Non-cognitive Skills and Learning Outcomes in Post-Disaster Recovery *

Esma Ozer and Faqiang Li†

September 27, 2025

Preliminary Version

We investigate the role of non-cognitive skills in shaping student outcomes after a major natural disaster. Using a new survey of over 1,600 high school students and their parents, along with LLM-based text analysis, we study outcomes related to mental health, academic performance, post-disaster migration, parental inputs, and new study environments. We then examine how these outcomes vary with students' non-cognitive skills, specifically grit and resilience. We find that earthquake exposure disrupted not only the physical environment but also mental well-being. The return to baseline non-cognitive skills appears stronger in high-destruction areas. Finally, while academic performance is more responsive to resilience, mental health outcomes are less so, suggesting that coping mechanisms or support systems beyond resilience play an important role to overcome psychological trauma.

*First version: September, 2025. We are very grateful for the guidance and useful comments from Caterina Calsamiglia, Michael Gechter, Kei Hirano, Brad Setzler, Andres Aradillas Lopez, Ewout Verriest, and Chloe Tergiman. We thank Alejandro Sanchez for his help in the survey design and Basit Zafar for kindly sharing the questionnaires he used in his research. This project would not have been possible without the support from the Turkish Ministry of Education. We especially thank Yusuf Altunhan and Emine Akdag for their collaboration and support. We thank Hatice Sanver-Gursoy for her help in designing questions in age-appropriate way. We thank Zeynep Yildiz and her team for coordinating this large-scale survey and Elif Deniz for field assistance. We thank Hatice Ilhan for her insights on the local context. We also acknowledge research assistance by OpenAI GPT and Anthropic Claude models. Ozer is grateful for the financial support from Penn State University's Economics Department. The Institutional Review Board approval was obtained from Penn State University (STUDY00023803). The Turkish Department of Education research registry document ID is 101920742 with approval date 05/14/2024.

†Ozer: Department of Economics, Penn State University (esmaozer@psu.edu); Li: INSE, Peking University (faqiangli.econ@gmail.com).

1 Introduction

Disruptions to education due to external shocks such as natural disasters including earthquakes and wildfires, pandemics such as COVID 19, weather changes due to climate change, or internal conflicts are expected to persist because of natural geological reasons, increasing population movements, and ecosystem changes (IPCC 2012, Caldara and Iacoviello, 2022, Williams et al., 2023). With these shocks continuing, the disaster costs and direct economic impacts are expected to be elevated. (E. Cavallo et al., 2010). It is not only the physical capital losses that matter but also the human capital losses, which are often more persistent (E. A. Cavallo & Noy, 2009; Dercon & Porter, 2014; Tincani, 2018).

Human capital losses can be direct (such as physical disruptions including losses, injuries, or disruption to education) (Belmonte et al., 2020) or indirect (such as mental health issues or trauma experience, migration, or changes in parental inputs) (Guariso & Björkman Nyqvist, 2023). Losses due to disruptions are also uneven across different groups or geographies, which can be due to differences in school disclosure policies, financial support systems, and parental inputs (Inui and Okudaira, 2022, Parolin and Lee, 2021a). While these can be directly observed and relevant mechanisms be tested, there might also other underlying mechanisms that can shape how individuals respond to these shocks. Understanding these mechanisms is important for designing effective policies to mitigate the negative impacts of such shocks on human capital accumulation.

One potential mechanism we focus on is non-cognitive skills. We ask whether traits such as resilience and grit help reduce the negative impacts of a major natural disaster on student outcomes. We also examine whether the importance of these skills depends on the severity of the shock, and whether they matter more for some outcomes such as academic performance compared to others such as mental health. In addition, we consider how non-cognitive skills may move together with other mechanisms, including parental inputs and migration decisions, that also shape student outcomes.

We study these questions in the context of the 2023 Türkiye earthquakes, which caused major physical and human capital losses. We use a new survey of more than 1,600 high school students and their parents from three provinces with different levels of earthquake exposure: low, medium, and high exposure. The survey builds on existing theoretical and empirical work on non-cognitive skills and educational outcomes, and was designed to capture a wide range of relevant dimensions. We collect information on pre and post earthquake mental health, academic performance, migration, parental inputs, time use, social networks, and study environments. We also ask open ended questions to capture student experiences that are not easily measured with structured questions. Using LLM based text analysis, we then extract patterns and sentiments from these open-ended responses.

We begin by documenting several descriptive patterns. Survey responses show that students in the affected provinces faced a wide range of disruptions after the earthquakes. Many reported separation from friends and relatives, changes in study environments, and loss of educational resources. Schooling was also disrupted, with closures and interruptions to tutoring. Families faced additional shocks, including housing damage and job loss. In short, the earthquakes led not only to physical infrastructure damage but also to serious disruptions in the social and educational environments in which students learn and develop.[1]

In addition to physical disruptions, students also experienced serious mental health challenges. Anxiety symptoms increased in the most affected provinces, and overall well-being declined across the sample. LLM based analysis of open-ended responses shows large psychological impacts even in the least affected province. More than half of the students in our sample mentioned psychological effects in their answers, followed by social disruptions. Academic concerns appeared less often, though they were still important for some students.

In the empirical analysis, we study how changes in academic and mental health outcomes vary with students' non-cognitive skills, focusing on resilience and grit. We are particularly interested in how the interaction between earthquake intensity and non-cognitive skills appears once we control for individual and parental baseline characteristics. The variation across provinces provides a natural control–treatment setting, with medium and high intensity provinces serving as treatment groups. For non-cognitive skills, we use a behavioral measure of resilience and grit. To address potential endogeneity, we use orthogonalized changes in study time and rely on a selection on observables assumption. We also conduct a broad set of robustness checks, including analyses for migration and recall bias.

Our results show that non-cognitive skills, especially resilience and grit, are strongly associated with students' ability to sustain academic performance after the earthquakes. Students with higher levels of resilience performed better on exams, even after controlling for pre disaster study time, household background, and parental inputs. The association is particularly pronounced in the high destruction province, suggesting that non-cognitive skills become more important in contexts where the shock is most severe. These findings highlight the role of behavioral traits in supporting academic recovery in the face of large external disruptions.

When we turn to mental health, the patterns are different. Anxiety levels rose and overall well being declined across all provinces, and the interaction between earthquake exposure and resilience is too small to offset the psychological distress in high impact areas. While students with higher resilience show somewhat better mental health outcomes, the effects are limited and statistically weaker. This indicates that while academic outcomes are responsive to perseverance and sustained effort, mental health disruptions may require different coping mechanisms and broader support systems beyond individual traits.

We also find important heterogeneity across subgroups. Female students benefit more consistently from resilience, with positive effects in both quantitative and verbal subjects, while male students show smaller and less consistent gains. Household resources also

[1] These patterns are based on our survey sample, which may be affected by post-migration selection. We address this issue in the empirical analysis.

matter, as students from more advantaged families experienced fewer disruptions. Finally, when we examine outcomes corresponding to almost a year later using student reported school grades, we continue to see positive associations between resilience and academic performance, particularly in the hardest hit regions. However, the medium term effects on mental health remain limited, reinforcing the idea that educational and psychological recovery follow different paths.

1.1 Related Literature

This project contributes to three strands of literature. The first relates to research on educational disruptions and their impact on student outcomes. Much of this literature draws on sudden adverse events such as natural disasters and pandemics to study the effects of schooling interruptions on students' academic and non-cognitive development. For example, Angrist et al. (2023) provides experimental evidence on how education policies implemented during COVID-19 can enhance system resilience to future shocks. A growing number of studies, across various contexts, investigate the learning losses associated with school closures during the pandemic. For instance, Alasino et al. (2024) in Mexico; Aucejo et al. (2020), Goldhaber et al. (2022), and Jack et al. (2023) in the U.S.; and Guariso and Björkman Nyqvist (2023) in India, all report consistent evidence that students experienced considerable learning losses due to pandemic-related school disruptions.[2]

While COVID-19 brought a global attention to educational disruptions, a substantial body of literature has previously examined how natural disasters impact student learning outcomes. For instance, Sacerdote (2012) examines the consequences of Hurricane Katrina for student achievement, while Gallagher et al. (2023) studies the impact of Hurricane Harvey on human capital investments. Andrabi et al. (2021) investigates the 2005 Pakistan earthquake, Paudel and Ryu (2018) the Nepal earthquakes, and Shidiqi et al. (2023) the effects of earthquakes in Indonesia. These studies also provides evidence on the negative effects on disasters on educational outcomes.[3]

While much of the work in this area focuses on the impacts of adverse events on primarily cognitive outcomes, there is relatively less work examining the role of underlying mechanisms such as noncognitive skills in shaping student outcomes post-shocks. The importance of noncognitive skills for both educational and labor market outcomes is well established (Alan et al., 2019; Cunha et al., 2010; Heckman et al., 2006), and there is also growing evidence on how these skills respond to adverse events (Biener & Landmann, 2023). However, the interaction of noncognitive skills with other mechanisms in the con-

[2] Other relevant COVID-19 studies in this area include, but are not limited to, Johnston et. al. (2021), Alan and Turkum (2024), Bacher-Hicks et al. (2021), Betthäuser et al. (2023), Boruchowicz et al. (2022), "The Economic Impacts of Learning Losses" (2020), Kuhfeld et al. (2020), Lichand et al. (2022), and Parolin and Lee (2021b).

[3] Additional relevant work on natural disasters includes (Belmonte et al., 2020; Bulaon & Shoji, 2022; Cerqua & Di Pietro, 2017; Morrill & Westall, 2023). Some studies, such as Imberman et al. (2012) and Tincani (2018) also use natural disasters as quasi-experiments to explore broader outcomes, such as classroom peer effects. A related body of literature examines how learning and human capital accumulation evolve during economic downturns. For an example, see Ersoy (2020).

text of extreme shocks remains underexplored. Chen et al. (2024) and Lytle and Shin (2023) highlight the role of noncognitive traits like grit in predicting success and hope for the future, particularly in the COVID-19 context.[4] This paper contributes to this line of work by providing evidence on the importance of resilience and endurance in the context of a natural disaster, and how these traits can serve as mechanisms to improve student outcomes following such disruptions.

Lastly, this paper also contributes to the literature on education policy. Much of the research in this area focuses on evaluating cognitive skills, such as how exams sort students based on their cognitive abilities (Brown et al., 2025; Jacob & Rothstein, 2016; Ozer et al., 2024). However, there is growing recognition of noncognitive skills' importance in labor markets (Deming, 2017) and in recovery from disruptions. This paper highlights the potential for integrating psychosocial interventions into curriculum design (Alan & Ertac, 2018; Alan & Kubilay, 2023) to enhance students' resilience and endurance, which can be crucial for their long-term success following educational disruptions.

Road Map: The remainder of this paper proceeds as follows. Section 2 describes our data sources and provides contextual background for the study. Section 4 presents our empirical methodology and main results, along with a comprehensive set of robustness checks. In Section 5, we explore the underlying mechanisms driving the observed effects. Section 6 concludes.

2 Background and Data

2.1 Context

On February 6, 2023, the southeastern part of Türkiye experienced two devastating earthquakes, of 7.5 and 7.8 magnitudes, one of the largest earthquakes in the world, and in the country's history. Figure A2 shows where Maras Quakes is placed in the distribution of earthquakes occurred since 1900. The map in Figure A3 shows the scope of the affected area, where shaking was felt in a wide area, covering 350,000 square kilometers. Effectively (as also reported by government agencies), the earthquakes affected 11 provinces and around 13 million people, corresponding to around 15 percent of the population. The earthquakes led to significant infrastructural damage and caused more than 50,000 fatalities. The damage being widespread, around 227,000 housing units were damaged or destroyed.[5] The earthquakes in addition caused the displacement (both temporary and permanent) of more than three million people. The estimated cost of the earthquakes is around 104 billion dollars.[6] While the majority of those who migrated returned back to

[4] Other related papers on the importance of noncognitive skills include Kreft et al. (2021). Some other studies explore mechanisms like parental inputs during post-shock periods, for example, see Inui and Okudaira (2022).

[5] Source: Bloomberg HT.

[6] Source: SBB. In addition, the rate of spending is predicted to be around 4 percent of the 2023 GDP. The following years the rate is expected to be around 2.5 percent. (Source: Bloomberg HT.)

their hometowns, the permanent migration, according to TURKSTAT statistics, is around 14 percent of the population in the affected provinces.

After the earthquake, education was one of the most severely affected areas, alongside the labor market. The disruption took many forms. Most directly, the regular schooling schedule was suspended. Students in the affected regions experienced a pause in education that lasted from two weeks to three months, depending on the severity of the quake.[7] Even after schools technically reopened, many students in the region did not resume full classroom hours or standardized exams, and both students and teachers faced large-scale displacement. In practice, this meant that students lost nearly an entire semester, typically running from February to June.

Open-ended responses from our survey suggest that, for many families, the immediate post-quake period was simply too overwhelming to focus on education. Those who relocated to other cities often had the chance to continue schooling there,[8] but for many who remained, educational access remained very limited. Figure A1 shows the different paths students might have taken after the earthquake and gives a sense of how the disruption played out in practice.[9] Many students experienced abrupt and significant changes in their learning environments, and struggled to maintain their usual learning habits in the post-earthquake period.

2.2 Survey Data

The data for this study come from a survey, designed in-house, conducted in provinces with differential earthquake impact, and complemented by government-provided data on earthquake destruction. The survey was conducted in May 2024, approximately fifteen months after the earthquake. We recruited students and their parents by reaching out to high schools in the provinces of Adıyaman (High Impact), Malatya (Medium Impact), and Elazığ (Low Impact). To encourage participation, we offered incentives: each student who completed the survey received 200 Turkish Liras if their parent also completed it, and 100 Liras if not. The incentive amounts were chosen based on two considerations: first, ensuring that compensation aligned with minimum wage standards for the time required; and second, feedback from a pilot survey with a small group of students and parents regarding their willingness to participate and the minimum amount they would expect. These considerations guided the final incentive structure.

2.2.1 Study Sample and Recruitment. The sampling closely follows the framework to identify the causal effect of noncognitive skills on students' academic performance.

[7] Nationwide, schools closed for two weeks, while in the most impacted areas, this period extended to three months.

[8] This option existed, but not all students were able or willing to enroll in new schools after moving.

[9] In host regions, many public and private schools, as well as tutoring centers, offered free education to affected students. In the disaster zone, the government set up temporary education centers and psychosocial support units. However, based on our field visits and conversations with local officials, these services were mostly limited to temporary housing areas and did not have a consistent presence in cities.

Location Choice— The goal with the location choice is to include a high, medium, and low earthquake impact region, where the low-impact region also serves as a control that was not meaningfully affected by the earthquake.[10] These regions are also selected to be comparable in terms of socioeconomic characteristics to help control for confounding factors unrelated to earthquake exposure. The definition of impact comes from observable measures of damage: the destruction rate (i.e., the share of unusable housing units), the death rate, and the injury rate. Table A1 shows these rates for the 11 affected provinces. Based on this information, we create a composite index that combines the three indicators.[11]

Provinces are then clustered into three groups, high, medium, and low impact, based on this composite index. Instead of mechanically assigning terciles, we use threshold values of 0.15 and 0.6, which reflect clear breaks in the empirical distribution of the index. This approach avoids misclassifying provinces like Malatya, which had high levels of destruction but comparatively moderate death and injury rates. From these clusters, we select three provinces for the study. Adıyaman is chosen as the high-impact region, given its severity across all dimensions and its similarity to the other regions in terms of socioeconomic background. Malatya, with high destruction but lower human toll, represents the medium-impact region. Elazığ is selected as the low-impact (control) region. It had the lowest composite index and was not meaningfully affected by the earthquake, while also being comparable to Malatya and Adıyaman in terms of underlying socioeconomic conditions.[12][13] In the remainder of the paper, we refer to these provinces as HIA (High-Impact Area), MIA (Medium-Impact Area), and Control (Low-Impact) provinces.

School and Grade Selection— In Turkey, high schools are broadly categorized into two types: exam schools and non-exam schools. Exam schools require students to take a standardized entrance exam and are generally considered higher quality, as they admit students based on academic performance. Non-exam schools, by contrast, do not require an entrance exam and are more accessible, though they typically prioritize students residing within the local neighborhood. Within both categories, there are narrower subtypes

[10] Note that the pre-analysis plan for this project included a pure control province, Kayseri. However, due to the survey timing and logistical reasons, the survey there was conducted fully online, which resulted in a very low take-up rate. Because of concerns about unhealthy data collection, we ended up dropping these observations and treat Elazig as the low-impact Control province.

[11] The composite index is the average of min-max normalized destruction, death, and injury rates. Each variable is scaled to the $[0, 1]$ range before averaging, so that no single component dominates the index due to differences in magnitude. This yields a unitless measure of relative impact severity across provinces.

[12] Note, however, that Elazığ and Malatya experienced a previous earthquake in 2019, which may have influenced how residents responded to the 2023 earthquake. We address this concern by controlling for pre-existing covariates in the analysis.

