummy	31,478	0.358	0.479	426	0.261	0.439
Number of Siblings	31,478	2.293	0.881	426	2.345	1.100
Sibling with Long-term Absenteeism Dummy (x100)	31,478	1.963	13.87	426	35.92	48.03
TV Time (t-1) (hours/day)	30,794	1.676	0.934	352	1.491	1.043
Study Time (t-1) (hours/day)	30,860	1.170	0.843	355	0.932	0.869
Game Time (t-1) (hours/day)	30,894	1.341	1.007	355	1.731	1.042
Overweight Dummy (t-1)	31,027	0.080	0.272	389	0.093	0.290
Obesity Dummy (t-1)	31,027	0.053	0.224	389	0.111	0.314
Number of Books at Home (t-1) (units of 100)	31,009	0.998	0.698	360	0.861	0.690
Class Size (x0.1)	31,478	3.279	0.410	426	3.279	0.412
Ratio of Long-term Absentees in Class	31,478	0.037	0.038	420	0.052	0.045

Junior high school students	(1) Grades 7–8 Non-long-term Absentee			(2) Grades 7–8 Long-term Absentee		
	Obs.	Mean	Std. dev	Obs.	Mean	Std. dev
Long-term Absenteeism Dummy	26,749	0.000	0.000	657	1.000	0.000
Japanese (t-1)	26,441	4.977	0.971	584	4.490	1.102
Mathematics (t-1)	26,441	4.968	0.952	582	4.344	1.080
Extraversion (t-1)	26,238	4.996	0.997	571	4.800	1.020
Agreeableness (t-1)	26,234	5.014	0.993	572	4.729	1.055
Conscientiousness (t-1)	26,248	4.998	0.993	570	4.699	1.014
Emotional Stability (t-1)	26,235	5.000	0.995	571	4.768	1.009
Openness (t-1)	26,237	4.976	0.997	569	5.062	1.032
Quasi-public Assistance Status Dummy	26,749	0.169	0.375	657	0.280	0.449
Public Assistance Status Dummy	26,749	0.023	0.150	657	0.085	0.279
Female Dummy	26,749	0.491	0.500	657	0.536	0.499
Single-parent Household Dummy	26,749	0.172	0.377	657	0.393	0.489
Age of the Youngest Sibling	26,749	11.162	3.109	657	10.778	3.538
First-born Child Dummy	26,749	0.367	0.482	657	0.387	0.487
Number of Siblings	26,749	2.298	0.875	657	2.317	1.017
Sibling with Long-term Absenteeism Dummy (x100)	26,749	1.245	11.09	657	28.01	44.94
TV Time (t-1) (hours/day)	26,401	1.673	0.932	577	1.483	1.084
Study Time (t-1) (hours/day)	26,406	1.146	0.801	579	0.907	0.849
Game Time (t-1) (hours/day)	26,410	1.528	1.028	579	1.882	1.059
Overweight Dummy (t-1)	26,514	0.072	0.258	637	0.102	0.303
Obesity Dummy (t-1)	26,514	0.047	0.211	637	0.102	0.303
Number of Books at Home (t-1) (units of 100)	26,445	0.966	0.743	584	0.978	0.775
Class Size (x0.1)	26,749	3.631	0.227	657	3.627	0.247
Ratio of Long-term Absentees in Class	26,749	0.089	0.050	656	0.096	0.051

Table 2. Determinants of Long-term Absenteeism

	(1)		(2)		(3)	
	Grades 5–8		Grades 5–6		Grades 7–8	
Japanese (t-1)	-0.001		0.001		-0.002	*
	(0.001)		(0.001)		(0.001)	
Mathematics (t-1)	-0.008	***	-0.005	***	-0.010	***
	(0.001)		(0.001)		(0.001)	
Extraversion (t-1)	-0.003	***	-0.002	**	-0.005	***
	(0.001)		(0.001)		(0.001)	
Agreeableness (t-1)	-0.003	***	-0.002	***	-0.003	***
	(0.001)		(0.001)		(0.001)	
Conscientiousness (t-1)	-0.001	**	-0.001		-0.002	**
	(0.001)		(0.001)		(0.001)	
Emotional Stability (t-1)	-0.002	***	-0.002	***	-0.003	**
	(0.001)		(0.001)		(0.001)	
Openness (t-1)	0.003	***	0.001		0.004	***
	(0.001)		(0.001)		(0.001)	
Quasi-public Assistance Status Dummy	0.001		0.002		0.000	
	(0.002)		(0.002)		(0.004)	
Public Assistance Status Dummy	0.031	***	0.033	***	0.029	**
	(0.008)		(0.010)		(0.013)	
Female Dummy	0.002		-0.002		0.006	***
	(0.001)		(0.001)		(0.002)	
Single-parent Household Dummy	0.016	***	0.010	***	0.022	***
	(0.002)		(0.003)		(0.004)	
Class Size (x0.1)	0.003		0.005	**	-0.002	
	(0.002)		(0.002)		(0.005)	
2020 year Dummy	-0.006	***	-0.005	**	-0.007	**
	(0.002)		(0.002)		(0.003)	
2021 year Dummy	-0.003	*	-0.003		-0.003	
	(0.002)		(0.002)		(0.003)	
2022 year Dummy	0.002		0.000		0.004	
	(0.002)		(0.002)		(0.004)	
2023 year Dummy	0.015	***	0.015	***	0.016	***
	(0.002)		(0.003)		(0.004)	
School Dummy	Y		Y		Y	
Grade Dummy	Y		Y		Y	
Observations	57,604		31,009		26,595	
R-squared	0.02		0.02		0.02	

Notes: Standard errors are between parentheses. **, *, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The full sample includes long-term absentees who did not participate in the Amakko Survey.

Table 3. Determinants of Long-term Absenteeism (With Additional Controls)

	+ Household composition & Time-use			+ Obesity status			+ Number of books			+ Class-level absenteeism rate		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Grades 5–8	Grades 5–6	Grades 7–8	Grades 5–8	Grades 5–6	Grades 7–8	Grades 5–8	Grades 5–6	Grades 7–8	Grades 5–8	Grades 5–6	Grades 7–8
Japanese (t-1)	-0.001	0.001	-0.002 **	0.000	0.001	-0.002 **	-0.001	0.001	-0.002 **	-0.001	0.001	-0.002 **
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Mathematics (t-1)	-0.006 ***	-0.004 ***	-0.007 ***	-0.005 ***	-0.004 ***	-0.007 ***	-0.005 ***	-0.004 ***	-0.007 ***	-0.005 ***	-0.004 ***	-0.007 ***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Extraversion (t-1)	-0.002 ***	-0.002 **	-0.003 ***	-0.002 ***	-0.001 **	-0.003 ***	-0.002 ***	-0.001 **	-0.003 ***	-0.002 ***	-0.001 **	-0.003 **
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Agreeableness (t-1)	-0.003 ***	-0.002 ***	-0.003 **	-0.002 ***	-0.002 ***	-0.003 **	-0.002 ***	-0.002 ***	-0.003 **	-0.002 ***	-0.002 ***	-0.002 **
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Conscientiousness (t-1)	-0.001	0.000	-0.002	-0.001 *	-0.001	-0.002	-0.001 *	-0.001	-0.002	-0.001	-0.001	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Emotional Stability (t-1)	-0.002 ***	-0.001 **	-0.003 **	-0.002 ***	-0.001 **	-0.003 **	-0.002 ***	-0.001 **	-0.003 **	-0.002 ***	-0.001 **	-0.002 **
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Openness (t-1)	0.002 ***	0.001	0.003 ***	0.002 ***	0.001	0.003 ***	0.002 ***	0.001	0.003 ***	0.002 ***	0.001	0.003 ***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Quasi-public Assistance Status Dummy	0.000	0.001	-0.002	-0.001	0.000	-0.002	-0.001	0.000	-0.002	-0.001	0.000	-0.003
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Public Assistance Status Dummy	0.016 **	0.019 **	0.014	0.016 **	0.022 **	0.011	0.016 **	0.021 **	0.010	0.014 *	0.020 **	0.009
	(0.007)	(0.009)	(0.011)	(0.008)	(0.009)	(0.011)	(0.008)	(0.010)	(0.011)	(0.008)	(0.009)	(0.011)
Female Dummy	0.003 **	-0.001	0.008 ***	0.003 **	-0.001	0.007 ***	0.003 **	-0.001	0.008 ***	0.003 **	-0.001	0.008 ***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Single-parent Household Dummy	0.012 ***	0.007 **	0.018 ***	0.012 ***	0.006 **	0.018 ***	0.012 ***	0.006 **	0.019 ***	0.013 ***	0.006 **	0.019 ***
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
Age of the Youngest Sibling	-0.001 ***	-0.001 *	-0.002 ***	-0.001 ***	-0.001 *	-0.002 ***	-0.001 ***	-0.001 *	-0.002 ***	-0.002 ***	-0.001 **	-0.002 ***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
First-born Child Dummy	-0.002	-0.003 *	-0.001	-0.002	-0.003 *	-0.001	-0.002	-0.003 *	-0.001	-0.002	-0.004 **	-0.001
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Number of Siblings	-0.005 ***	-0.003 ***	-0.007 ***	-0.005 ***	-0.003 **	-0.007 ***	-0.005 ***	-0.003 **	-0.007 ***	-0.005 ***	-0.003 ***	-0.007 ***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Sibling with Long-term Absenteeism Dummy	0.002 ***	0.002 ***	0.003 ***	0.002 ***	0.002 ***	0.003 ***	0.002 ***	0.002 ***	0.003 ***	0.002 ***	0.001 ***	0.003 ***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
TV Time (t-1) (hours/day)	-0.002 ***	-0.002 **	-0.003 *	-0.002 ***	-0.002 **	-0.003 *	-0.002 ***	-0.002 **	-0.003 *	-0.002 ***	-0.002 **	-0.002 *
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Study Time (t-1) (hours/day)	-0.001	0.000	-0.002	-0.001	0.000	-0.002	-0.001	0.000	-0.002	-0.001	0.000	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Game Time (t-1) (hours/day)	0.003 ***	0.001 *	0.004 ***	0.002 ***	0.001 *	0.003 ***	0.002 ***	0.001 *	0.004 ***	0.002 ***	0.001 *	0.004 ***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Overweight Dummy (t-1)				0.001	0.000	0.002	0.001	0.000	0.002	0.001	0.000	0.002
				(0.002)	(0.003)	(0.004)	(0.002)	(0.003)	(0.004)	(0.002)	(0.003)	(0.004)
Obesity Dummy (t-1)				0.011 ***	0.005	0.018 ***	0.010 ***	0.005	0.018 ***	0.010 ***	0.004	0.017 ***
				(0.003)	(0.003)	(0.005)	(0.003)	(0.003)	(0.005)	(0.003)	(0.003)	(0.005)
Number of Books at Home (t-1)							0.002 ***	0.000	0.003 **	0.002 ***	0.000	0.003 **
							(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ratio of Long-term Absentees in Class										0.022	0.044	0.001
										(0.017)	(0.027)	(0.025)
Class Size (x0.1)	0.004 *	0.005 **	-0.001	0.003	0.004 *	0.000	0.003	0.004 *	0.000	0.003	0.003	-0.001
	(0.002)	(0.002)	(0.005)	(0.002)	(0.002)	(0.005)	(0.002)	(0.002)	(0.005)	(0.002)	(0.002)	(0.005)
2020 year Dummy	-0.005 ***	-0.004 **	-0.007 **	-0.005 ***	-0.004 **	-0.006 **	-0.005 ***	-0.004 **	-0.006 **	-0.005 ***	-0.004 **	-0.006 *
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
2021 year Dummy	-0.004 **	-0.004 *	-0.003	-0.004 **	-0.004 *	-0.003	-0.004 **	-0.004 *	-0.003	-0.004 ***	-0.004 **	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
2022 year Dummy	-0.001	-0.003	0.002	-0.001	-0.003	0.002	-0.001	-0.003	0.002	-0.002	-0.004 *	0.001
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
2023 year Dummy	0.011 ***	0.012 ***	0.011 ***	0.011 ***	0.012 ***	0.011 ***	0.011 ***	0.012 ***	0.011 ***	0.010 ***	0.009 ***	0.010 **
	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.004)
School Dummy	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Grade Dummy	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	57,323	30,818	26,505	56,889	30,527	26,362	56,858	30,524	26,334	56,825	30,503	26,322
R-squared	0.07	0.06	0.09	0.08	0.07	0.09	0.07	0.07	0.09	0.07	0.06	0.09

Notes: Standard errors are between parentheses. **, *, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The full sample includes long-term absentees who did not participate in the Amakko Survey.