[13] Another important consideration in the selection of provinces is that the analysis is restricted to schools located in the city centers, as they account for the majority of enrollment. As a result, we focus only on the severity of earthquake impact in the central districts of each province. For example, while the overall destruction rate in Gaziantep is similar to that of Elazığ, and its death rate—at 0.18%—is close to Malatya's, the most severely affected areas in Gaziantep are two rural districts outside the city center. The central part of Gaziantep was largely unaffected by the earthquake, and thus does not qualify as a high- or even medium-impact region within the scope of this study.

such as Science, Social Sciences, Anatolian, and Vocational schools.[14]

In this study, we focus only on Anatolian and Science high schools, as they are the most common across provinces and more comparable in terms of curriculum and student composition. These schools also tend to attract students with relatively similar educational aspirations, making them more homogeneous with respect to future academic trajectories.[15] Additionally, we limit the sample to schools located in the city centers of the selected provinces, where the urban student population is more comparable across regions in terms of socioeconomic background. To ensure a representative sample, we classify schools in each category (exam and non-exam) into three performance tiers and select one school randomly per tier, resulting in approximately six schools per province. When originally selected schools declined participation, we replaced them with similar schools based on cutoff scores, in consultation with the local Department of Education. In total, we survey 18 high schools across the three provinces, 5 HIA, 7 MIA, and 6 Control. Full details on the selection procedure are provided in Appendix B.1.1.

Regarding grade level, we focus on 9th-grade students, as they were in a critical transition period between middle and high school at the time of the earthquake. This cohort is particularly relevant for studying the earthquake's impact on academic performance, as they had recently taken the high-stakes centralized exam at the end of 8th grade, which determines high school placement. The standardized nature of this exam provides a consistent measure of academic performance across schools.[16]

In each school, 9th-grade enrollment varies, but the average is approximately 120 students per school (roughly four classrooms, with 20–30 students per classroom). In lower-performing schools, class sizes tend to be larger. In schools with many students, we surveyed three or four classrooms, depending on overall enrollment and logistical feasibility.[17]

Implementation— Student surveys were conducted in person using an online platform in a classroom setting. On the scheduled survey day, the research team visited each school. The school administration informed students in advance and instructed them

[14] Science and Social Sciences schools are typically exam schools, with the former emphasizing science and mathematics and the latter focusing on social sciences. Anatolian schools are non-exam schools that provide a general academic curriculum with an emphasis on language education, while Vocational schools focus on occupational training.

[15] For example, one most common vocational high school is religious high schools, and resilience could be driven much differently there. The values taught there could boost student behavior much differently than a regular education.(Siddique, 2024)

[16] An alternative would have been to focus on 12th-grade students, who also take a major standardized test, the university entrance exam. However, during the study period, most 12th graders (of the earthquake cohort) had already graduated from high school and were likely placed in university, making them logistically difficult to reach.

[17] In four schools, only a subset of classrooms was surveyed. Classroom selection varied by context: in one case, we selected three classrooms located on the same floor to minimize logistical challenges such as unstable internet access; in others, we followed a random selection of classes. One concern with this is that if there are tracking policies in place, students might differ systematically, however, as told by the school administration, there were no tracking policies in place for the schools with specific classroom selection.

to bring their phones.[18] For students without phones, the research team provided devices on-site. To ensure uninterrupted survey access, the research team brought mobile hotspots, backup devices, and chargers.

Although computer labs were considered, they proved infeasible: many schools lacked functioning labs, and those available were too small (typically 15–20 seats) to accommodate the average class size. Rotating students through multiple lab sessions would have significantly extended the data collection period beyond the timeframe approved by the Department of Education.[19]

This hybrid model, using online tools in a proctored classroom environment, combined the advantages of both in-person and digital data collection. The in-person setup reduced distractions and helped create a focused environment, improving data quality and response rates. The online platform (Qualtrics) allowed us to implement real-time quality controls, such as preventing random clicking or overly fast completion. These features were explained to students at the outset, reinforcing the importance of thoughtful participation and qualifying for the incentive. The exact classroom script is included in Appendix B.2.2.

In a small number of cases, students completed the survey at home. This occurred either because students preferred not to use school-provided devices for privacy reasons or due to technical constraints (e.g., limited devices or connectivity issues). Follow-up reminders were issued to ensure survey completion.[20]

The parent survey was conducted online. On the same day students completed their survey, parents received a survey link via the school administration and were asked to fill in within two days. The parental communication template is included in Appendix B.2.3. A contact number was shared for technical support. Students whose parents had not completed the survey received up to three reminders: two from the school administration and one from our team, using contact details provided by the students. The final dataset includes 1,645 students and 1,306 parents. Of these, 1,152 are matched student-parent pairs. Table 1 presents summary statistics for some variables, broken down by province. One potential concern with our sample is that migration and displacement after the earthquakes may have influenced who was available to participate in the survey. Although some permanent migration did occur, the summary statistics indicate that the sample remains broadly representative of the affected population.[21]

[18] Prior to data collection, notices were distributed to students and parents via school administrators.Students and parents were informed that participation was voluntary, could be withdrawn at any time, and were given contact details for follow-up or opt-out requests. Approximately 78% of schools allowed students to bring their phones, provided they were kept in lockers outside class hours.

[19] While using phones in the classroom poses potential challenges such as distractions from social media or other apps. We addressed this by implementing in-class proctoring and informing students that the survey platform detects response quality and attention. Issues related to internet access or battery life were mitigated by the research team through the use of mobile hotspots, backup devices, and chargers.

[20] In the robustness analysis, we control for whether the survey was completed in the classroom or at home, allowing for potential differences in response behavior.

[21] For example, according to TURKSTAT (2023), provincial GDP levels prior to the earthquake in 2022 were similar in Control and MIA provinces, while HIA had a lower level. That is, HIA's lower income levels are not only driven by the post-earthquake migration selection.

Table 1: Summary Statistics Table

	All	**HIA**	**MIA**	**Control**
Female	0.53	0.55	0.52	0.52
	(0.01)	(0.02)	(0.02)	(0.02)
Public School (Pre-Q)	0.88	0.94	0.85	0.83
	(0.01)	(0.01)	(0.01)	(0.02)
Household Income (Pre-Q)	3.39	3.31	3.43	3.39
	(0.05)	(0.1)	(0.08)	(0.1)
Home Damage	1.13	1.57	1.31	0.65
	(0.03)	(0.05)	(0.04)	(0.04)
School Damage	0.99	1.23	1.21	0.51
	(0.03)	(0.06)	(0.04)	(0.05)
Tutoring (Pre-Q)	0.52	0.49	0.52	0.6
	(0.01)	(0.03)	(0.02)	(0.02)
No of Movements (Post-Q)	1.14	1.5	1.57	0.39
	(0.03)	(0.07)	(0.05)	(0.04)
N	1642	404	663	456

Notes: The table reports summary statistics for the full sample and by province. Reported values are means, with standard errors in parentheses. "Pre-Q" refers to variables measured before the earthquake (e.g., school type, household income, tutoring), while "Post-Q" refers to variables measured after the earthquake (e.g., number of movements). Household income is measured on a standardized scale from the survey. Home and school damage variables indicate the average reported intensity of earthquake damage.

2.2.2 Survey Structure. Figure A4 illustrates the flow of the student and parent surveys. In this section, we describe the main content of the key survey blocks. These blocks are designed to serve two purposes: first, to identify the causal impact of noncognitive skills on educational outcomes in the aftermath of the earthquake; and second, to explore the potential mechanisms through which these effects operate. The survey includes the following blocks:

Background Information: For students, this block includes questions about their demographic information such as age, gender, which school they attend, and their current residence. For parents, this block includes additional questions about their education, occupation, and household income.

Earthquake Exposure Experience: This module includes questions about the earthquake exposure of the respondent. For students, we ask about their experience regarding their post-earthquake living conditions, migration status, and school disruption. For parents, we ask additional questions about their household's employment status, job loss, income and property loss.

Mental Health: This module aims to get information on the mental health and psychological well-being of the respondents. We use validated scales to measure the well-being of the respondents post-earthquake. The measures come from General Anxiety Disorder

(GAD-7) and CAPSAW scale.[22] The parental survey includes questions about both the parents and their children's mental health.

Education and Test Outcomes: This block asks students about their educational outcomes such as their grades before and after the earthquake and high school entrance exam test scores after the earthquake. Note that our sample consists of schools of different quality. Some of those schools are non-exam schools, for which there is no requirement to take high school entrance exams. While for these groups of students, the high school entrance exam scores are missing, we address this issue by using the Fall 2023 grades on Mathematics and Turkish Literature, which are comparable across schools up to a considerable degree. The Department of Education initiated a new program for 9^{th}-grade students: students take the same Mathematics and Turkish Literature exams. While the grading is done by the teachers in the schools, there is a certain level of standardization in the grading process. This eases the comparison of the students' performance across schools even in the absence of high school entrance exam scores.

Time-Use: Following the literature that uses time diaries, we ask students and their parents about students' time use before and after the earthquake. We ask questions about how much time they spend studying and other activities such as hobbies, social media, and watching TV/video. Using these questions, we construct a behavioral measure of noncognitive skills.

Noncognitive Skills: This module aims to collect information regarding the noncognitive skills of the students before and after the earthquake. The measure of noncognitive skills is based on a set of questions validated in other studies. Specifically, we use the CDRS resilience scale, and Duckworth's grit scale.

Parental Inputs: This module focuses on parental involvement in students' education. Parents are asked about the extent to which they supported their child's exam preparation after the earthquake, including specific forms of support such as helping with homework, providing a quiet study environment, and offering encouragement. In addition, students are also asked parallel questions about their perceptions of parental involvement.

Social Network: In this section, questions aim to understand the potential impacts of the social network loss post-earthquake on the students' general well-being and academic performance.

New Study Environment: As explained above, many students had to change their study environment due to the earthquake. For example, some students stayed with ten other people in the same home. This module asks students about their new study environment and how it affected their study habits and performance.

[22] The CAPSAW scale includes items such as *"Are you happy in general?", "Do you think people care about you?", "Do you feel safe in general?", "Can you do the things you want to do in your life?", "If you have a problem, can you find a way to deal with it?", and "Do you think you are helpful to other people?".* Responses are recorded on a 1-5 Likert scale where 1 represents "Always" and 5 represents "Never". We reverse the scale coding so that higher values indicate stronger mental health. We then calculate the mean response across all items to create a composite index for mental health assessment.

2.2.3 Survey Data Quality. The survey was designed to ensure high data quality. The time spent distribution plots in Figure A5 and Figure A6 show that most students and parents completed the survey within a reasonable time frame, suggesting they were engaged and attentive throughout the process. We also implemented several quality checks to assess the reliability of the data; the results of these checks are summarized in Table A3. The majority of respondents passed the attention checks. One potential concern regarding data quality is that many questions asked about both the pre- and post-earthquake periods. To account for this, we asked respondents to rate their recall confidence on a scale from 1 to 5, where 1 indicates low confidence and 5 indicates high confidence. Most respondents reported high confidence in their recall. Finally, the majority also reported putting in high effort when completing the survey, around 4 on a scale from 1 to 5, where 1 indicates low effort and 5 indicates high effort. Although data quality indicators do not significantly vary across quake intensity levels, we still conduct robustness checks to ensure the results are not driven by potential quality issues.

2.3 Earthquake Exposure Measure

In the main analysis, we use two measures for treatment. The first measure is a province dummy variable indicating whether the student is from a high- or medium-impact province. The second measure is the destruction level of the earthquake in the neighborhood where the student lives. After the earthquakes, state officials inspected every building in the affected provinces and assigned a destruction level to each building. Citizens could view the status of their building through an online e-government portal. We used this portal to scrape building-level data and calculated the destruction level for each neighborhood. The destruction level is a categorical variable ranging from 0 to 4, where 0 indicates no damage and 4 indicates complete destruction.[23] Figure A12 shows the geographical distribution of destruction levels in the affected provinces we study.

3 Descriptive Facts

This section provides a diagnosis of the physical and mental disruptions to a student's life caused by the earthquake using the survey data.

[23] According to the scraped data, the number of evaluated buildings is 196,019 for Malatya (compared to a total of 178,987 registered buildings reported by the government), 128,782 for Adıyaman (compared to a total of 120,496 registered buildings), and 66,317 for Elazığ. While the number of evaluated buildings for the first two provinces approximately matches the official registered numbers, it slightly exceeds them. The reason for this discrepancy could be due to the inclusion of unregistered buildings or differences in data update cycles. In contrast, the number for Elazığ is significantly lower than the government-reported figure of approximately 123,713. This gap might be explained by Elazığ being less affected by the earthquakes, resulting in fewer inspections. See "2023 Kahramanmaraş and Hatay Earthquakes Report" (n.d.) for details.

3.1 Physical Disruptions

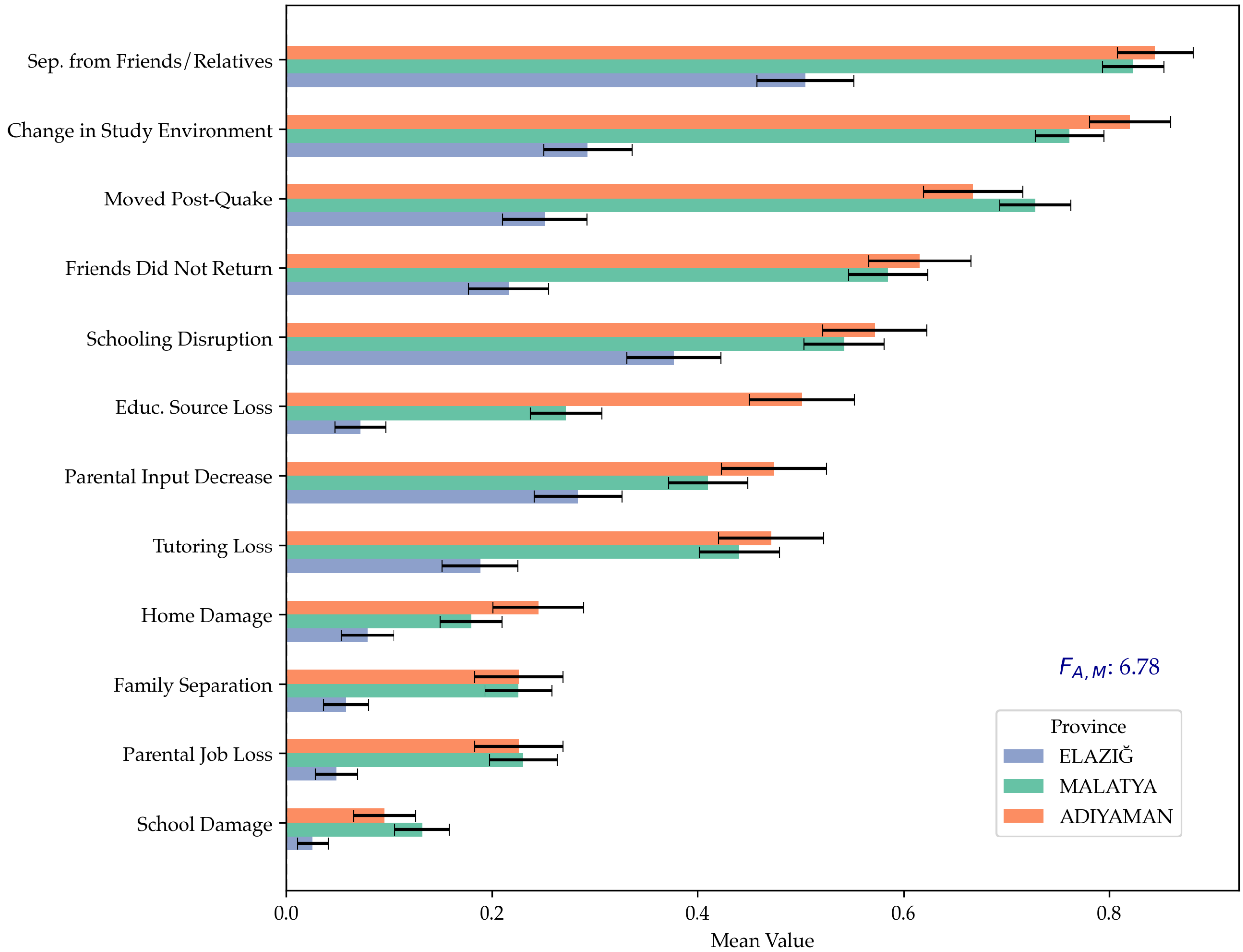
Figure 1 shows the physical disruptions caused by the earthquake across three locations with low, medium, and high levels of destruction. First, we observe a large impact on students' environments, especially in the high destruction area. More than 80% of the sample from this region report disruptions in their study environment and network access. This also includes separation from family and friends, and a general breakdown in the stability of their surroundings. Second, relocation after the earthquake is quite common in the medium and high destruction areas. Around 70% of the sample in each of these areas report having relocated, but interestingly, the relocation rate is actually higher in the medium destruction area than in the high destruction area.[24] In the high destruction area, many students also report losing access to educational resources, and about 60% of the population reports experiencing school disruption. It's worth noting that the low impact region also saw some level of school disruption, which could be explained by the nationwide school closure policy that was implemented after the earthquake.[25] In both high and medium impact areas, students also report being physically separated from their parents — often because the father stayed behind for work while the rest of the family relocated to another city.[26] Home damage is reported by more than 20% of respondents in the high destruction area and close to 20% in the medium destruction area. School damage, on the other hand, is reported less frequently, likely due to strict building codes in place. Overall, these disruption patterns confirm that the earthquake had its most severe impact in HIA. A joint F-test comparing the medium and high destruction regions shows that the difference in means is statistically significant, with $F = 6.78$. Note that these numbers are purely descriptive and might partly reflect the sample selection due to post-earthquake migration patterns, which we address in the robustness section.