Table 4. Comparison of Results: Restricted Sample (Including Only Amakko Survey Respondents) vs. Full Sample (Including Non-respondents)

	Restricted Sample		Restricted Sample		Full Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
	Grades 5–6	Grades 7–8	Grades 5–6	Grades 7–8	Grades 5–6	Grades 7–8
Japanese (t-1)	0.001	-0.002 **				
	(0.001)	(0.001)				
Mathematics (t-1)	-0.004 ***	-0.007 ***				
	(0.001)	(0.001)				
Extraversion (t-1)	-0.001 **	-0.003 **				
	(0.001)	(0.001)				
Agreeableness (t-1)	-0.002 ***	-0.002 **				
	(0.001)	(0.001)				
Conscientiousness (t-1)	-0.001	-0.002				
	(0.001)	(0.001)				
Emotional Stability (t-1)	-0.001 **	-0.002 **				
	(0.001)	(0.001)				
Openness (t-1)	0.001	0.003 ***				
	(0.001)	(0.001)				
Quasi-public Assistance Status Dummy	0.000	-0.003	0.001	0.002	0.003	0.011
	(0.002)	(0.003)	(0.002)	(0.003)	(0.004)	(0.007)
Public Assistance Status Dummy	0.020 **	0.009	0.023 **	0.025 **	0.081 ***	0.114 ***
	(0.009)	(0.011)	(0.010)	(0.012)	(0.016)	(0.016)
Female Dummy	-0.001	0.008 ***	-0.001	0.004 *	0.000	0.004
	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.004)
Single-parent Household Dummy	0.006 **	0.019 ***	0.010 ***	0.023 ***	0.029 ***	0.063 ***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.006)	(0.007)
Age of the Youngest Sibling	-0.001 **	-0.002 ***	-0.001 **	-0.002 ***	-0.002 ***	-0.006 ***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
First-born Child Dummy	-0.004 **	-0.001	-0.003	-0.006 **	-0.011 ***	-0.037 ***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)
Number of Siblings	-0.003 ***	-0.007 ***	-0.004 ***	-0.008 ***	-0.011 ***	-0.021 ***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
Sibling with Long-term Absenteeism Dummy (x100)	0.001 ***	0.003 ***	0.002 ***	0.003 ***	0.004 ***	0.006 ***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
TV Time (t-1) (hours/day)	-0.002 **	-0.002 *				
	(0.001)	(0.001)				
Study Time (t-1) (hours/day)	0.000	-0.002				
	(0.001)	(0.001)				
Game Time (t-1) (hours/day)	0.001 *	0.004 ***				
	(0.001)	(0.001)				
Overweight Dummy (t-1)	0.000	0.002				
	(0.003)	(0.004)				
Obesity Dummy (t-1)	0.004	0.017 ***				
	(0.003)	(0.005)				
Number of Books at Home (t-1)	0.000	0.003 **				
	(0.001)	(0.001)				
Ratio of Long-term Absentees in Class	0.044	0.001	0.054 *	-0.019	0.041	-0.162 **
	(0.027)	(0.025)	(0.031)	(0.022)	(0.044)	(0.065)
Class Size (x0.1)	0.003	-0.001	0.002	0.000	0.006 *	-0.022 *
	(0.002)	(0.005)	(0.002)	(0.006)	(0.004)	(0.012)
2020 year Dummy	-0.004 **	-0.006 *	-0.003 *	-0.004	-0.004	-0.005
	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.005)
2021 year Dummy	-0.004 **	-0.003	-0.004 **	-0.004	-0.001	0.009
	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.007)
2022 year Dummy	-0.004 *	0.001	-0.004	0.004	-0.001	0.023 **
	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.008)
2023 year Dummy	0.009 ***	0.010 **	0.011 ***	0.017 ***	0.015 ***	0.028 ***
	(0.002)	(0.004)	(0.002)	(0.004)	(0.004)	(0.009)
School Dummy	Y	Y	Y	Y	Y	Y
Grade Dummy	Y	Y	Y	Y	Y	Y
Observations	30,503	26,322	31,872	27,386	32,640	29,285
R-squared	0.06	0.09	0.07	0.10	0.18	0.23

Notes: Standard errors are between parentheses. **, *, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The restricted sample excluded students that are classified as a long-term absentee in a given year and did not participate in the Amakko Survey in the preceding year. The full sample includes long-term absentees who did not participate in the Amakko Survey.

Table 5. Transitions in Reasons for Long-term Absenteeism

Prior Year Reason (N)	Obs.	Current Year Reason for Absenteeism (N) School refusal	Family/ Personal Circumstances	Others	Sickness
School refusal	1,536	1,265	92	2	168
Family/Personal Circumstances	543	212	236	6	77
Others	41	5	9	25	2
Sickness	1,125	425	82	3	583

Prior Year Reason (%)	Obs.	Current Year Reason for Absenteeism (%) School refusal	Family/ Personal Circumstances	Others	Sickness
School refusal	1,536	82.4	6.0	0.1	10.9
Family/Personal Circumstances	543	39.0	43.5	1.1	14.2
Others	41	12.2	22.0	61.0	4.9
Sickness	1,125	37.8	7.3	0.3	51.8

Table 6. Determinants of Long-term Absenteeism by Reason

	Elementary School Students			Junior High School Students		
	(1)	(2)	(3)	(4)	(5)	(6)
	School refusal	Family/Personal Circumstances	Sickness	School refusal	Family/Personal Circumstances	Sickness
Japanese (t-1)	0.0008	-0.0003	0.0003	-0.0005	-0.0006 *	-0.0012
	(0.0006)	(0.0004)	(0.0006)	(0.0005)	(0.0003)	(0.0009)
Mathematics (t-1)	-0.0008	-0.0004	-0.0023 **	-0.0007	-0.0006	-0.0054 ***
	(0.0005)	(0.0005)	(0.0009)	(0.0006)	(0.0007)	(0.0009)
Extraversion (t-1)	-0.0001	-0.0001	-0.0010 **	-0.0008 ***	0.0001	-0.0023 **
	(0.0003)	(0.0003)	(0.0004)	(0.0002)	(0.0003)	(0.0009)
Agreeableness (t-1)	-0.0006 *	-0.0003	-0.0006 *	-0.0016 ***	0.0002	-0.0008
	(0.0004)	(0.0003)	(0.0004)	(0.0005)	(0.0004)	(0.0009)
Conscientiousness (t-1)	-0.0004	0.0000	-0.0003	-0.0006	0.0000	-0.0010
	(0.0003)	(0.0003)	(0.0004)	(0.0004)	(0.0002)	(0.0011)
Emotional Stability (t-1)	0.0000	0.0000	-0.0013 ***	0.0000	0.0000	-0.0024 **
	(0.0003)	(0.0003)	(0.0004)	(0.0003)	(0.0003)	(0.0008)
Openness (t-1)	0.0000	0.0002	0.0008	0.0005 *	0.0010 **	0.0010
	(0.0002)	(0.0003)	(0.0006)	(0.0002)	(0.0004)	(0.0009)
Quasi-public Assistance Status Dummy	-0.0014	0.0014	-0.0002	-0.0005	-0.0004	-0.0018
	(0.0009)	(0.0011)	(0.0015)	(0.0011)	(0.0007)	(0.0031)
Public Assistance Status Dummy	-0.0041 *	0.0120 ***	0.0119 **	0.0010	0.0151	-0.0064
	(0.0023)	(0.0044)	(0.0058)	(0.0028)	(0.0089)	(0.0062)
Female Dummy	-0.0001	-0.0001	-0.0006	0.0001	0.0001	0.0077 ***
	(0.0006)	(0.0005)	(0.0009)	(0.0009)	(0.0006)	(0.0018)
Single-parent Household Dummy	0.0025 *	0.0013	0.0024	0.0009	0.0015 **	0.0166 ***
	(0.0012)	(0.0011)	(0.0019)	(0.0014)	(0.0007)	(0.0027)
Age of the Youngest Sibling	-0.0002	-0.0002 *	-0.0007 **	-0.0003	-0.0001	-0.0016 ***
	(0.0002)	(0.0001)	(0.0003)	(0.0002)	(0.0002)	(0.0005)
First-born Child Dummy	-0.0006	-0.0004	-0.0027 **	-0.0009	-0.0003	-0.0003
	(0.0009)	(0.0007)	(0.0012)	(0.0007)	(0.0006)	(0.0022)
Number of Siblings	0.0001	-0.0009 **	-0.0026 ***	-0.0011 *	-0.0006	-0.0053 ***
	(0.0006)	(0.0004)	(0.0007)	(0.0006)	(0.0006)	(0.0013)
Sibling with Long-term Absenteeism Dummy (x100)	0.0003 ***	0.0003 ***	0.0009 ***	0.0004 ***	0.0003 **	0.0021 ***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
TV Time (t-1) (hours/day)	0.0001	-0.0008 **	-0.0010 *	-0.0002	0.0001	-0.0024 *
	(0.0003)	(0.0003)	(0.0006)	(0.0003)	(0.0003)	(0.0012)
Study Time (t-1) (hours/day)	-0.0001	-0.0005	0.0002	-0.0006	-0.0002	-0.0011
	(0.0004)	(0.0005)	(0.0006)	(0.0005)	(0.0003)	(0.0012)
Game Time (t-1) (hours/day)	0.0007 *	0.0001	0.0004	0.0009 *	-0.0001	0.0026 ***
	(0.0004)	(0.0002)	(0.0005)	(0.0005)	(0.0003)	(0.0009)
Overweight Dummy (t-1)	-0.0021 ***	0.0004	0.0015	-0.0010	-0.0009	0.0040
	(0.0006)	(0.0011)	(0.0024)	(0.0011)	(0.0013)	(0.0042)
Obesity Dummy (t-1)	0.0026	0.0022	-0.0008	0.0000	0.0008	0.0167 ***
	(0.0021)	(0.0017)	(0.0020)	(0.0020)	(0.0022)	(0.0038)
Number of Books at Home (t-1)	0.0000	0.0007 *	-0.0005	0.0002	0.0004	0.0023 **
	(0.0004)	(0.0004)	(0.0006)	(0.0003)	(0.0005)	(0.0010)
Ratio of Long-term Absentees in Class	0.0213 *	0.0000	0.0230	-0.0007	-0.0046	0.0068
	(0.0117)	(0.0074)	(0.0202)	(0.0105)	(0.0073)	(0.0178)
Class Size (x0.1)	0.0012	0.0008	0.0011	0.0008	-0.0034 **	0.0020
	(0.0007)	(0.0012)	(0.0014)	(0.0013)	(0.0014)	(0.0038)
2020 year Dummy	0.0014	0.0000	-0.0050 ***	0.0021 *	0.0007	-0.0085 ***
	(0.0008)	(0.0008)	(0.0013)	(0.0012)	(0.0006)	(0.0027)
2021 year Dummy	0.0007	-0.0004	-0.0045 ***	0.0024 **	0.0010	-0.0065 **
	(0.0008)	(0.0008)	(0.0016)	(0.0011)	(0.0007)	(0.0028)
2022 year Dummy	0.0004	-0.0004	-0.0037 **	0.0011	0.0023 *	-0.0017
	(0.0008)	(0.0009)	(0.0018)	(0.0007)	(0.0011)	(0.0031)
2023 year Dummy	0.0013	0.0025 **	0.0054 ***	0.0028 **	0.0024 **	0.0056
	(0.0008)	(0.0012)	(0.0015)	(0.0012)	(0.0010)	(0.0041)
School Dummy	Y	Y	Y	Y	Y	Y
Grade Dummy	Y	Y	Y	Y	Y	Y
Observations	30,503	30,503	30,503	26,322	26,322	26,322
R-squared	0.01	0.01	0.04	0.01	0.01	0.07

Notes: Standard errors are between parentheses. **, *, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 7. Results of Blinder–Oaxaca Decomposition

Elementary School Students	2019 β_B		2023 β_A		Endowments $(x_A - x_B)\beta_B$	Coefficients $x_B(\beta_A - \beta_B)$	Share	Interactions $(x_A - x_B) \times (\beta_B - \beta_A)$
Japanese (t-1)	0.003	*	-0.001		0.000	-1.789	-226	-0.001
Mathematics (t-1)	-0.004	**	-0.007	***	0.005	-1.848	-233	-0.002
Extraversion (t-1)	-0.002		-0.003		0.001	-0.290	-37	0.000
Agreeableness (t-1)	-0.002	*	-0.003		0.000	-0.387	-49	0.000
Conscientiousness (t-1)	0.000		-0.002		0.001	-0.872	-110	-0.001
Emotional Stability (t-1)	-0.002		-0.002		0.001	-0.229	-29	0.000
Openness (t-1)	0.001		0.000		0.000	-0.362	-46	0.001
Quasi-public Assistance Status Dummy	0.002		-0.003		0.001	-0.081	-10	-0.002
Public Assistance Status Dummy	0.025	***	0.042	***	-0.038	0.027	3	0.015
Female Dummy	0.001		-0.003		0.005	-0.151	-19	-0.006
Single-parent Household Dummy	0.003		0.015	**	-0.019	0.168	21	0.015
Age of the Youngest Sibling	0.000		-0.004	***	-0.005	-3.255	-411	0.005
First-born Child Dummy	0.000		-0.011	**	0.011	-0.377	-48	-0.011
Number of Siblings	-0.003	*	-0.011	***	-0.002	-1.881	-238	0.002
Sibling with Long-term Absenteeism Dummy (x100)	0.001	***	0.002	***	0.363	0.277	35	-0.132
TV Time (t-1) (hours/day)	0.000		-0.005	**	0.108	-0.742	-94	-0.101
Study Time (t-1) (hours/day)	-0.001		-0.003		0.072	-0.172	-22	-0.038
Game Time (t-1) (hours/day)	0.003	**	0.002		0.051	-0.087	-11	0.013
Overweight Dummy (t-1)	-0.002		0.000		0.001	0.022	3	-0.004
Obesity Dummy (t-1)	-0.010		0.027	***	0.052	0.224	28	-0.071
Number of Books at Home (t-1) (units of 100)	0.002		0.001		-0.012	-0.138	-17	-0.019
Ratio of Long-term Absentees in Class	0.077	*	-0.021		-0.057	-0.574	-72	0.272
Class Size (x0.1)	0.003		0.012	*	-0.161	2.845	359	0.122
6th grade Dummy	0.003		-0.001		-0.001	-0.182	-23	0.003
School District Dummies					-0.034	0.858	108	0.103
Constant	0.031		0.129	***	0.000	9.787	1236	0.000
Total	0.931		2.226	1.295	0.342	0.792		0.161
Share (%)					26.39	61.15		12.46