[24] Relocation can happen both within and across cities, as reported in the survey.

[25] Since the survey captures the timing of school disruption, we can check whether these disruptions align with the earthquake or the national policy.

[26] For those who report family separation, we can use the parental survey to identify the reasons for staying behind.

Figure 1: Physical Disruptions by Province



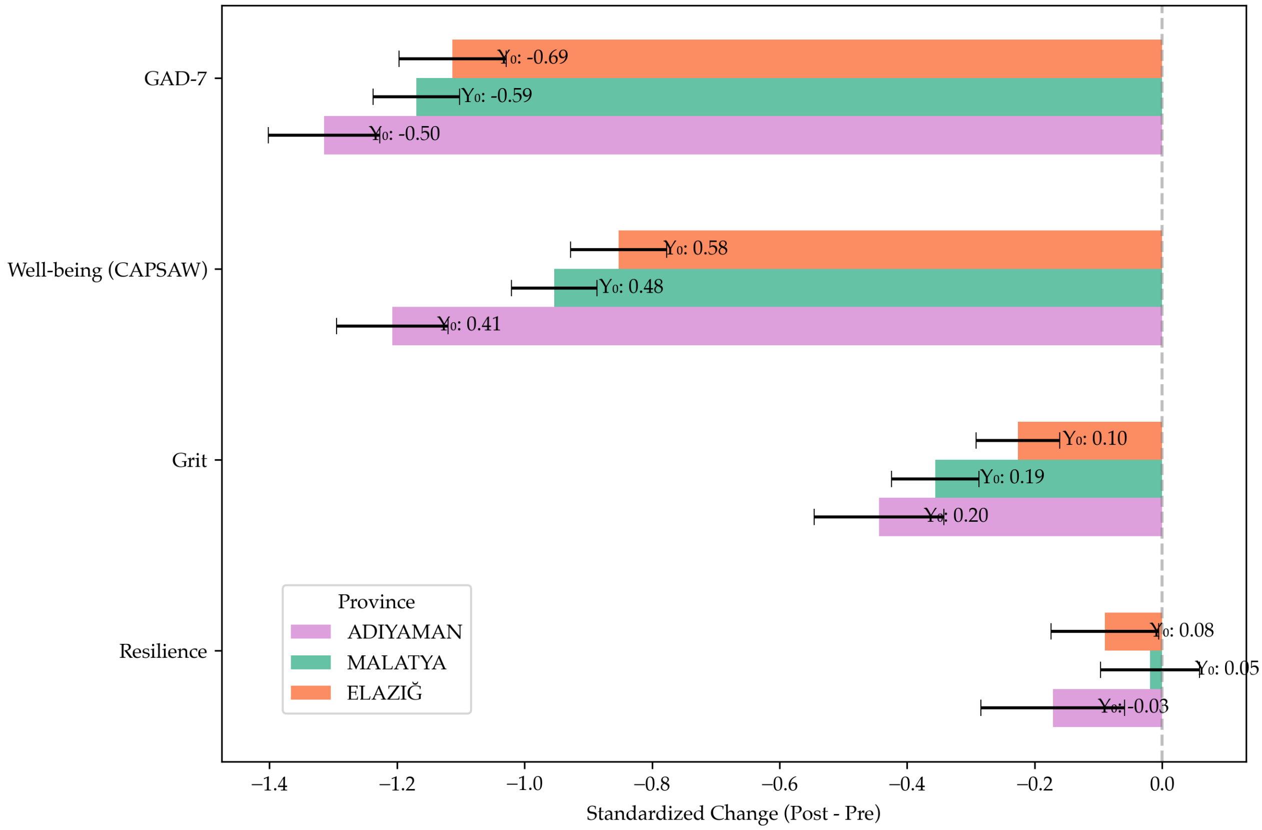
Notes: This figure shows the physical disruptions caused by the earthquake. Horizontal bars are 95% confidence intervals. All variables are in dummy form.

3.2 Mental Disruptions

Figure 2 shows the mental disruptions caused by the earthquake across three locations with low, medium, and high levels of destruction. We present results across four indices. The first is the GAD-7 index,[27] which measures anxiety. The second is a well-being index constructed from CAPSAW modules. The third and fourth are the grit and resilience indices, respectively. The figure suggests that the earthquake had a significant impact on students' mental health, with the largest disruptions observed in the high destruction area. Importantly, the plot also includes the initial (pre-earthquake) values of these indices, as reported by students (denoted by Y_0). Across the three locations, students in the high destruction area had the lowest initial levels. The empirical analysis accounts for these baseline differences. Among the non-cognitive skills, grit and resilience, we observe modest declines in the high destruction region.

[27] In the original Likert scale, a higher GAD-7 score indicates higher anxiety. For comparability with the other indices, we reverse-coded it so that higher values reflect better outcomes.

Figure 2: Mental Disruptions by Province



Notes: This figure shows the mental disruptions caused by the earthquake. Horizontal bars are 95% confidence intervals.

3.2.1 LLM Analysis of Open Ended Questions In the survey, we included open-ended questions to better understand students' feelings and thoughts about the earthquake that might not be captured by structured survey questions. As noted by Haaland et. al. (2024), open-ended questions have recently gained popularity in economics literature, particularly in development and labor economics, as they effectively elicit top-of-mind responses.[28] We employ LLM-supported text analysis to analyze these responses, which allows for a systematic labeling of the students' experiences and emotions.[29] Details regarding the LLM implementation are provided in Appendix C. Table 2 presents the descriptive results of the LLM analysis by province. Accordingly, the biggest disruptions both psychological, social, academic, and material are observed in the high destruction area, HIA, with more than 60% of the respondents reporting psychological disruptions, 20% reporting social disruptions, and 8% reporting academic disruptions, and 19% reporting material disruptions, which includes losses like property or income.

[28] Although the open-ended questions were placed at the end of the survey to prevent cognitive overload in this lengthy instrument, approximately 90% of students still provided responses despite these questions being optional. Additionally, since the survey was conducted in proctored classroom settings, the concern of LLM-generated answers is minimal compared to other studies.

[29] While manual coding is an option, it becomes impractical at scale and subject to researcher fatigue and bias. We instead, provided the LLM with a structured prompt fine-tuned for this and asked it to classify the survey responses into predefined categories.

Table 2: Open Ended Text - LLM Assigned Labels

	HIA		MIA		Control	
	Mean	SD	Mean	SD	Mean	SD
Psychological	0.664	0.473	0.557	0.497	0.511	0.500
Social Disruption	0.202	0.402	0.177	0.382	0.075	0.264
Academic	0.086	0.281	0.079	0.269	0.048	0.215
Physical/Material	0.193	0.395	0.091	0.287	0.044	0.205
Minimal Impact	0.081	0.274	0.181	0.386	0.315	0.465
Positive Growth	0.064	0.245	0.063	0.244	0.051	0.220
N	405		662		454	

Notes: Mean and standard deviation of LLM-based open-ended text labels across selected provinces. Each label is a binary indicator for whether the response was assigned that label. Multiple labels can be assigned to a single response.

Additional descriptive statistics about the earthquake's impact on student outcomes can be found in Appendix D.

4 Empirical Analysis

The goal of this paper is to quantify how non-cognitive skills shape the impact of earthquakes on students' post-quake outcomes, such as academic performance and mental health. Do students with higher grit or resilience fare better in the aftermath of an earthquake? To answer this, we define treatment, δ, as an increase in exposure to the earthquake, and examine how treatment effects vary by students' non-cognitive skill levels. The main empirical specification is:

$$Y_i = \alpha + \beta_1 \delta_i + \beta_2 NC_i + \beta_3 \delta_i \times NC_i + \mathbf{X_i \Gamma} + \epsilon_i \tag{1}$$

where Y_i is the outcome of interest for student i, δ_i is the earthquake exposure measure, NC_i is the non-cognitive skills measure, and $\mathbf{X_i}$ is a vector of pre-earthquake controls. This framework is essentially a difference-in-differences model. The coefficient β_1 captures the average effect of earthquake exposure, while β_2 reflects differences in outcomes by non-cognitive skill level. The key parameter, β_3, measures how the effect of earthquake exposure varies with non-cognitive skills, that is, whether students with higher grit or resilience are more protected against the negative impacts of the earthquake. The identification strategy leverages variation across provinces: Control province serves as the control, while HIA and MIA provide variation in exposure. This design allows me to estimate β_3 credibly, though several identification challenges remain, as discussed below.

4.1 Identification Challenges and Assumptions

The first threat to identification is potential correlation between pre-quake neighborhood quality and academic success. One might argue that better-off places might suffer less from the earthquake and also have better schools. Figure A11 shows the relationship between the neighborhood destruction level and household income. First, we observe that poorer neighborhoods clearly suffered more from the earthquake, in the impacted provinces. To address any concern, our baseline regressions control for household income and pre-earthquake academic and non-cognitive skill measures.

The second threat is migration-related. Our data includes only those who stayed and those who returned to their hometowns after temporary displacement. Permanent migrants are missing. If the decision to migrate is correlated with potential outcomes, OLS estimates will be biased. To put it into perspective, MIA's population decreased by 8.6% and HIA's by 4.8% in the year following the earthquake.[30]

Now, the question is whether those who stayed or returned are systematically different from those who did not. To address this, we included a set of questions in our survey asking our available subgroup about the people they know who never returned. First, note that we explain our main results for the available subgroup only; however, the full cohort ATE might differ. For example, those who moved might have different non-cognitive skills than those who stayed. To address this concern we included detailed questions about migration in our survey. We then use the available information to establish bounds on the full cohort ATE. Section 4.6 documents the migration issue and provides robustness checks as well as bounds on the full cohort ATE.

Third, our measures are inevitably subject to measurement error, specifically in the form of recall bias. Our survey asks students to self report their pre-earthquake, post-quake and current (the survey time) mental health, skills, study times, and scores. Students might incorrectly recall their pre-earthquake levels. If this differs by treatment status, it can bias our estimates, otherwise it can be treated as a classical measurement error. In the robustness checks, we address the recall bias concern with several strategies.

4.2 Measurement

4.2.1 Behavioral Grit/Resilience Measure Equation (1) requires a measure of students' non-cognitive skills, which in this study we interpret as behavioral resilience or grit. We proxy these traits using the change in students' self-reported weekly study hours before and after the earthquake. Students who sustain their study time, or exhibit smaller declines relative to peers, are considered more resilient. Our approach parallels that of Chen et al. (2024), who rely on a behavioral grit index collected via a web-based platform.[31]

[30] The numbers are from the Turkish Statistical Institute. Note that before quakes, MIA's net annual out-migration rate was 0.4% and 0.7 for HIA. Note that Control area recorded an increase in population by 2.2%, due to incoming migrants from the affected provinces.

[31] While we collected a rich set of self-reported psychological trait measures prior to the earthquake, we ultimately rely on our behavioral proxy. This choice is motivated by both empirical and conceptual

However, unlike their pre-shock measure, ours is constructed from both pre- and post-earthquake data, which raises a natural concern that the measure may itself be affected by the treatment, as observed in Figure A9. To address this, we residualize the raw change in study time, $\Delta \text{StudyTime}_i$, by regressing it on province-level earthquake exposure (δ_i) and a rich set of pre-quake controls X_i, including academic performance, study habits, and household inputs. The resulting residual,

$$\hat{r}_i = \Delta \text{StudyTime}_i - \hat{\mathbb{E}}\left[\Delta \text{StudyTime}_i \mid \delta_i, X_i\right],$$

captures variation in students' post-quake academic effort that is orthogonal to both the intensity of exposure and pre-existing characteristics. We standardize this residual to have mean zero and unit variance, and refer to it as our behavioral grit/resilience measure, NC_i. While NC_i is not a direct psychological measure of grit,[32] it serves as a revealed-behavior proxy for students' ability to sustain effort under adversity. Our empirical analysis focuses on the *heterogeneity* in earthquake impacts by NC_i, rather than on the average treatment effect alone.[33]. Our causal interpretation relies on the assumption of selection on observables, namely, that conditional on δ_i and X_i, the residual variation in NC_i is not confounded by unobserved determinants of test scores.

We conduct two diagnostic tests to assess the validity of our behavioral resilience proxy. First, we examine its correlation with pre-quake characteristics such as middle school GPA, the pre-quake resilience index, and province dummies and find all correlations to be near zero, as expected mechanically by construction. This supports the interpretation that the proxy is orthogonal to pre-determined observables. Second, we estimate a placebo version of equation (1) using middle school GPA (standardized) as the outcome. We find that neither the coefficient on NC_i nor its interactions with province dummies are statistically significant.[34] These tests help rule out confounding from pre-existing academic performance.

4.2.2 Self Reported Mental Health and Non-cognitive Skills To measure mental health and non-cognitive skills, we include several self-reported pre-earthquake measures collected via standardized psychological scales, as detailed in Section 2.2.2. For analytical tractability, we construct indices for each psychological trait using principal component analysis (PCA), taking the first component as the index value for each trait. Table A6 presents the factor loadings for each item in these psychological scales, along with their

considerations. Empirically, we find that self- and parent-reported non-cognitive measures exhibit weak correlations with academic outcomes in our setting (e.g., < 0.05 for grit and resilience with exam scores; see Table A8). These findings are consistent with prior work highlighting the limitations of self-reported measures (e.g., Bertrand and Chen et al., 2024) and the value of behavioral proxies. Conceptually, while our measure is survey-based, it is arguably more objective in nature, as it captures quantitatively revealed effort.

[32] For example, a direct measure would involve survey instruments such as Duckworth's grit scale.

[33] Our residualization strategy follows the framework in Chetty et al. (2014)

[34] The coefficient on NC_i is -0.009 with a p-value of 0.86. The interaction coefficients are -0.04 ($p = 0.57$) for HIA and -0.07 ($p = 0.24$) for MIA.

corresponding eigenvalues. These indices serve among the control variables in our regression specifications to account for pre-existing differences in psychological attributes.

4.3 Main Results

We focus on two main post-earthquake outcomes: academic performance and mental health. By post-earthquake, we mean the semester right after the disaster (Spring 2023) until the centralized high school entrance exam.

4.3.1 Academic Performance Table 3 presents the estimates from regressions of standardized exam scores on our behavioral non-cognitive skills measure, NC_i, the residualized study time change ΔST, and its interaction with earthquake exposure, proxied by province dummies. A full set of controls is included. Columns (1) in each subject panel show results using only student-reported controls, while Columns (2) include additional parent-reported controls, such as pre-quake household income, which explains the smaller sample size in those columns.

Across all subjects (Math, Turkish, Science, and History) we find a strong positive association between the behavioral non-cognitive skills measure and exam performance. That is, students who reduced their study time less performed better on the exams, controlling for pre-quake study time. A one standard deviation increase in the behavioral non-cognitive skills measure is associated with a 0.19 SD increase in Math scores, 0.11 in Turkish, 0.18 in Science, and 0.08 in History.

We also find meaningful heterogeneity by earthquake exposure, which is central to our analysis. The returns to staying resilient appear higher in the high-destruction province of HIA compared to the medium-destruction province of MIA. The interaction coefficients $\Delta ST \times$HIA range from 0.11 to 0.19 standard deviations, while those for $\Delta ST \times$MIA range from 0.06 to 0.15. The province dummies, relative to the baseline, show negative point estimates. The home damage indicator is also negatively associated with scores, ranging from –0.06 to –0.10 SDs. Other controls like middle school GPA and baseline study time behave as expected, showing positive associations with exam scores. Our regressions also control for whether the student temporarily relocated after the earthquake.

4.3.2 Mental Health Table 4 presents the results of regressions of post-earthquake mental health outcomes on the behavioral non-cognitive skills measure, NC_i, and its interaction with earthquake exposure. As mental health outcomes, we consider two indices: the GAD-7 anxiety index and a well-being index constructed from the CAPSAW modules.[35] In all columns, we control for baseline well-being and anxiety scores, as well as a rich set of pre-quake controls, including gender and baseline academic performance measures. Columns (2) in addition include parental controls such as household income and parents' academic input on students' education. Across both outcome measures, there

[35] For interpretability concerns, we reverse the GAD-7 index so that higher values indicate better outcome rather than higher anxiety.