Junior High School Students	2019 β_B		2023 β_A		Endowments $(x_A - x_B)\beta_B$	Coefficients $x_B(\beta_A - \beta_B)$	Share	Interactions $(x_A - x_B) \times (\beta_B - \beta_A)$
Japanese (t-1)	-0.005	*	-0.003		0.000	1.040	90.8	0.000
Mathematics (t-1)	-0.007	**	-0.012	***	-0.006	-2.547	-222.2	0.003
Extraversion (t-1)	0.000		-0.005	*	-0.008	-2.566	-223.9	0.009
Agreeableness (t-1)	-0.003	*	-0.003		-0.003	0.369	32.2	-0.001
Conscientiousness (t-1)	-0.002		-0.003		-0.001	-0.536	-46.7	0.000
Emotional Stability (t-1)	0.000		-0.005	**	-0.005	-2.666	-232.7	0.005
Openness (t-1)	0.001		0.004	*	-0.001	1.771	154.6	0.001
Quasi-Public Assistance Status Dummy	-0.003		0.004		-0.007	0.109	9.5	0.013
Public Assistance Status Dummy	-0.003		0.060	***	-0.057	0.109	9.5	0.060
Female Dummy	0.010	***	0.010	*	0.017	-0.022	-2.0	0.001
Single-parent Household Dummy	0.024	***	0.029	***	-0.038	0.082	7.2	0.007
Age of the Youngest Sibling	-0.001		-0.002	*	-0.007	-1.489	-129.9	0.004
First-born Child Dummy	-0.001		-0.001		0.000	-0.020	-1.8	0.000
Number of Siblings	-0.007	***	-0.010	***	0.010	-0.627	-54.7	-0.003
Sibling with Long-term Absenteeism Dummy (x100)	0.003	***	0.003	***	0.256	-0.116	-10.1	0.044
TV Time (t-1) (hours/day)	0.002		-0.008	***	0.237	-1.494	-130.4	-0.284
Study Time (t-1) (hours/day)	-0.005	**	0.001		-0.008	0.653	57.0	0.079
Game Time (t-1) (hours/day)	0.005	***	0.003		0.065	-0.470	-41.0	0.071
Overweight Dummy (t-1)	0.011		0.005		0.013	-0.050	-4.3	0.014
Obesity Dummy (t-1)	0.029	***	0.023	**	0.029	-0.031	-2.7	0.007
Number of Books at Home (t-1) (units of 100)	0.002		0.007	*	0.056	0.464	40.5	-0.037
Ratio of Long-term Absentees in Class	-0.048		-0.073		-0.272	-0.280	-24.4	0.091
Class Size (x0.1)	0.009		0.006		0.005	-1.370	-119.5	0.003
8th grade Dummy	0.009	**	0.015	***	-0.005	0.282	24.6	0.002
School District Dummies					0.019	-0.573	-50.0	-0.008
Constant	0.071		0.182	***	0.000	11.122	970.5	0.000
Total	1.811		3.325	1.514	0.289	1.146		0.079
Share (%)					19.07	75.69		5.24

Note: **, *, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

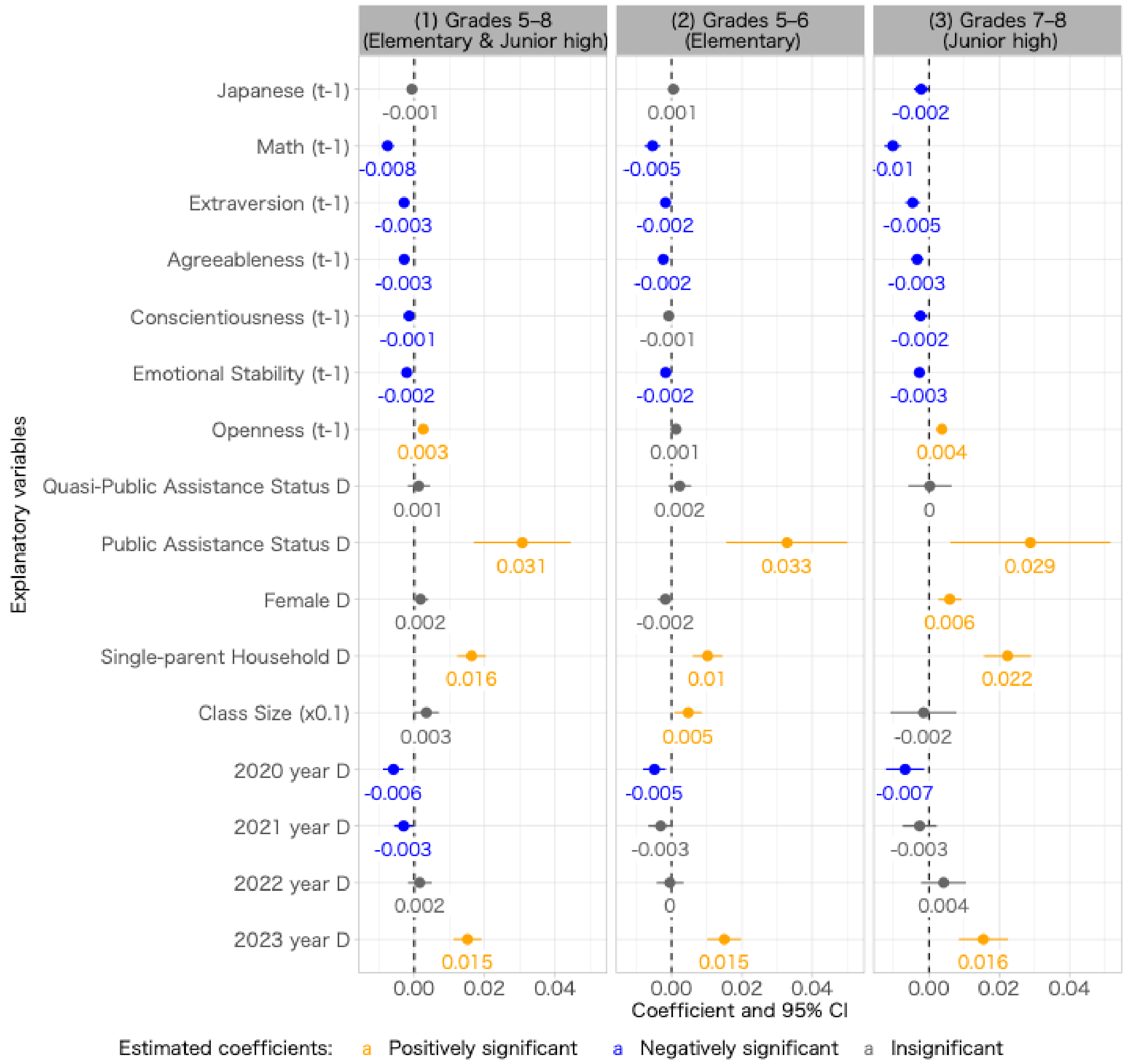

Figure 1. Coefficient plot of explanatory variables for the probability of long-term absenteeism (visualizing Table 2)

Note: The dots represent coefficients and the bars 95% confidence intervals. Estimations were conducted using pooled ordinary least squares on panel data for individual students from fiscal years 2019 to 2023, including fixed effects for year, school, grade, and class size. Class-clustered, robust standard errors were used.