Table 3: Effect of Non Cognitive Skills on Exam Scores

	Math		Turkish		Science		History	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Mid School Gpa	0.06**	0.04	0.06**	0.04	0.06**	0.05	0.03	0.03
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Pre Quake Study Time	0.06***	0.06***	0.03**	0.03	0.04***	0.05***	0.04***	0.05***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
ΔST	0.19***	0.14**	0.11***	0.11**	0.18***	0.15***	0.08**	0.10*
	(0.04)	(0.05)	(0.04)	(0.05)	(0.03)	(0.05)	(0.04)	(0.05)
Home Damage	-0.07**	-0.06*	-0.10***	-0.09**	-0.06**	-0.06*	-0.07**	-0.06*
	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Moved Dummy	0.17***	0.17**	0.27***	0.28***	0.23***	0.24***	0.21***	0.16**
	(0.06)	(0.07)	(0.06)	(0.08)	(0.06)	(0.07)	(0.06)	(0.07)
Province HIA	-0.06	-0.04	-0.06	0.02	-0.10	-0.09	-0.23***	-0.17**
	(0.07)	(0.08)	(0.07)	(0.08)	(0.07)	(0.08)	(0.07)	(0.09)
Province MIA	-0.24***	-0.21***	-0.19***	-0.09	-0.25***	-0.22***	-0.34***	-0.27***
	(0.06)	(0.08)	(0.07)	(0.08)	(0.07)	(0.08)	(0.07)	(0.08)
ΔST × HIA	0.16***	0.19***	0.11*	0.11*	0.12**	0.11	0.15***	0.12*
	(0.05)	(0.07)	(0.06)	(0.07)	(0.05)	(0.07)	(0.06)	(0.07)
ΔST × MIA	0.13**	0.15**	0.11**	0.06	0.11**	0.13**	0.12**	0.07
	(0.05)	(0.07)	(0.05)	(0.07)	(0.05)	(0.06)	(0.05)	(0.07)
Constant	-0.06	-0.71***	-0.22***	-0.68***	-0.09	-0.57***	0.02	-0.41***
	(0.07)	(0.10)	(0.07)	(0.12)	(0.07)	(0.11)	(0.07)	(0.11)
Student Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parental Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	1403	1013	1403	1013	1403	1013	1403	1013
R^2	0.138	0.222	0.101	0.127	0.117	0.172	0.083	0.111

Notes: This table shows the regressions of exam scores.

is a strong positive association between ΔST and post-quake mental health outcomes. Turning to the interaction coefficients, we do not find statistically significant heterogeneity by earthquake exposure, however the direction of the coefficients is consistent with our hypothesis that the interaction term for high exposure is larger than that for medium exposure. While the main effects of medium and high exposure are negative compared to baseline, the interaction terms are small and insignificant. These findings suggest that in high-impact areas, the psychological burden of the disaster may be too large for the behavioral resilience alone to offset. That is, while academic performance is more responsive to resilient behaviors, mental health might require different coping mechanisms or support systems beyond sustained effort.

Table 4: Regression of Mental Health Outcomes

	Reversed GAD (Post)		**CAPSAW (Post)**	
	(1)	(2)	(1)	(2)
Mid School GPA	-0.08***	-0.07**	-0.05**	-0.05
	(0.02)	(0.03)	(0.02)	(0.03)
Study Time (Pre)	-0.03**	-0.04***	-0.03**	-0.03***
	(0.01)	(0.01)	(0.01)	(0.01)
ΔST	0.14***	0.22***	0.16***	0.18***
	(0.05)	(0.06)	(0.04)	(0.05)
Home Damage	-0.03	-0.02	-0.03	-0.04
	(0.03)	(0.03)	(0.02)	(0.03)
Moved Dummy	-0.14***	-0.14**	-0.07	-0.09
	(0.05)	(0.06)	(0.05)	(0.06)
HIA	-0.23***	-0.23***	-0.41***	-0.42***
	(0.07)	(0.08)	(0.06)	(0.08)
MIA	-0.01	-0.02	-0.11**	-0.11
	(0.06)	(0.08)	(0.06)	(0.07)
HIA $\times$ ΔST	0.03	-0.08	0.00	-0.04
	(0.07)	(0.08)	(0.06)	(0.08)
MIA $\times$ ΔST	-0.01	-0.11	-0.05	-0.06
	(0.06)	(0.07)	(0.05)	(0.07)
Constant	0.45***	0.32***	0.35***	0.21**
	(0.06)	(0.10)	(0.06)	(0.09)
Student Controls	Yes	Yes	Yes	Yes
Parental Controls	No	Yes	No	Yes
Observations	1430	1030	1430	1030
R^2	0.249	0.255	0.308	0.316

Notes: Higher values of the reversed GAD index correspond to better mental health (lower anxiety). The key predictor is ΔST: the residualized, standardized change in study time.

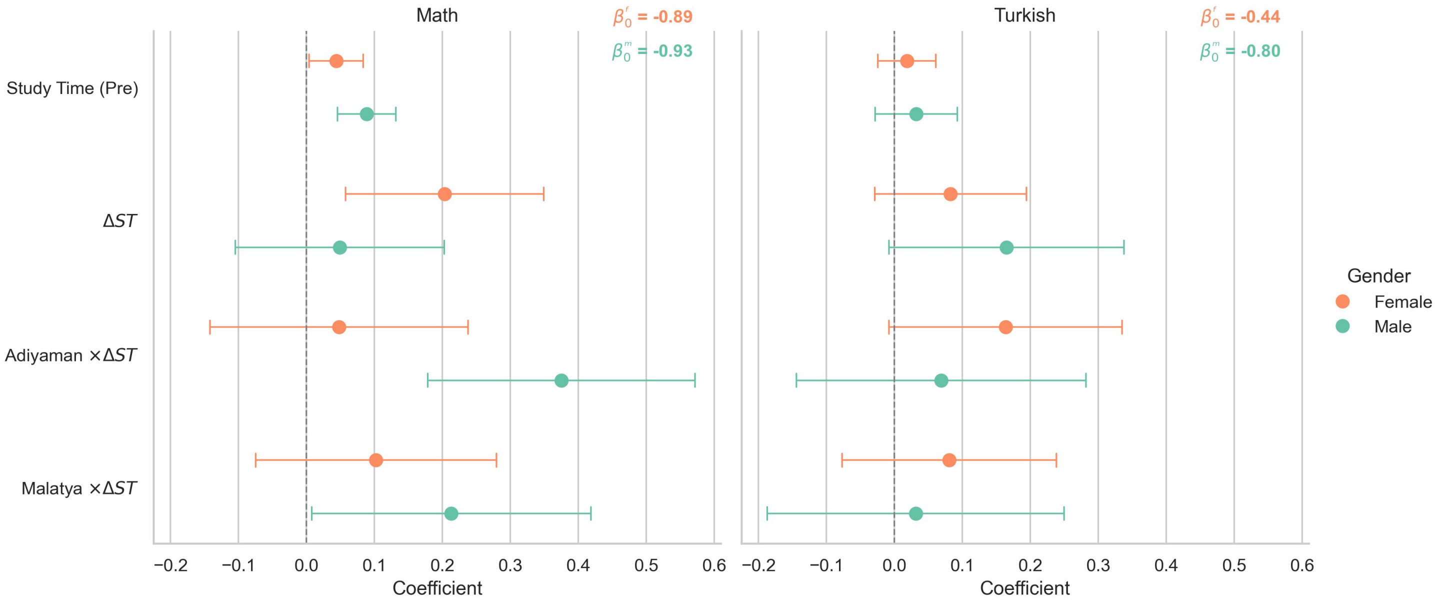
4.4 Heterogeneity Analysis

We explore whether the relationship between behavioral non-cognitive skills and post-earthquake outcomes varies by student characteristics. Specifically, we focus on two key dimensions: gender and pre-quake household income. The ability to cope with adversity may differ across demographic and socioeconomic sub-groups, and understanding these differences in detail can inform targeted interventions.

4.4.1 Student Gender Figure 3 presents coefficients from the main regression specification, estimated separately by gender. We focus on Math and Turkish subjects as representative measures of quantitative and verbal skills, respectively. All regressions include a rich set of student- and parent-level controls: baseline academic performance, parental income and academic input index, post-quake temporary relocation status, and baseline well-being indices. Full regression results are reported in Table A9. The key findings are as follows. Pre-quake study time is positively associated with exam scores for both genders, with slightly larger effects for males. The effect of behavioral non-cognitive skills captured by the change in study time is stronger for females in Math, and stronger for males in Turkish (possibly due to lower pre-quake baselines for males), with estimated effects of 0.19 and 0.15 standard deviations, respectively.[36] Among males, the interaction coefficients are larger in Math but statistically significant only in the high-impact region. For females, interaction terms are statistically significant across both subjects and regions, suggesting that girls benefit more consistently from behavioral non-cognitive engagement. These patterns imply that post-disaster recovery policies emphasizing behavioral skill-building may be particularly effective for the female student population.

[36] This finding is consistent with findings in the existing literature.

Figure 3: Academic Outcome Regression Results: Gender

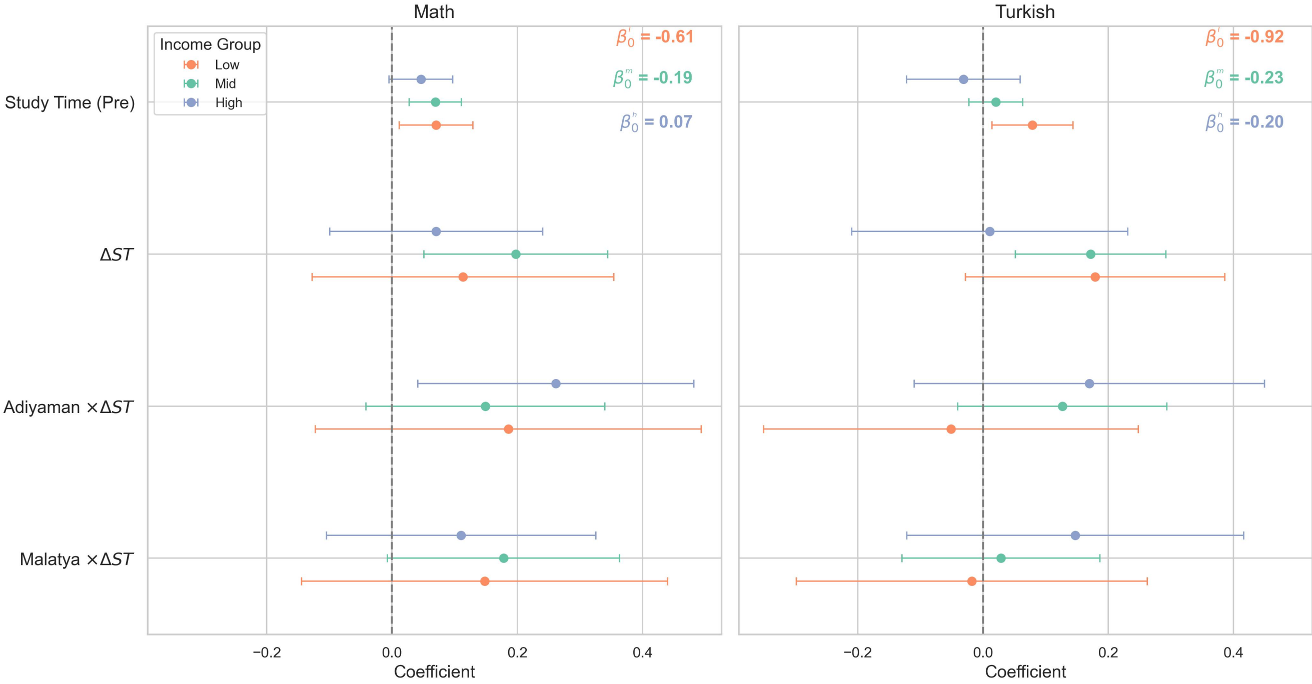


Notes: This figure presents coefficients from the main regression specification, estimated separately by gender. Horizontal bars are 95% confidence intervals.

4.4.2 Household Income. Figure 4 shows the heterogeneity of the main results across household income groups. We categorize households into three income groups: low, middle, and high, based on their self-reported pre-quake household income.[37] All regressions include a rich set of student- and parent-level controls as mentioned above. Full regression results are reported in Table A10. The key findings are as follows. The return to sustained effort is concentrated in the middle-income group, with a 0.18 standard deviation increase in Math and 0.16 in Turkish. Among the low-income group, Turkish responds more to sustained effort, while Math is less responsive. High-income students show minimal sensitivity overall. The interaction coefficients vary more across income groups. High-income students in high-impact regions benefit more in both subjects, especially in Math, with a 0.27 standard deviation increase. Among the low-income group, none of the interaction terms are statistically significant, though the Math coefficients are directionally consistent with the hypothesis that behavioral non-cognitive skills are more effective in high-impact areas. For Turkish, there is no clear pattern. For middle-income students, both interaction terms are positive and statistically significant in Math. For Turkish, while not significant, signs are aligned with the hypothesis. Overall, behavioral gains in earthquake areas appear most pronounced among middle-income students, after controlling for baseline characteristics.

[37] The income groups are defined as follows: low income is below 10,000 TL, middle income is between 10,000 and 30,000 TL, and high income is above 30,000 TL.

Figure 4: Academic Outcome Regression Results: Household Income



Notes: This figure presents coefficients from the main regression specification, estimated separately by household income groups. Horizontal bars are 95% confidence intervals.

4.5 Recovery Period Outcomes

The survey was conducted more than a year after the earthquake, allowing us to study its medium-term effects on students' academic performance and mental health. For academic outcomes, we use students' scores from the fall exams of the 2023–2024 academic year. In Turkey, the semester ends in January, meaning these results reflect outcomes nearly a year after the February 6 earthquakes. While we collected all fall exam scores, we focus on Math and Turkish in this section, as these are more comparable across students for two reasons. First, as part of a new policy, the Ministry of Education standardized the Math and Turkish exams for all 9th-grade students across Turkey.[38] Second, the survey explicitly instructed students to report accurate scores by checking their grades in the school management system.[39] For mental health outcomes, our survey asks about students' "current" mental health status (which was assessed at the time of the survey) corresponding to roughly sixteen months after the earthquake. We use the same GAD-7 anxiety index and CAPSAW well-being index as in previous sections.

Table 5 presents the results of regressions for our main equation using fall exam scores as the outcome variable. The results are similar to those from the earlier period following the earthquake. All signs and directions of the coefficients are consistent with the

[38] That is, all schools administered the same exam. However, some variation in grading may persist, as not all questions are multiple choice.

[39] On the survey screen, students were told: "Please log in to the online school portal at this moment and check your exam scores." Unfortunately, we were not able to cross-check the high school entrance exam scores since they are not readily available in the online system.

previous findings. The interaction terms suggest that the behavioral non-cognitive skills measure, ΔST, continues to be positively associated with exam performance, particularly in the high-impact region of HIA, with a 0.15 standard deviation increase in Math and 0.23 in Turkish. The interaction term for MIA is also positive but smaller in magnitude, 0.11 standard deviations in Math and 0.10 in Turkish, though the latter is not statistically significant. Table 6 reports the results of regressions on mental health outcomes. While students in the high-impact region continue to exhibit lower mental health scores, the interaction terms are not statistically significant. This suggests a limited role of behavioral non-cognitive skills in shaping mental health outcomes during the recovery period. All regressions include the standard set of controls.

Table 5: Recovery Period Academic Outcomes

	Fall Math		**Fall Turkish**	
	(1)	(2)	(1)	(2)
Mid School GPA	0.16***	0.16***	0.16***	0.17***
	(0.03)	(0.03)	(0.03)	(0.04)
Study Time (Pre)	0.03**	0.04**	0.03**	0.04**
	(0.01)	(0.01)	(0.01)	(0.02)
ΔST	0.17***	0.15***	0.14***	0.09*
	(0.03)	(0.05)	(0.03)	(0.04)
Home Damage	-0.08***	-0.05*	-0.07**	-0.05
	(0.03)	(0.03)	(0.03)	(0.03)
HIA	-0.12*	-0.09	-0.15**	-0.16**
	(0.07)	(0.08)	(0.07)	(0.08)
MIA	-0.30***	-0.26***	-0.31***	-0.25***
	(0.07)	(0.08)	(0.06)	(0.07)
HIA $\times$ ΔST	0.14**	0.15**	0.15**	0.23***
	(0.05)	(0.06)	(0.06)	(0.07)
MIA $\times$ ΔST	0.11**	0.11*	0.07	0.10
	(0.05)	(0.06)	(0.05)	(0.06)
Constant	-0.01	-0.51***	-0.11	-0.54***
	(0.07)	(0.10)	(0.07)	(0.10)
Student Controls	Yes	Yes	Yes	Yes
Parental Controls	No	Yes	No	Yes
Observations	1430	1030	1430	1030
R^2	0.134	0.193	0.131	0.180

Notes: This table shows regressions of fall-term exam scores on non-cognitive skills and controls. Robust standard errors in parentheses.

Table 6: Regression of Fall Mental Health Outcomes

	GAD (Reversed)		**CAPSAW (current)**	
	(1)	(2)	(1)	(2)
Mid School GPA	0.01	-0.01	-0.00	-0.00
	(0.03)	(0.03)	(0.02)	(0.03)
Study Time (Pre)	-0.01	-0.01	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)
ΔST	0.10***	0.12**	0.09**	0.10**
	(0.04)	(0.06)	(0.03)	(0.05)
Home Damage	-0.03	-0.04	-0.03	-0.03
	(0.02)	(0.03)	(0.02)	(0.03)
HIA	-0.21***	-0.13*	-0.26***	-0.29***
	(0.06)	(0.07)	(0.06)	(0.07)
MIA	0.01	0.04	0.00	-0.04
	(0.05)	(0.07)	(0.05)	(0.06)
HIA $\times \Delta ST$	0.04	0.02	-0.02	-0.02
	(0.06)	(0.08)	(0.06)	(0.07)
MIA $\times \Delta ST$	0.04	0.02	0.03	0.01
	(0.05)	(0.07)	(0.05)	(0.06)
Constant	0.37***	0.18*	0.18***	0.05
	(0.06)	(0.09)	(0.06)	(0.09)
Student Controls	Yes	Yes	Yes	Yes
Parental Controls	No	Yes	No	Yes
Observations	1430	1030	1430	1030
R^2	0.351	0.358	0.405	0.398

Notes: This table reports regressions of post-quake mental health outcomes on residualized study time change (ΔST), home damage, and controls. GAD is reversed for interpretation purposes.