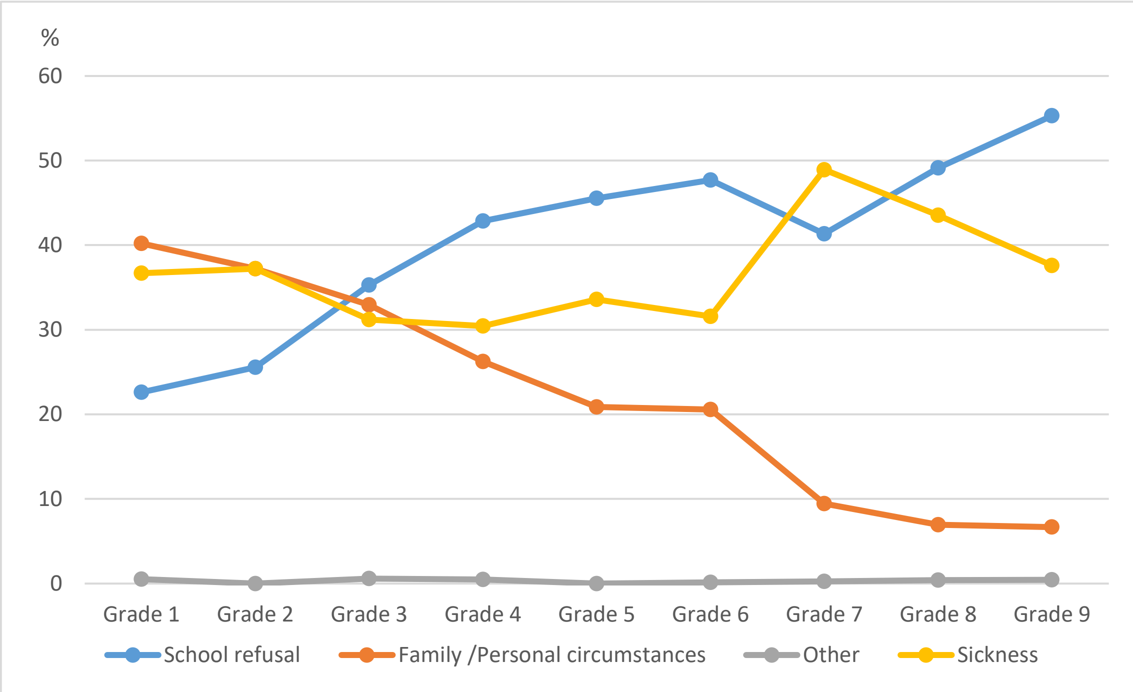

Figure 2. Distribution of reasons for long-term absenteeism by grade

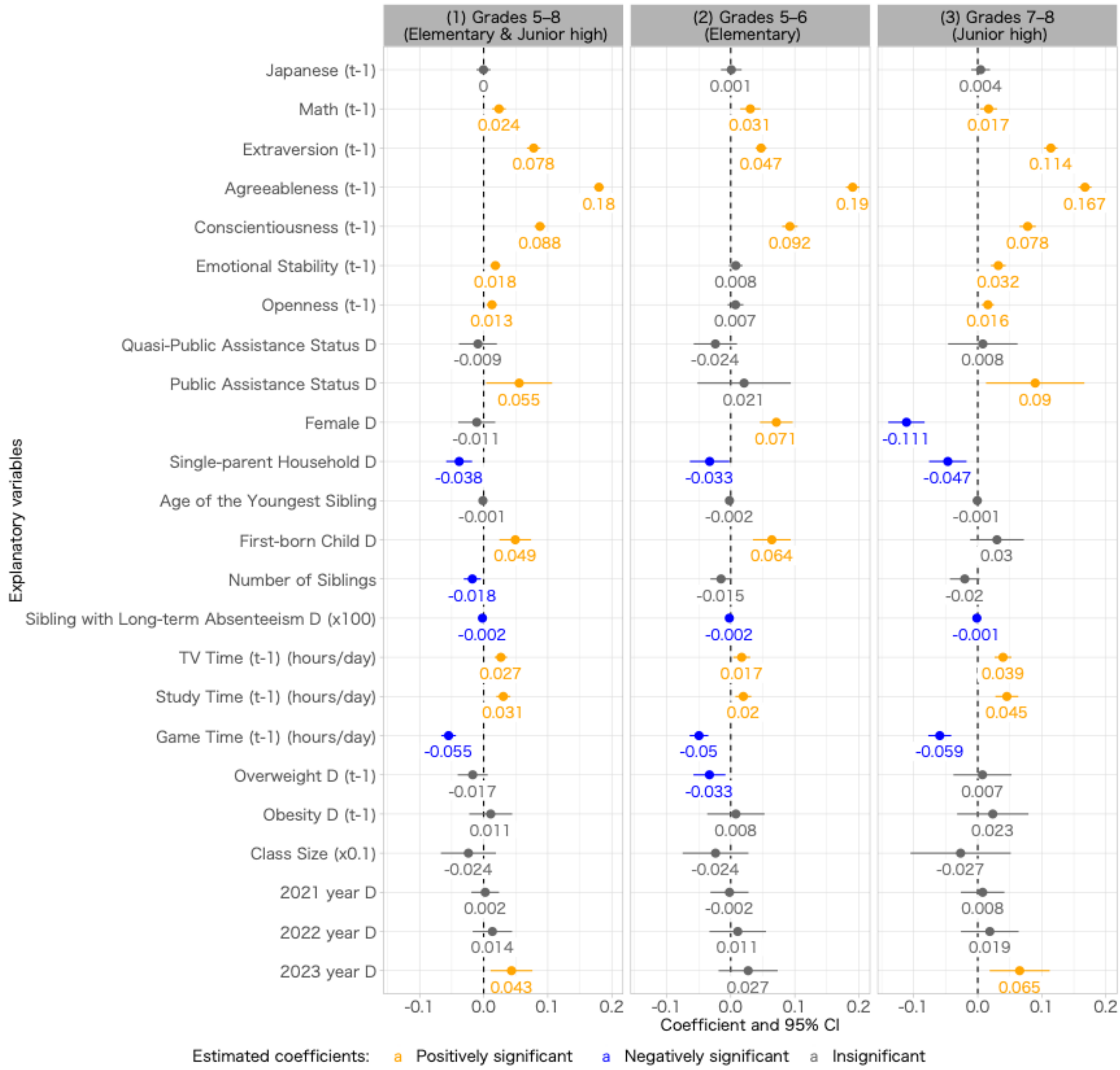

Figure 3. Coefficient plot of explanatory variables for "I enjoying going to school" as the dependent variable

Note: The dots represent coefficients and the bars 95% confidence intervals. Estimations were conducted using pooled ordinary least squares on panel data for individual students from fiscal years 2019 to 2023, including fixed effects for year, school, grade, and class size. Class-clustered, robust standard errors were used.

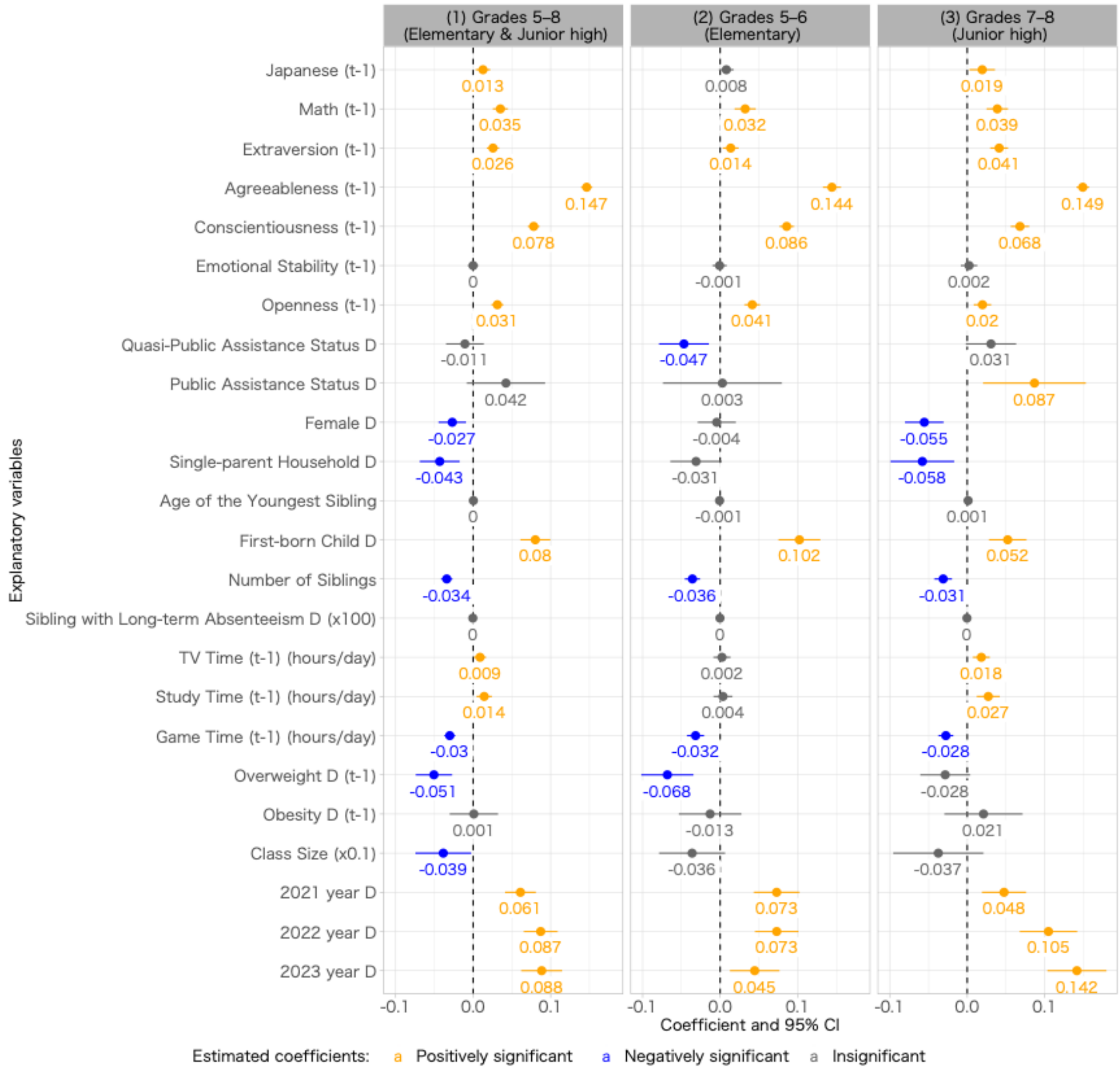

Figure 4. Coefficient plot of explanatory variables for "I feel that my teachers have recognized and valued me" as the dependent variable

Note: The dots represent coefficients and the bars 95% confidence intervals. Estimations were conducted using pooled ordinary least squares on panel data for individual students from fiscal years 2019 to 2023, including fixed effects for year, school, and grade. Class-clustered, robust standard errors were used.