4.6 Robustness Checks

4.6.1 Migration Status. One potential threat to identification is the non-random selection of students in the treatment and control groups due to the high post-quake migration rates in the treatment provinces. Just to document clearly, there are three groups of populations, those always stay, those who temporarily relocated in the aftermath of the quake but came back home, and those who permanently left the quake region. First let's document the migration patterns. Tables A11 and A12 present the migration patterns for the two impacted provinces, HIA and MIA. The first table reports the share of students who moved, the average number of moves among movers, and the share of those who relocated to a big city. The second table provides additional detail on migration dynamics across different move episodes.[40] This includes the average number of people they stayed with, the type of housing, and the stated reason for each move. Temporary relocation rates are notably high: 66.7% of students in HIA and 72.5% in MIA moved at least once. On average, movers reported 2.23 and 2.14 moves in HIA and MIA, respectively. The share of moves involving relocation to a big city was 14.6% in HIA and 21.5% in MIA. This difference may reflect variation in access to resources or urban networks across provinces. Move duration also varies across moves. The average duration of the first move was 1.23 months in HIA and 1.34 months in MIA, while the second move was longer, averaging 2.63 months and 3.03 months, respectively. First moves generally involved larger household sizes, suggesting that students initially relocated with extended family. In terms of housing types, the most common type during the first move was staying with relatives, reported by 39.1% of students in HIA and 40.5% in MIA. The second most common was renting a house, followed by staying in a village house.[41] The common reason for migrating after the earthquake was to escape unsafe living conditions, cited by 65.4% of movers in HIA and 58.3% in MIA. Educational motives were also salient, accounting for approximately 18 of first moves in both provinces. Psychological reasons were more frequently cited in MIA, suggesting potential differences in resilience or stress responses across the two regions. In summary, migration patterns are dynamic, with substantial variation in both initial and subsequent responses to the earthquake. The differences between provinces reveal potential heterogeneity in coping strategies and post-quake living patterns.

After noting down the migration patterns for the group of students who temporarily relocated, we now turn to understand the always stayers' reason for staying. Figure A14 presents the reasons provided by students who stayed in their hometown and never relocated. The most commonly reported reason is parent's employment, followed by a sense of belonging to the hometown. Additional reasons include ownership of property or agricultural land, lack of a social network outside. These results suggest that relocation decision is also influenced by both inside and outside factors.

Tables A13 and A14 complement our analysis in that we ask students and parents

[40] In the survey, students were asked to report their migration history, with up to five moves recorded.
[41] A village house refers to a residence in a rural area, typically owned by the student's family.

about their perceptions of people who left the region permanently. We asked respondents whether they think the permanent leavers are higher income, younger, more educated, and less resilient. Across all these measures in both provinces, the mean responses are around 3 on a 5-point Likert scale, corresponding to neutral midpoint. However, when we break down these perception responses across those who are always stayers and those who temporarily relocated, movers tend to have more neutral responses, while non-movers specifically in HIA tend to agree with the statements. For MIA, the differences across the two groups are minimal, with both groups having similar, neutral perceptions about the leavers. These analyses suggest potential selection into migration, specifically in HIA, where the earthquake's impact was most severe.

We address the selection as follows. First, we compute lower and upper bounds for the main interaction terms using Lee (2009) bounds, constructed the admin-reported attrition rates in the two provinces.[42] The results together with the baseline estimates are provided in Table A15. The interaction terms stay positive and the directions across the two provinces are consistent with the baseline estimates.

To understand, how the covariates explain selection into migration we estimate a logit model where the outcome variable is the relocation status. Results are presented in Table A16. The results align with people's perceptions of those who left the region. Specifically, we find that income is positively associated with the probability of leaving, while resilience is negatively associated. We form inverse probability weights (IPW) using the predicted probabilities from the logit model.[43] Table A17 presents the results of the main regression using IPW weights. The main interest variables, which are the interaction terms, remain positive and statistically significant. The province specific terms are now estimated to be closer to zero, which is expected since we are now controlling for the selection.

4.6.2 Potential Recall Bias. Our input and outcome measures are self-reported and come from both pre- and post-earthquake periods. This raises the concern of potential recall bias, which could take the form of non-classical measurement error if it varies systematically by treatment status, potentially biasing our estimates. We address this concern as follows. First, we document differences in recall confidence across provinces. Table A18 reports summary statistics for recall confidence from three perspectives: students' confidence in their own recall, parents' confidence in their own recall, and parents' confidence about their child's recall. On average, student confidence is 3.6 on a 5-point scale, with no notable differences across provinces. Parents report higher confidence in their own recall, around 4.2 on average, but lower confidence about their child's recall, around 3.0. To address potential bias from differential recall, we re-estimate our main regressions including a control for recall confidence. Results are presented in Table A19. The interaction terms remain positive and statistically significant. Moreover, recall confi-

[42] We trim the upper and lower tails of the outcome distribution in Control province until the retained share matches the post-quake presence rate in the two provinces.

[43] We calibrate province specific intercept shifts to match the attrition rates in the two provinces.

dence itself is positively associated with outcomes, suggesting recall quality matters, but in a way that does not undermine the main results.[44] Third, we re-run our main regressions using the parent-reported study time data to construct the residualized behavioral NC measure. These results are reported in Table A20. The interaction terms remain positive and of comparable magnitude, supporting the robustness of our findings.

4.6.3 Data Collection Robustness. We specifically check the robustness of our results to the data quality across three dimensions: (i) the survey was conducted in proctored classroom settings, (ii) the survey which was filled only by attentive respondents, and (iii) the implementation induced variation like the classroom environment. Details about each of these robustness checks are provided in Appendix G.3 and results are presented in Table A21. Accordingly, under (i) and (ii), we find that the results are robust both in magnitude and significance. Under (iii), while the coefficients lose significance, the directions of the coefficients remain the same.

5 Mechanisms

`in progress...`

6 Conclusion

This study adds to the literature on educational disruptions and human capital by using a large-scale field survey to produce several important lessons for resilience in education. First, we document the wide-ranging disruptions that students face after a major natural disaster, including not only physical disruptions but also mental health challenges. We provide novel evidence on how non-cognitive skills, particularly resilience and grit, help students sustain academic performance in the face of such shocks. While we observe impacts on academic outcomes, mental health outcomes appear less responsive to these traits. This highlights the need for a broader set of interventions to support psychological recovery, beyond individual behavioral traits. Closing student gaps after major shocks will likely require a combination of academic support, mental health services, and community rebuilding efforts. While there have been attempts to integrate psychosocial interventions post-quakes in our context, our results suggest that these government-led efforts do not effectively reach those in need.

Overall, our findings underline the importance of considering both cognitive and non-cognitive dimensions in education policy, especially in contexts prone to disruptions. Future research could explore additional mechanisms that support recovery and resilience,

[44] In an alternative specification, we also interact recall confidence with the treatment terms. These interactions are not statistically significant, suggesting that recall confidence does not meaningfully alter the treatment effect. Our identifying assumption here is that students truthfully report their confidence, in other words, they are aware of what they do and do not recall.

and identify policies that can effectively prevent gaps between affected and non-affected students from emerging.

References

2023 Kahramanmaraş and Hatay Earthquakes Report. (n.d.).

Alan, S., Boneva, T., & Ertac, S. (2019). Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit*. *The Quarterly Journal of Economics*, *134*(3), 1121–1162.

Alan, S., & Ertac, S. (2018). Fostering Patience in the Classroom: Results from Randomized Educational Intervention. *Journal of Political Economy*, *126*(5), 1865–1911.

Alan, S., & Kubilay, E. (2023). Impersonal trust in a Just and Unjust world: Evidence from an educational intervention. *Journal of Development Economics*, *162*, 103044.

Alan, S., & Turkum, B. (2024). Abstract reasoning, theory of mind and character development in the school. *Journal of Economic Behavior & Organization*, *221*, 307–326.

Alasino, E., Ramírez, M. J., Romero, M., Schady, N., & Uribe, D. (2024). Learning losses during the COVID-19 pandemic: Evidence from Mexico. *Economics of Education Review*, *98*, 102492.

Andrabi, T., Daniels, B., & Das, J. (2021). Human Capital Accumulation and Disasters: Evidence from the Pakistan Earthquake of 2005. *Journal of Human Resources*.

Angrist, N., Ainomugisha, M., Bathena, S. P., Bergman, P., Crossley, C., Cullen, C., Letsomo, T., Matsheng, M., Marlon Panti, R., Sabarwal, S., & Sullivan, T. (2023, May). *Building Resilient Education Systems: Evidence from Large-Scale Randomized Trials in Five Countries*. 31208.

Aucejo, E. M., French, J., Ugalde Araya, M. P., & Zafar, B. (2020). The impact of COVID-19 on student experiences and expectations: Evidence from a survey. *Journal of Public Economics*, *191*, 104271.

Bacher-Hicks, A., Goodman, J., & Mulhern, C. (2021). Inequality in household adaptation to schooling shocks: Covid-induced online learning engagement in real time. *Journal of Public Economics*, *193*, 104345.

Belmonte, A., Bove, V., D'Inverno, G., & Modica, M. (2020). School infrastructure spending and educational outcomes: Evidence from the 2012 earthquake in Northern Italy. *Economics of Education Review*, *75*, 101951.

Betthäuser, B. A., Bach-Mortensen, A. M., & Engzell, P. (2023). A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic. *Nature Human Behaviour*, *7*(3), 375–385.

Biener, C., & Landmann, A. (2023). Recovery mode: Non-cognitive skills after the storm. *World Development*, *164*, 106152.

Boruchowicz, C., Parker, S. W., & Robbins, L. (2022). Time Use of Youth during a Pandemic: Evidence from Mexico. *World Development*, *149*, 105687.

Brown, C., Kaur, S., Kingdon, G., & Schofield, H. (2025). Cognitive Endurance as Human Capital*. *The Quarterly Journal of Economics*, *140*(2), 943–1002.

Bulaon, J. P., & Shoji, M. (2022). Disaster Exposure in Childhood and Adult Noncognitive Skill: Evidence from the Philippines. *MPRA Paper*, (112913). Retrieved April 9, 2025, from https://ideas.repec.org//p/pra/mprapa/112913.html

Caldara, D., & Iacoviello, M. (2022). Measuring Geopolitical Risk. *American Economic Review*, *112*(4), 1194–1225.

Cavallo, E., Powell, A., & Becerra, O. (2010). Estimating the Direct Economic Damages of the Earthquake in Haiti. *The Economic Journal*, *120*(546), F298–F312.

Cavallo, E. A., & Noy, I. (2009). The Economics of Natural Disasters: A Survey. Retrieved November 5, 2023, from https://publications.iadb.org/en/publication/economics-natural-disasters-survey

Cerqua, A., & Di Pietro, G. (2017). Natural disasters and university enrolment: Evidence from L'Aquila earthquake. *Applied Economics*, *49*(14), 1440–1457.

Chen, D. L., Ertac, S., Evgeniou, T., Miao, X., Nadaf, A., & Yilmaz, E. (2024). Grit and academic resilience during the COVID-19 pandemic. *npj Science of Learning*, *9*(1), 1–10.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*, *104*(9), 2593–2632.

Chinoy, S., Nunn, N., Sequeira, S., & Stantcheva, S. (2023, September). *Zero-sum Thinking and the Roots of US Political Differences*. 31688.

Cunha, F., Heckman, J. J., & Schennach, S. M. (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, *78*(3), 883–931.

Deming, D. J. (2017). The Growing Importance of Social Skills in the Labor Market*. *The Quarterly Journal of Economics*, *132*(4), 1593–1640.

Dercon, S., & Porter, C. (2014). Live Aid Revisited: Long-Term Impacts of the 1984 Ethiopian Famine on Children. *Journal of the European Economic Association*, *12*(4), 927–948.

The economic impacts of learning losses. (2020, September 9). OECD. Retrieved April 9, 2025, from https://www.oecd.org/en/publications/the-economic-impacts-of-learning-losses_21908d74-en.html

Ersoy, F. (2020). The effects of the great recession on college majors. *Economics of Education Review*, *77*(100).

Gallagher, E. A., Billings, S. B., & Ricketts, L. R. (2023). Human Capital Investment after the Storm. *The Review of Financial Studies*, *36*(7), 2651–2684.

Goldhaber, D., Kane, T. J., McEachin, A., Morton, E., Patterson, T., & Staiger, D. O. (2022, May). *The Consequences of Remote and Hybrid Instruction During the Pandemic*. 30010.

Guariso, A., & Björkman Nyqvist, M. (2023). The impact of the COVID-19 pandemic on children's learning and wellbeing: Evidence from India. *Journal of Development Economics*, *164*, 103133.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, *24*(3), 411–482.

Imberman, S. A., Kugler, A. D., & Sacerdote, B. I. (2012). Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees. *American Economic Review*, *102*(5), 2048–2082.

Inui, T., & Okudaira, H. (2022). Parental Investment after Adverse Event: Evidence from the Great East Japan Earthquake. *Discussion papers*, (22049). Retrieved April 9, 2025, from https://ideas.repec.org//p/eti/dpaper/22049.html

Jack, R., Halloran, C., Okun, J., & Oster, E. (2023). Pandemic Schooling Mode and Student Test Scores: Evidence from US School Districts. *American Economic Review: Insights*, *5*(2), 173–190.

Jacob, B., & Rothstein, J. (2016). The Measurement of Student Ability in Modern Assessment Systems. *Journal of Economic Perspectives*, *30*(3), 85–108.

Kreft, C., Huber, R., Wuepper, D., & Finger, R. (2021). The role of non-cognitive skills in farmers' adoption of climate change mitigation measures. *Ecological Economics*, *189*, 107169.

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020). Projecting the Potential Impact of COVID-19 School Closures on Academic Achievement. *Educational Researcher*, *49*(8), 549–565.

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies*, *76*(3), 1071–1102.

Lichand, G., Doria, C. A., Leal-Neto, O., & Fernandes, J. P. C. (2022). The impacts of remote learning in secondary education during the pandemic in Brazil. *Nature Human Behaviour*, *6*(8), 1079–1086.

Lytle, A., & Shin, J. E. L. (2023). Resilience and grit predict fewer academic and career concerns among first-year undergraduate students during COVID-19. *Social Psychology of Education*, *26*(1), 227–240.

Morrill, M., & Westall, J. (2023). Heterogeneity in the educational impacts of natural disasters: Evidence from Hurricane Florence. *Economics of Education Review*, *94*, 102373.

Ozer, E., Krishna, K., & Akyol, P. (2024). Do Time-Constraints Matter? How, Why, and For Whom? Retrieved April 9, 2025, from https://esmaozer.github.io/research/time-constraints/time-constraints.pdf

Parolin, Z., & Lee, E. K. (2021a). Large socio-economic, geographic and demographic disparities exist in exposure to school closures. *Nature Human Behaviour*, *5*(4), 522–528.

Parolin, Z., & Lee, E. K. (2021b). Large socio-economic, geographic and demographic disparities exist in exposure to school closures. *Nature Human Behaviour*, *5*(4), 522–528.

Paudel, J., & Ryu, H. (2018). Natural disasters and human capital: The case of Nepal's earthquake. *World Development*, *111*, 1–12.

Sacerdote, B. (2012). When the Saints Go Marching Out: Long-Term Outcomes for Student Evacuees from Hurricanes Katrina and Rita. *American Economic Journal: Applied Economics*, *4*(1), 109–135.

Shidiqi, K.-A., Di Paolo, A., & Choi, Á. (2023). Earthquake exposure and schooling: Impacts and mechanisms. *Economics of Education Review*, *94*, 102397.

Siddique, A. (2024). Behavioral consequences of religious schooling. *Journal of Development Economics*, *167*, 103237.

Tincani, M. M. (2018). Heterogeneous Peer Effects in the Classroom.

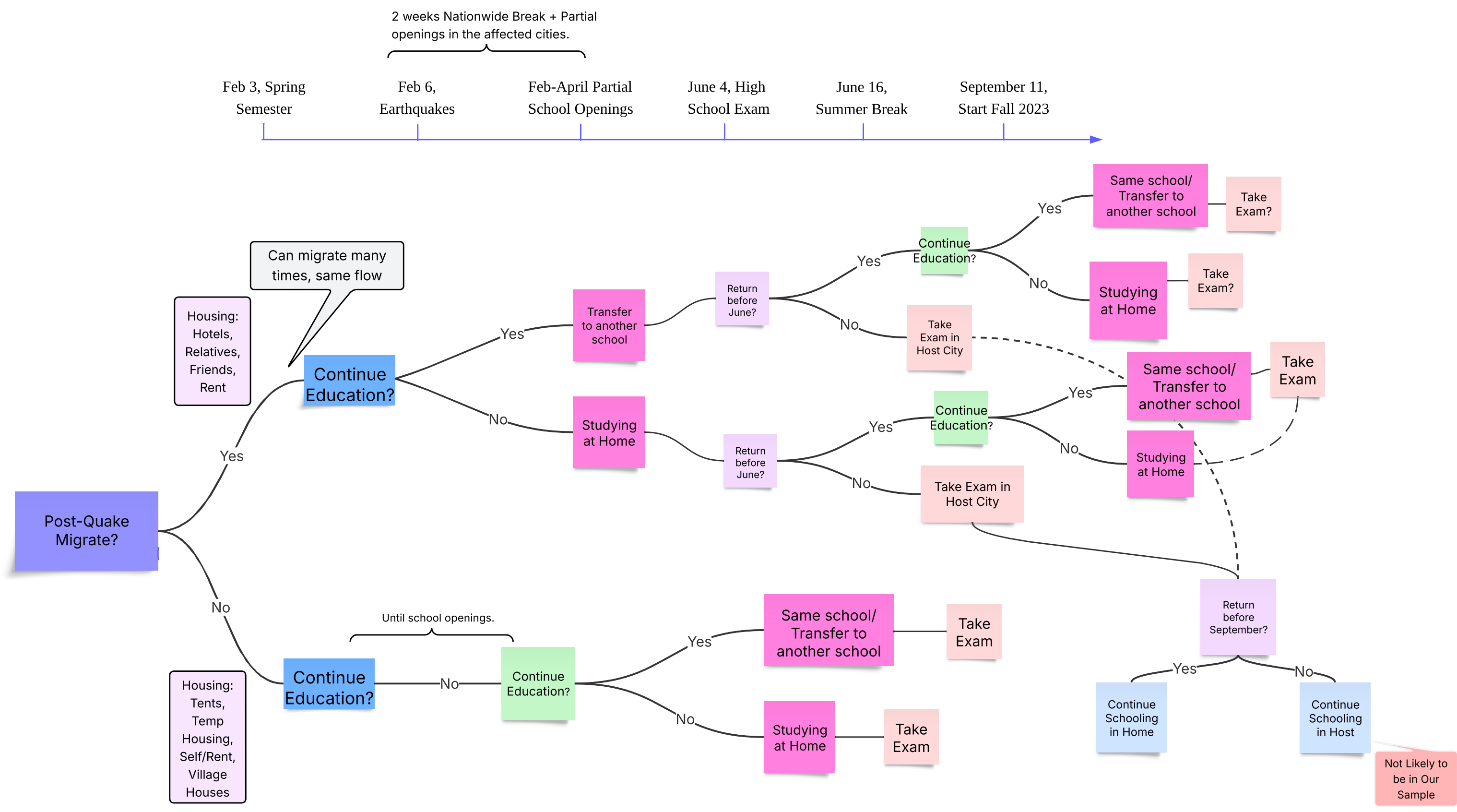
Williams, B. A., Jones, C. H., Welch, V., & True, J. M. (2023). Outlook of pandemic preparedness in a post-COVID-19 world. *npj Vaccines*, *8*(1), 178.

Appendices

A Background

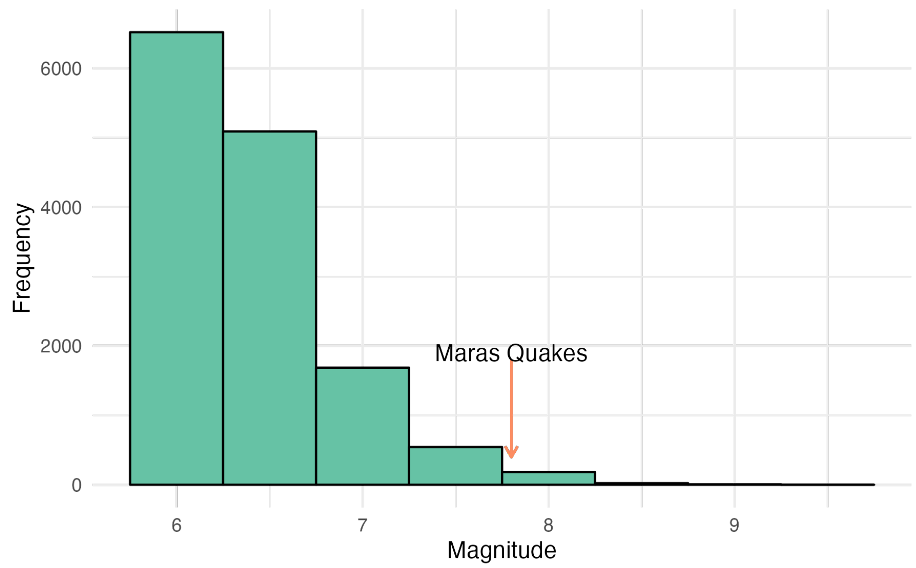
Figures A2 and A3 show the earthquake frequency over time and the intensity map of the February 6, 2023 earthquake.

Figure A1: Student Education Paths After the Earthquake



Notes: This figure shows the different paths students might have taken after the earthquake. The timeline on top marks the key disruptions: the February 6 earthquakes, partial school openings in spring, the June high school exam, and the September return for the new school year. The diagram summarizes possible paths; actual experiences varied across students, and some may not fit neatly into one branch

Figure A2: Earthquakes with *mag*> 6.0, 1900-2024



Source: USGS Earthquake Catalog.

Figure A3: February 6, 2023 Earthquake Intensity

Source: New York Times.

B Survey

The survey was developed, in house, in collaboration with several parties including: economists, psychologists, educators, Department of Education professionals, and those counseling professionals who voluntarily worked in temporary housing psychosocial support centers built after the earthquake. The role of people in the field specifically helped in designing trauma-related questions since they had first-hand experience with the affected population. The input of psychologists was also essential to design questions with ethical considerations in mind. We incorporated the multi-disciplinary feedback to the revisions after the pilot testing with a small group of students (around 30 students) and their parents.

B.1 Sample Recruitment

Table A1: Earthquake Impact by Province

Province	Population	Housing Units	Destruction %	Death %	Injury %	Composite Index
Adana	2,301,089	972,561	0.303	0.0197	0.324	0.039
Adıyaman	647,658	216,744	25.9	1.29	2.70	0.955
Diyarbakır	1,776,852	563,295	1.53	0.0233	0.0508	0.021
Elazığ	603,789	292,406	3.48	0.0008	0.0628	0.043
Gaziantep	2,143,632	893,558	3.26	0.182	0.622	0.151
Hatay	1,621,757	847,380	25.4	1.49	1.90	0.893
Kahramanmaraş	1,157,976	481,362	20.6	1.09	0.798	0.602
Kilis	147,879	74,976	3.35	0.0500	0.510	0.108
Malatya	812,395	345,536	20.7	0.172	0.794	0.397
Osmaniye	555,112	243,436	6.61	0.179	0.401	0.166
Şanlıurfa	2,125,133	718,063	0.859	0.0160	0.420	0.057

Notes: The destruction rate is defined as the number of buildings that were destroyed or damaged divided by the total number of buildings in the province. The death and injury rates are defined as the number of fatalities or injuries divided by the total population. The composite index is the unweighted average of the min-max normalized destruction, death, and injury rates, scaled to the [0,1] range. Statistics obtained from "2023 Kahramanmaraş and Hatay Earthquakes Report" (n.d.).

B.1.1 Detailed School Selection To ensure comparability and feasibility, we restrict the sample to schools located in the central districts of each province. There are three reasons for this choice. First, schools in city centers are significantly easier to access within the limited fieldwork window, whereas reaching rural schools would require substantially more time and resources. Second, enrollment is generally higher in urban areas, ensuring a more robust sampling frame.[45] Third, students in rural areas are more likely to participate in family-based agricultural and livestock work, making them potentially systematically different in terms of time use and school engagement.

Within each province, we select schools using stratified sampling based on school performance. For exam schools, stratification is based on entrance exam cutoff scores from the 2023 high school placement system. These schools do not recruit by neighborhood, so student characteristics can vary widely. In the central districts, cutoff scores range from 388 to 480 in MIA (8 schools), 407 to 477 in Control province (7 schools), and 422 to 463

[45] Rural to urban enrollment ratio is approximately 0.15 (KODA, 2016).

in HIA (3 schools). We define three strata: High-Performance Schools (HPS) with cutoffs above 450, Medium (MPS) with 425–450, and Low (LPS) with cutoffs below 425. While the range for HPS is wide, the number of such schools per province is limited, generally one to three. One exam school in HIA opted out participation.

For non-exam schools, stratification is based on middle school GPA cutoffs, typically ranging from 50 to 100. We define HPS as schools with GPA cutoffs above 85, MPS as 70–85, and LPS as below 70 (capped at 55 to facilitate school participation). This classification ensures at least three eligible schools per province, with the exception of HIA, where only two schools met the criteria in the LPS tier. In cases where originally selected schools declined participation, substitutions were made with comparable schools based on cutoff scores and in consultation with the local education authorities. While the final list of schools is not identical to the initial sampling frame, replacement schools were matched closely in terms of performance and characteristics, preserving the integrity of the stratified design. In total, we survey 18 high schools across the three provinces: 5 in HIA, 7 in MIA, and 6 in Control province. The slightly higher number of schools in MIA reflects both its larger school population and logistical flexibility during fieldwork, which allowed for the inclusion of an additional non-exam school within the Medium sampling strata, which has the highest number of schools.

B.2 Implementation Details

B.2.1 Planning and Timeline Our implementation took a careful planning given the short time window and provinces being apart from each other. Our implementers were based in MIA province. We started implementation in MIA province and completed survey data collection here in the first to a large extent. Second, in the second week, schools in HIA and MIA provinces (which are around 1,5 hours of driving each from MIA) were visited. We approach the school administration either in person or via phone and both tell about our research, implementation procedure and how they could help us both for the classroom implementation and contacting to parents. This second step generally occured 1-3 days before the actual school visit. Then, as a third step, implementers visit schools for in person implementation for which the details are listed in the next section. After this step, the parental link distribution comes and this occurs either the same day evening or the next day during the day. This depends on the school administration's communication tool with the parents. If they communicate over a text message system used in the school, then they distribute the link the next day. If they have a WhatsApp group with parents, then the link is distributed the same day evening. Then, we send reminders to those respondents' with incomplete survey or students whose parents didn't fill in the survey. After we make sure reminders sent three times, then we created a list of students with parental filling and without it. Then, the payments are distributed to students in closed envelopes one week after this step.

B.2.2 Students' Survey We describe a usual procedure for classroom and online surveys. Note that there are two types of implementation type: both (i) the speech and implementation done in the classroom by our team, (ii) the speech done by our team in the classrooms but the survey conducted online. We, below, explain in classroom implementation separately.

i. **In-Classroom Implementation.** Once the implementers arrive in the school, they meet with the school manager and the counseling teachers. The counseling teachers, who have a broader view of their school, teachers, and the students let in advance us know about any considerations we should be aware such as students with special needs.[46] Then the two implementers and the counseling teacher visited the 9th-grade classrooms and gave a brief speech and answers clarification questions, as well as took note of those students without phones, battery, or Internet so they can be addressed either in the classroom or later if the student is willing to fill out the survey later. In this first classroom visit, the Wi-Fi information is provided and phones/chargers given for anyone in need.[47] Luckily, in most schools there were not any problem with Wi-Fi given the classrooms were in the same corridor and not apart. However, when classrooms were apart, the Wi-Fi hotspot was located in the corridor where the majority of the classrooms are and for the classroom without Wi-Fi access, an additional mobile hotspot was provided by the school administration devices. If Wi-Fi still has problems due to other unseen technical reasons, then implementers with personal Internet access shared their mobile hotspot with students.

   Once this step is done, then the next classroom hour, the survey link is distributed through the school administration to the students via the common WhatsApp groups. Luckily, each classroom has their WhatsApp announcement group where school admin communicates them.

   Before the link distribution, each classroom has their teacher or proctor. Teachers also were informed about the study by the school admin and they were told to provide an exam-like conditions to prevent student communication or noise.[48] Most of the time, each classroom had one proctor. During the implementation, the implementers and counseling teachers were active and mobile to answer any clarification questions without too much interfering with the details though.

   Students who are done with the survey leave their phone in the boxes, and then either stay silently or leave if it is break time. When it is break time and there are still students who didn't complete the survey, the implementers and counseling teachers replaced the existing proctor teacher. If there are still delays and students who

[46] While not common, there were students with visual impairments or language barriers. Whenever this is the case, an implementer helped these students' completing the survey without any problem.

[47] Time to time, the number of extra devices were less than the number of students who do not bring their phone to school. For these students, either a second-round was there for them to complete the survey with extra devices or they could be take-home if they are not willing to do it in classroom.

[48] In few instances, the teachers assigned for that classroom hour were absent and not present in the school. When this happened, one of the implementers proctored.

continue their survey (could be due to several technical reasons such as slow Internet), then the teacher assigned for the next classroom hour continued proctoring. The smoothness of these transtions were maximized through communication with teachers in advance. Implementers leave the school after making sure no one with incomplete survey left.

The speech text (translated to English) is given by:

Hello everyone. [Introduce implementer briefly.] We're here to conduct a survey about your experiences related to the earthquake. Your answers will help us better understand how students have been affected. You'll receive a cash reward for completing the survey.

In the next class hour, you'll complete the survey on your phone. It should take around 25–30 minutes. During the survey, please remain seated and silent. No talking or communication with others is allowed. Your teacher will supervise the session. Do not open any other apps—turn off notifications and mute your phone now. [Pause 1 minute.]

Does anyone not have a phone or enough battery? [Implementer takes note of who needs devices or will complete the survey at home.] We have a few backup phones—if you'd prefer to complete the survey later at home instead, please let us know.

A few quick but important rules: this survey uses smart design to detect response quality. If your answers are rushed or random, you may not receive your reward. If you're caught using other apps, your teacher will give you one warning. After that, you'll be disqualified from completing the survey.

We take this process seriously. We want to ensure both quality and fairness. Later today, your parents will also receive a survey link. If both you and your parent complete the survey properly, you'll receive 200TL in the coming days. We prefer your mother to fill it out, but your father or another parent can also do so. Please do not discuss the survey with them while they complete it.

Your answers—and your parents'—are completely confidential. We do NOT share your responses with teachers or parents, and vice versa. This data is only used for academic purposes.

The survey is straightforward. You should be able to complete the questions on your own without technical help. In addition, as laid out before, it is a voluntary survey, and you can choose not to participate. If you do not want to take the survey, please let the implementer know now. If you have any questions or concerns, feel free to ask the implementer.

Now, please check your internet connection and open the link shared in your class chat group. If needed, you can use the Wi-Fi we provide. You can also find the link written on the board.

[Wait 2–3 minutes, confirm students are ready.]

Thank you for your attention—please begin the survey now.

ii. **Online Implementation.** For a total of three schools in HIA and Control province, the survey was administered online. A modified version of the speech above was done by the research team in the first school visit. Students received the survey link through their school's WhatsApp group and were asked to complete it between 7–9 p.m. that evening. To assist with questions or technical issues, we provided two callback numbers and ensured availability during the survey window.

B.2.3 Parents' Survey All parents' survey was conducted online. The same day or the following day of when students took the survey, the parental link was distributed by the school administration. In the case that parent does not have a smart phone, we told students that their parents can use the student's phone but should let the research team know. The description sent to parents is included in the following text:

Dear Parents, as you know, with the permission of the Ministry of Education, a research team has launched a survey to assess the educational situation following the earthquake. We kindly ask you (preferably the student's mother, if possible) to complete the survey using the link provided, within two days Please read each question carefully and answer based on your own personal views.

Table A2: Summary of Implementation Variables

Variable	Mean	SD
School Surveyed Online	0.220	0.420
Classroom Size	31.720	3.220
Implementer Recall Confidence	1.350	0.720
Survey During Mid-Courses	0.740	0.440
Teacher Change	0.630	0.480
Break Noise	0.400	0.530
Student Noise	0.610	0.670
Teacher Helpfulness	0.820	0.380
Managerial Cooperation	1.550	0.590
Wi-Fi Working	0.870	0.330
Post-Submission Leaving Classroom	0.440	0.500
No Charger Demand	0.170	0.380

It is very important that you do not communicate with your child while completing the survey. As a token of appreciation for your and your child's valid participation, your child will receive a 200 Turkish Lira allowance. Your participation is valuable and essential to the success of this research. Thank you in advance for your time and cooperation. If you have any questions, feel free to contact our research coordinator at 5xxxxxxxxx. Survey Link: [link]

B.2.4 Quality Check

Table A3 summarizes the attention check variables across provinces.

Table A3: Data Quality and Attention Checks

Variable	All	HIA	MIA	Control
Progress (%)	96.13	94.32	96.9	96.6
	(0.42)	(0.96)	(0.57)	(0.73)
Duration (min)	45.31	37.83	41.28	34.7
	(0.72)	(0.92)	(0.81)	(0.79)
Total Attention Score	1.83	1.81	1.8	1.9
	(0.01)	(0.02)	(0.02)	(0.01)
Recall Confidence	3.65	3.65	3.7	3.59
	(0.02)	(0.04)	(0.04)	(0.05)
Effort Level	4.09	4.04	4.11	4.11
	(0.02)	(0.04)	(0.03)	(0.03)

Notes: This table shows the attention check variables across provinces.