Appendix: Data Sources, Definitions, and Notes for Variables Used in the Analysis

Variables	Data Sources	Definitions and Notes
Long-term Absenteeism Dummy	Long-term Absenteeism Records	Dummy variable coded as 1 if the student was absent for 30 days or more between April and February.
Japanese/Mathematics (t-1)	Amakko Survey	Standardized score (T-score) for each academic year and grade, divided by 10 for the previous academic year.
Extraversion (t-1)	Amakko Survey	**Elementary School Students**: The score obtained by taking the mean of the self-assessments “I like to stand out” and “I am rather quiet (reversed),” standardizing it as a T-score within each fiscal year and grade, and then dividing by 10 (using values from the previous year). **Junior High School Students**: The score obtained by taking the mean of the self-assessments “I think I am active and outgoing” and “I think I am reserved and quiet (reversed),” standardizing it as a T-score within each fiscal year and grade, and then dividing by 10 (using values from the previous year).
Agreeableness (t-1)	Amakko Survey	**Elementary School Students**: The mean of the assessments “I am a caring person” and “I have few friends I can rely on (reversed),“ standardized as a T-score within each academic year and grade, then divided by 10 (value from the previous year). **Junior High School Students**: The mean of the assessments “I think I am a kind person who is considerate of others“ and “I think I tend to be dissatisfied with others and easily get into trouble (reversed),“ standardized as a T-score within each academic year and grade, then divided by 10 (value from the previous year).
Conscientiousness (t-1)	Amakko Survey	**Elementary School Students**: The mean of the self-assessments “I work hard at everything“ and “I often give up in the middle of tasks (reversed),“ standardized as a T-score within each academic year and grade, then divided by 10 (mean value from the previous year). **Junior High School Students**: The mean of the self-assessments “I think I am reliable and strict with myself“ and “I think I am sloppy and careless (reversed),“ standardized as a T-score within each academic year and grade, then divided by 10 (value from the previous year).
Emotional Stability (t-1)	Amakko Survey	**Elementary School Students**: The mean of the assessments “I am always worried about making mistakes“ and “I worry about even small things,“ standardized as a T-score within each academic year and grade, then divided by 10 (mean value from the previous year). **Junior High School Students**: The mean of the assessments “I think I am prone to worrying and easily get flustered“ and “I think I am calm and

		emotionally stable (reversed),“ standardized as a T-score within each academic year and grade, then divided by 10 (value from the previous year).
Openness (t-1)	Amakko Survey	**Elementary School Students**: The mean of the assessments “I enjoy daydreaming and thinking about various things“ and “I know a lot about many things,“ standardized as a T-score within each academic year and grade, then divided by 10 (value from the previous year). **Junior High School Students**: The mean of the assessments “I like new things and have unusual ideas“ and “I think I am an ordinary person lacking imagination (reversed),“ standardized as a T-score within each academic year and grade, then divided by 10 (value from the previous year).
Quasi-public Assistance Status Dummy	School Expense Subsidies	A dummy variable coded as 1 if the student’s household received Quasi-Public Assistance support during the survey year. This status is granted to households recognized by the local Board of Education as experiencing a level of financial hardship comparable to that of Indigent (yo-hogo) persons as defined under Article 6, Paragraph 2 of the Public Assistance Act.
Public Assistance Status Dummy	School Expense Subsidies	A dummy variable coded as 1 if the student’s household received public assistance (seikatsu-hogo) during the survey year. This includes households recognized by the local Board of Education as experiencing a level of financial hardship comparable to those receiving public assistance as defined under Article 6, Paragraph 2 of the Public Assistance Act.
Female Dummy	Amakko Survey, Long-term Absenteeism Records	Dummy variable coded as 1 if the student is female.
Single-parent Household Dummy	Basic Resident Register	Dummy variable for single-parent households (including both mother-child and father-child households).
Age of the Youngest Sibling	Basic Resident Register	Age of the youngest sibling in the household during the survey year.
First-born Child Dummy	Basic Resident Register	Dummy variable coded as 1 if the student is the eldest child in a household with two or more children.
Number of Siblings	Basic Resident Register	Total number of children in the household.
Sibling with Long-term Absenteeism Dummy (x100)	Basic Resident Register	Dummy variable coded as 1 if at least one sibling was absent for 30 days or more in the previous year (multiplied by 100 for descriptive statistics).
Number of Books at Home (t-1) (units of 100)	Amakko Survey	Previous year’s number of books at home, excluding textbooks and magazines. Values are divided by 100 for analysis.
TV Time (t-1) (hours/day)	Amakko Survey	Daily time spent watching TV was calculated based on responses regarding usage on weekdays and holidays respectively. The response options—rarely

		watch, about 30 min, about 1h, about 1.5h, about 2h, about 2.5h, and 3h or more—were assigned values of 0, 30, 60, 90, 120, 150, and 180 minutes, respectively. From these values, the total weekly time was calculated and then divided by 7 to obtain the average daily time.
Study Time (t-1) (hours/day)	Amakko Survey	Daily study time (including time spent at cram schools or with private tutors) was calculated based on responses regarding usage on weekdays and holidays respectively. The response options—rarely study, about 30 min, about 1h, about 1.5h, about 2h, about 2.5h, and 3h or more—were assigned values of 0, 30, 60, 90, 120, 150, and 180 minutes, respectively. From these values, the total weekly time was calculated and then divided by 7 to obtain the average daily time.
Game Time (t-1) (hours/day)	Amakko Survey	Daily time spent playing video or smartphone games was calculated based on responses regarding usage on weekdays and holidays respectively. The response options—rarely play, about 30 min, about 1h, about 1.5h, about 2h, about 2.5h, and 3h or more—were assigned values of 0, 30, 60, 90, 120, 150, and 180 minutes, respectively. From these values, the total weekly time was calculated and then divided by 7 to obtain the average daily time.
Class Size (x0.1)	Amakko Survey	The average number of students per class, calculated for each grade within each school by dividing the total number of students by the number of classes during the survey year.
Ratio of Long-term Absentees in Class	Long-term Absenteeism Records, Student Enrollment Records	The proportion of long-term absentees within each class during the survey year, calculated excluding the student him/herself.
Overweight Dummy (t-1)	Height and Weight	Dummy variable coded as 1 if the student's Rohrer Index is between 145 and 160.
Obesity Dummy (t-1)	Height and Weight	Dummy variable coded as 1 if the student's Rohrer Index is 160 or higher.