B.3 Survey Flow

The survey flow is given in Figure A4. In terms of the flow, we prioritized the order of exposure, mental health, and noncognitive skills over educational outcomes and time use since the former potentially requires more care to recall and is more subjective while a test outcome or GPA is more objective and can be verified by the admin data if shared.

Each block informs the respondent about the block content and in detail gives instruction on which period to recall. We have different types of questions across the survey including multiple choice, Likert-scale, slider, and open-ended questions. We included open-ended questions so that the participants could freely express themselves on what they experienced post-quake which are not captured by our questions. We used conditional logic when applicable. For example, movement questions not displayed for respondents reporting zero movements.

B.4 Survey Completion Time Distribution

Tables A5 and A6 show the survey completion time distribution for students and parents, respectively.

B.5 Implementation Figures

Figure A7: In-classroom Implementation

B.6 Survey Attrition

The survey collection aimed to get responses from both students and their parents. There are some students whose parents did not fill out the survey and there are some parents whose children did not fill out the survey. I dropped the latter since the main focus group is students and whose parents fill out the survey. Around 28% of students' parents did not fill out the survey, meaning that for 1098 students we have responses from both students and parents.

In addition, among those who started the survey, 6.44% of students and 8.94% of parents did not complete the survey. The predictors of these within survey attrition are given in Table. Among those who quit the survey, the most common survey location is the earthquake exposure block, similarly for parents that the earthquake exposure block and the first block asking about education and occupation are the common places where they quit the survey. Only focusing on the first set of questions as regressors, Table A5 shows that for students, being in the Medium and Low performance schools increases the probability of survey completion, while for parents, there is no significant predictor of survey completion.

Table A4: Predictors of Parental Participation

Variable	**Coefficient**	**Std. Error**
Intercept	0.19	(0.15)
Province: Control	-0.22***	(0.03)
Province: MIA	-0.01	(0.03)
Staying With Mom	0.00	(0.05)
Staying With Dad	-0.01	(0.03)
Male	0.01	(0.02)
Stud. Completed Survey	0.48***	(0.13)
Academic Ability	0.00	(0.00)
Admin. Cooperation	0.04	(0.03)
N	1371	
R^2	0.06	

Notes: This table reports OLS estimates where the dependent variable is if student's parent filled out the survey. Robust standard errors are reported in parentheses. Statistical significance is indicated by stars: $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

C LLM Analysis

To optimize the classification prompt, we used a few-shot learning approach, using LLMs to classify chat messages into predefined binary categories. We implemented an ensemble method using Llama 3.1, Qwen, and Gemma.[49] Each model independently labeled the dataset based on the optimized prompt. Final labels were assigned through a majority vote, improving robustness and reducing both model-specific and researcher-specific

[49] The implementation leverages Together AI's serverless API for inference. The models used are Gemma-2 Instruct 27B, Meta Llama 3.1 70B Instruct Turbo, and Qwen 2.5 Coder 32B Instruct.

Table A5: Linear Probability Models Predicting Survey Completion

Variable	Coef.	(SE)
Panel A: Student Completion		
Intercept	0.91***	(0.02)
Male	-0.00	(0.01)
Province: Control	0.04**	(0.02)
Province: MIA	0.04**	(0.02)
Panel B: Parent Completion		
Intercept	0.88***	(0.03)
Mom	-0.04	(0.02)
Other Relative	-0.12	(0.11)
Province: Control	0.04	(0.03)
Province: MIA	0.03	(0.02)

Notes: This table reports OLS estimates where the dependent variable is if surveyor completed the survey. Statistical significance is indicated by stars: $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

biases. The training is based on a small set of labeled data (100 responses, 40 training, 60 validation) to ensure the model's ability to generalize across different contexts. The accuracy rate on the validation set is 0.95.[50]

C.1 LLM Classification Prompt

The prompt used for the classification task for all models is as follows:

[50] The training is done using Qwen 2.5 72B model.

You are a highly accurate classification model. Your task is to classify Turkish-language survey responses written by students after a major earthquake. These open-ended responses describe how students were affected. Please label each response using binary indicators (1 if present, 0 if not). Each category can independently be marked as 1 or 0. If the response clearly refers to a category, assign 1; otherwise, assign 0. If the text is unclear or missing or consists only of \6666", \7777", \8888", or \9999", skip the example and do not return any output. Categories: • Psychological (1/0): The response reflects emotional, mental distress, including fear, trauma, sadness, anxiety. Do NOT assign this category if the emotional content is only about others (e.g., friends or family being sad or scared), unless the respondent also expresses their own psychological state. Do NOT assign 1 for: Positive psychological growth (e.g., \psikolojik gelişim"), Academic/motivational loss not tied to distress, Explicit denials of psychological impact. – SocialDisruption (1/0): The message reflects disruption in the respondent's own social connections, such as Separation from friends or family, Social isolation or detachment, Reduced interaction with peers or the community, Feeling alone, withdrawn, or disconnected. Do NOT assign SocialDisruption = 1 for General emotional mentions of others (e.g., \arkadaşım etkilendi", \annem üzüldü"), Descriptions of others' psychological states without indicating social disconnection. – Academic (1/0): The message indicates academic or performance-related effects, such as exam stress, lower motivation to study, difficulty concentrating, or changes in school engagement. Do NOT assign if School is only mentioned as a place (e.g., \okuldan uzak kaldım"), or if there is no reference to learning, exams, studying, or cognitive effort. – Physical/Material (1/0): The response discusses material or physical consequences, such as housing loss, financial difficulties, physical harm, or deaths in the family. – MinImpact (1/0): The message states that the earthquake had no effect or very minimal effect on the respondent's life or feelings. Assign MinImpact = 1 for vague or neutral responses such as \normal etkiledi", \çok fark etmedi", or \orta derecede etkiledi" only if there is no clear domain-specific impact. If any domain is explicitly mentioned, assign that instead. – PositiveGrowth (1/0): The response shows resilience or positive transformation, including signs of personal growth, maturity, or finding new perspective after the disaster. Do NOT assign 1 for Simple recovery statements (e.g., \toparlandım") without evidence of growth.

Examples: Psychological \Psikolojik olarak çöktük." \Korku etkisindeydik." \Stres yaşadım, ağladım."

SocialDisruption \Babamdan iki ay ayrı kaldım." \Arkadaşlarımdan uzak kalmak zordu." \Şehrimizi terk etmek zorunda kaldık."

Academic \Sınav stresi de eklendi." \Performansım düştü." \Motivasyonum kalmadı."

Material \Evimiz hala yapılmadı." \Bazı akrabalarımızı kaybettik." \Maddi sıkıntı çektik."

MinImpact \Hiç etkilemedi." \Beni pek etkilemedi, pek bir şey değişmedi."

PositiveGrowth \Olgunlaştım diyebilirim." \Yeni şeylerin farkına vardım." \Hayata farklı bakmaya başladım."

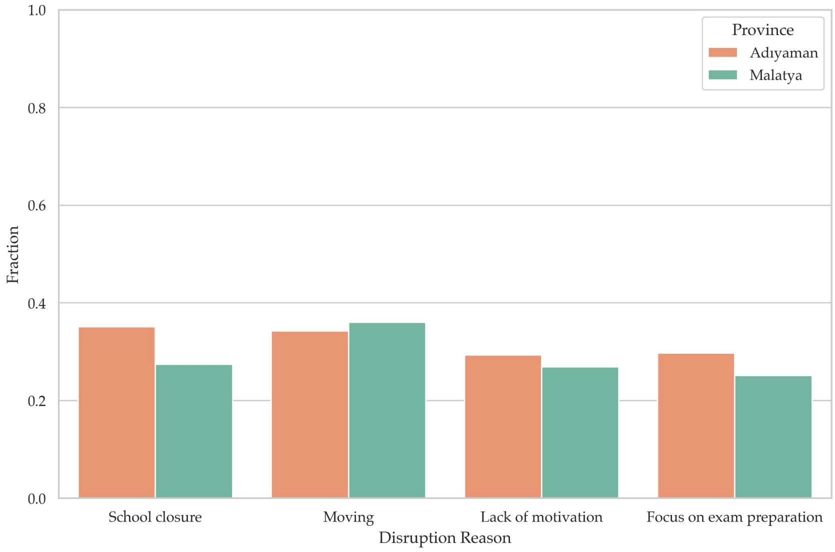
Message: "message"

ONLY return a JSON object with integers 0 or 1. Do NOT include explanations or extra text.

Example output: "Psychological": 1, "SocialDisruption": 0, "Academic": 1, "Material": 0, "MinImpact": 0, "PositiveGrowth": 0

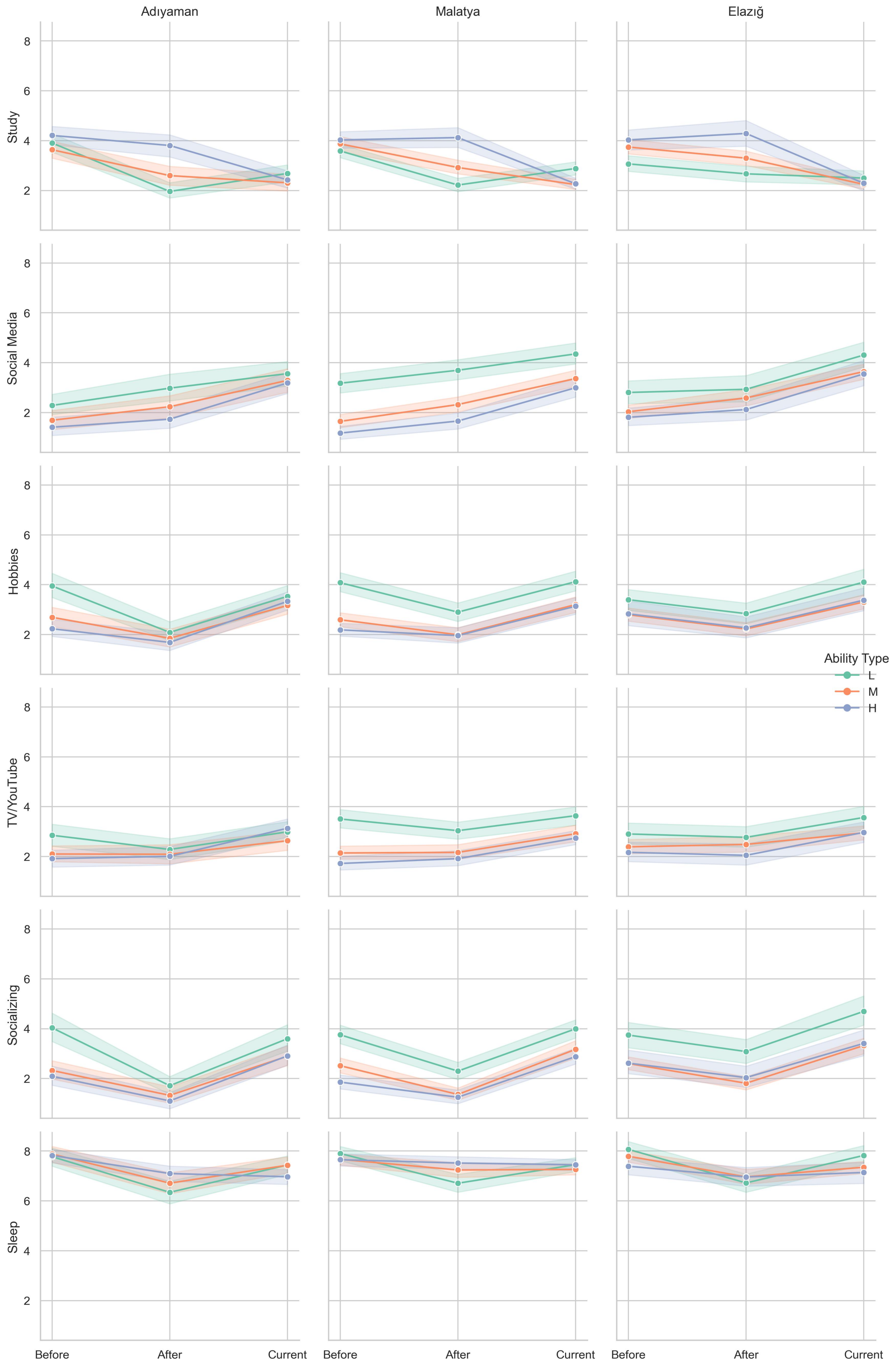
D Additional Descriptive Facts

Figure A8: Disruption Reasons by Province



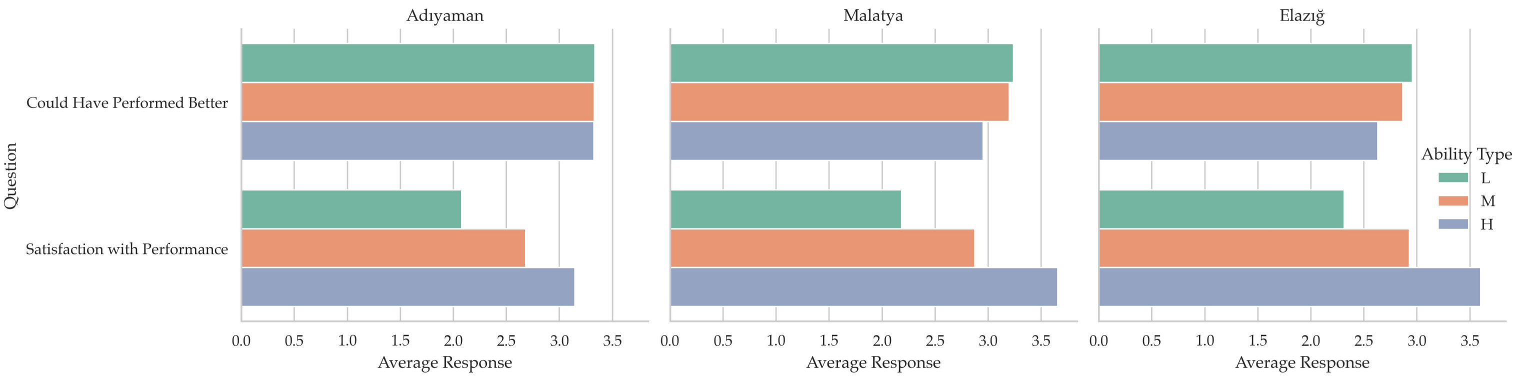
Notes: This figure shows the fraction of school disruption reasons by High and Medium impact provinces.

Figure A9: Time Use Before and After the Earthquake



Notes: This figure shows the time use before and after the earthquake.

Figure A10: Exam Performance Satisfaction

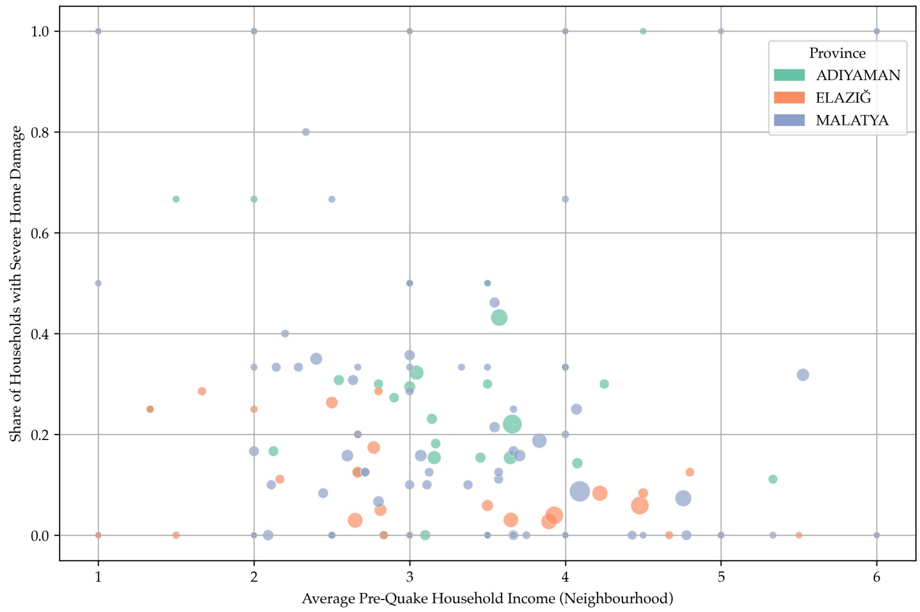


Notes: This figure shows the exam performance satisfaction by province.

E Empirical Analysis

E.1 Identification Details

Figure A11: Neighborhood Income vs. Home Destruction



Notes: This figure shows scatter plot of household income and self reported home destruction at the neighborhood level.

Table A6: PCA Factor Loadings for Psychological Scales

		Eigenvalues		Factor Loadings			
Index	**Variable**	**1st PC**	**2nd PC**	**1st PC**	**2nd PC**	**Cronbach's α**	**KMO**
GAD		2.5	0.77			0.74	0.8
	Felt nervous or anxious			0.47	0.09		
	Couldn't stop worrying			0.48	-0.17		
	Had trouble relaxing			0.48	-0.01		
	Easily annoyed or irritable			0.39	0.76		
	Felt afraid something awful might happen			0.4	-0.63		
CAPSAW		2.74	0.95			0.73	0.83
	Felt good about self			0.4	0.13		
	Felt happy			0.43	-0.23		
	Felt people cared			0.38	-0.14		
	Felt safe			0.38	-0.32		
	Capable of doing things			0.41	-0.16		
	Dealt with problems well			0.37	0.23		
	Helpful to others			0.24	0.86		
Resilience		3.61	1.0			0.79	0.89
	Adaptable to change			0.26	0.46		
	Dealt with problems well			0.37	0.11		
	Used humor to cope			0.25	0.55		
	Coped with stress			0.31	-0.16		
	Achieving goals			0.34	-0.4		
	Stayed focused			0.34	-0.41		
	Saw self as strong person			0.39	-0.14		
	Handled unpleasant feelings			0.34	-0.05		
	Empathized with others			0.21	0.05		
	Communicated effectively			0.3	0.3		
Grit		2.02	1.21			0.57	0.66
	Distracted by new ideas			0.15	0.56		
	Hard worker			0.43	-0.28		
	Not persistent with goals			0.24	0.6		
	Difficulty maintaining focus			0.43	0.33		
	Finished what was started			0.54	-0.16		
	Diligent			0.5	-0.32		
Parent-reported GAD		2.36	0.7			0.77	0.76
	Felt nervous			0.52	-0.31		
	Couldn't stop worrying			0.52	0.22		
	Easily annoyed			0.49	-0.58		
	Felt afraid			0.46	0.72		
Parent-reported Resilience		4.38	0.95			0.85	0.91
	Adaptable			0.29	-0.22		
	Dealt with problems			0.34	-0.2		
	Used humor			0.26	-0.35		
	Coped with stress			0.27	-0.36		
	Achieved goals			0.35	-0.07		
	Stayed focused			0.34	-0.01		
	Strong person			0.36	0.06		
	Handled feelings			0.34	0.01		
	Empathized			0.28	0.65		
	Communicated effectively			0.31	0.48		
Parent-reported Grit		1.84	1.31			0.55	0.57
	Distracted by ideas			0.21	0.57		
	Not persistent			0.38	0.55		
	Difficulty focusing			0.48	0.2		
	Finished tasks			0.56	-0.36		
	Diligent			0.52	-0.44		

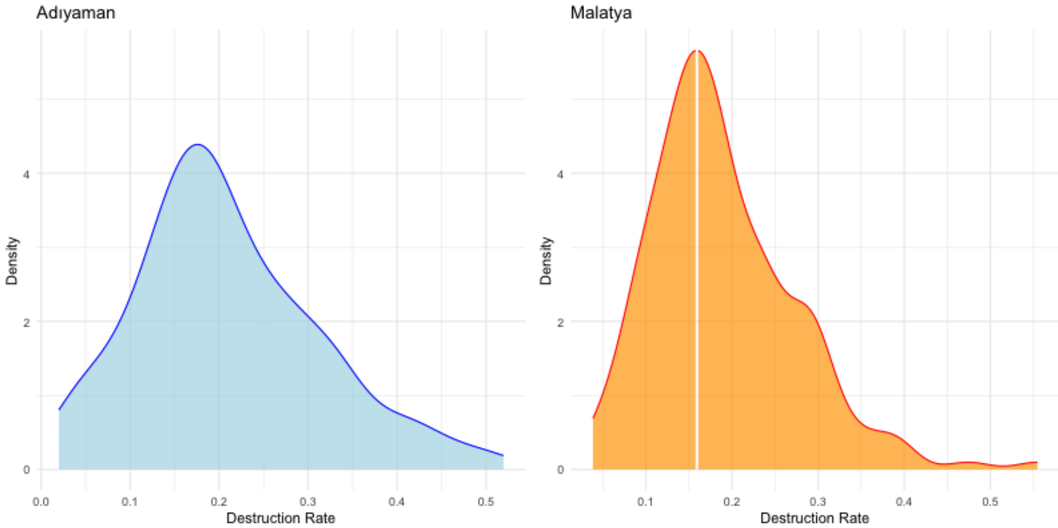
Table A7: Correlations Between Self- and Parent-Reported PCA Scores

Scale	**Pre**	**Post**
GAD	0.23	0.23
Grit	0.27	0.36
Resilience	0.28	0.35

Table A8: Correlations Between PCA-Based Psychological Traits and Academic Scores

Score	**Student Grit**	**Student Resilience**	**Parent Rep. Grit**	**Parent Rep. Resilience**
9th Grade Math	0.02	-0.00	0.11	0.18
9th Grade Turkish	0.03	0.00	0.09	0.18
Exam History	0.09	0.01	0.11	0.17
Exam Math	0.05	0.01	0.12	0.17
Exam Science	0.05	0.04	0.09	0.18
Exam Turkish	0.00	0.02	0.06	0.14
Middle School GPA	0.05	0.01	0.05	0.13
N	1463	1460	983	974

Figure A12: Distribution of Destruction Rate



Notes: This figure shows the distribution of destruction rates based on scraped damage data.

E.1.1 Destruction Rate

F Additional Results

F.1 Heterogeneity

Table A9: Heterogeneity by Gender: Academic Outcomes

	Math	Turkish
Panel A: Female		
Mid School GPA	0.11*** (0.03)	0.07* (0.04)
Study Time (Pre)	0.04** (0.02)	0.02 (0.02)
ΔST	0.19*** (0.05)	0.08* (0.05)
Home Damage	-0.07* (0.04)	-0.13*** (0.04)
HIA $\times$ ΔST	0.15* (0.08)	0.20** (0.08)
MIA $\times$ ΔST	0.16** (0.07)	0.14** (0.07)
Constant	-0.19** (0.08)	0.04 (0.09)
Observations	742	742
R^2	0.157	0.107
Panel B: Male		
Mid School GPA	0.03 (0.04)	0.05 (0.04)
Study Time (Pre)	0.09*** (0.02)	0.04 (0.02)
ΔST	0.17*** (0.05)	0.15** (0.06)
Home Damage	-0.08* (0.04)	-0.08 (0.05)
HIA $\times$ ΔST	0.20*** (0.08)	0.03 (0.09)
MIA $\times$ ΔST	0.09 (0.09)	0.07 (0.08)
Constant	-0.21** (0.09)	-0.29*** (0.10)
Observations	661	661
R^2	0.114	0.048

Notes: Regressions are estimated separately by gender. Coefficients with robust standard errors in parentheses. Stars denote significance at the 10%, 5%, and 1% levels. Additional controls: parental academic input index, temporary location, and CAPSAW well-being index.

Table A10: Heterogeneity by Household Income: Academic Outcomes

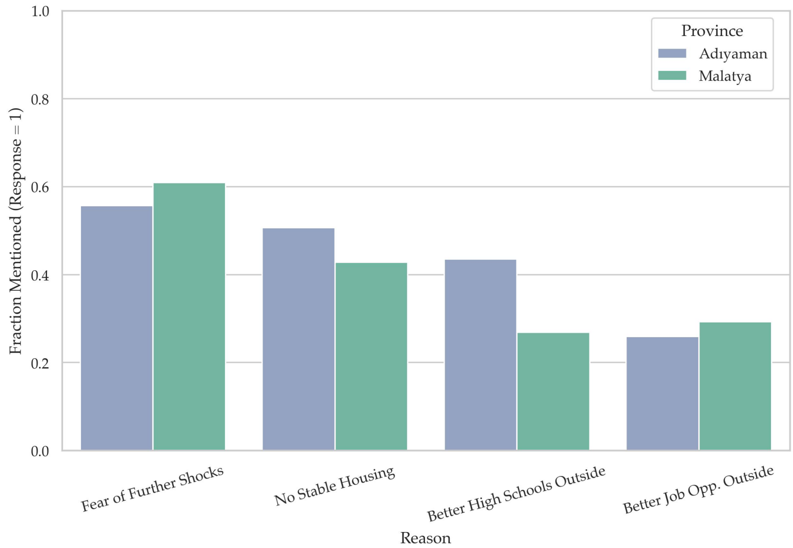
	Math	Turkish
Panel A: Income Low		
Mid School GPA	0.04 (0.06)	0.01 (0.07)
Study Time (Pre)	0.07** (0.03)	0.08** (0.03)
ΔST	0.11 (0.12)	0.18* (0.10)
Home Damage	-0.01 (0.05)	-0.14** (0.06)
HIA $\times$ ΔST	0.19 (0.16)	-0.06 (0.15)
MIA $\times$ ΔST	0.15 (0.15)	-0.02 (0.14)
Constant	-0.60*** (0.15)	-0.82*** (0.17)
Observations	320	320
R^2	0.101	0.128
Panel B: Income Mid		
Mid School GPA	0.01 (0.05)	0.02 (0.05)
Study Time (Pre)	0.07*** (0.02)	0.02 (0.02)
ΔST	0.18** (0.07)	0.16** (0.06)
Home Damage	-0.10** (0.05)	-0.07 (0.05)
HIA $\times$ ΔST	0.16* (0.10)	0.14 (0.09)
MIA $\times$ ΔST	0.19** (0.09)	0.04 (0.08)
Constant	-0.13 (0.11)	-0.18 (0.11)
Observations	488	488
R^2	0.165	0.089
Panel C: Income High		
Mid School GPA	0.12*** (0.04)	0.11* (0.06)
Study Time (Pre)	0.05* (0.03)	-0.03 (0.05)
ΔST	0.06 (0.09)	0.00 (0.11)
Home Damage	-0.11 (0.07)	-0.07 (0.08)
HIA $\times$ ΔST	0.27** (0.11)	0.18 (0.14)
MIA $\times$ ΔST	0.13 (0.11)	0.16 (0.14)
Constant	0.13 (0.16)	-0.16 (0.20)
Observations	205	205
R^2	0.138	0.196

Notes: Regressions are estimated separately by household income group. Robust standard errors in parentheses. Stars denote significance at the 10%, 5%, and 1% levels. Additional controls included in regressions: gender, parental academic input index, temporary location, and CAPSAW well-being index.

G Robustness

G.1 Migration

Figure A13: Reasons Friends Did not Return



Notes: This figure shows the reasons why friends did not return to their hometowns after the earthquake.

Table A11: Migration Summary by Province

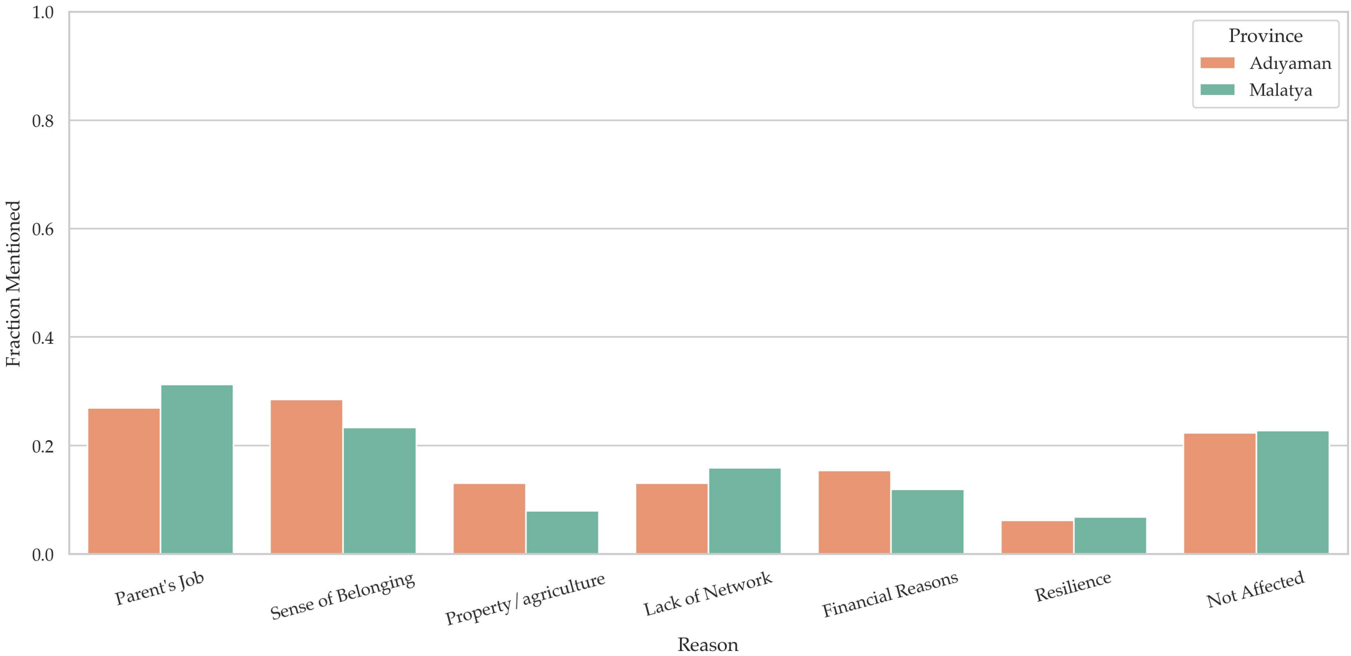
	HIA	MIA
Moved (Share)	0.667	0.725
Moved (SD)	0.472	0.447
Num. Moves — moved	2.233	2.144
Any Move to Big City	0.146	0.215

Table A12: Migration Patterns by Move Number After the Earthquake

	HIA			MIA		
	1st Move	2nd Move	3rd+ Move	1st Move	2nd Move	3rd+ Move
Migration Patterns						
Mean Num. of Indvs	6.517	5.132	5.263	5.864	4.575	5.491
Moved to Big City (%)	17.490	5.882	7.627	26.907	7.902	9.434
Avg Duration (months)	1.228	2.627	1.577	1.337	3.027	1.714
Housing Type Share (%)						
Self House	5.645	32.323	35.965	10.000	44.602	32.000
Rented House	17.742	20.202	21.053	23.182	22.159	22.000
Village House	16.532	3.535	0.877	10.227	4.261	6.000
Temporary House	4.839	12.121	17.544	2.727	6.250	10.000
Tent	4.435	6.061	3.509	1.364	1.989	2.000
Philanthropists	9.677	5.556	5.263	8.182	3.693	2.000
Relatives' House	39.113	18.182	14.912	40.455	17.045	24.000
Hotels	2.016	2.020	0.877	3.864	0.000	2.000
Move Reason Share (%)						
Back to home	0.000	3.922	0.847	0.000	1.907	3.774
Education	17.490	38.235	27.119	18.220	33.243	30.189
Psychological reasons	7.224	13.725	9.322	12.924	11.444	1.887
Escape unsafe living conditions	65.399	10.784	5.932	58.263	7.629	13.208
Financial reasons	0.760	1.961	0.847	1.059	4.360	7.547
Other	9.125	31.373	55.932	9.534	41.417	43.396

Notes: Statistics show patterns across different move numbers. 3rd+ Move combines third, fourth, and fifth moves. Housing and reason shares are percentages within each move category. Duration calculations: 1st move is from first to second move, 2nd move is from second to third move. Same-month moves are assumed to be 15 days (0.5 months).

Figure A14: Reasons for Not Moving



Notes: This figure shows the reasons for not moving across all locations

Table A13: Reasons Why Friends Did Not Return to Hometown After Earthquake by Province

	HIA		MIA	
Reason for Not Returning	N	Percentage	N	Percentage
Lack of stable housing in their hometown	240	50.4%	377	43.2%
Better employment opportunities outside	240	26.2%	377	29.2%
Better high school options outside	240	42.9%	377	27.1%
Fear of further shocks	240	55.8%	377	60.7%
Lack of social options	240	19.6%	377	17.0%

Notes: Based on responses from individuals whose friends did not return to their hometown after the earthquake. Percentage calculated as frequency divided by valid responses (excluding missing values) within each province. Respondents could select multiple reasons.

Table A14: Mean Perceptions About People Who Left, by Province and Migration Status

	HIA						MIA					
	Non-movers			Movers			Non-movers			Movers		
Statement	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N
Leavers are higher income	3.39	1.02	99	3.18	1.08	182	3.11	1.11	140	3.09	1.17	354
Leavers are younger	3.22	1.01	99	2.99	1.09	182	2.91	1.09	140	3.07	1.13	354
Leavers are more educated	3.22	1.11	99	3.01	1.19	182	3.09	1.15	140	3.04	1.13	354
Leavers are less resilient	3.19	1.04	99	3.35	1.06	182	3.11	1.05	140	3.16	1.11	354

Notes: Table reports mean Likert-scale responses (1 = Strongly Disagree to 5 = Strongly Agree) to statements about people who permanently left the region, by province and respondent migration status. Non-movers are those who have never moved. Movers are those who have moved at least once.

Table A15: Lee Bounds versus Baseline Estimates

Province	Term	Math			Turkish		
		Baseline	LB	UB	Baseline	LB	UB
HIA	ΔST × Prov	0.186	0.186	0.247	0.114	0.107	0.128
HIA	Main	-0.040	-0.203	-0.040	0.017	-0.183	0.094
MIA	ΔST × Prov	0.153	0.183	0.214	0.062	0.064	0.093
MIA	Main	-0.210	-0.411	-0.107	-0.087	-0.292	-0.000

Notes: Baseline columns report coefficients from the full-sample regression (Table 3, specification (2)). LB and UB are Lee (2009) lower and upper bounds under differential attrition of 8.6% (MIA) and 4.8% (HIA), with Control as the control province.

Table A16: Mover/Stayer Logit

Variable	Coefficient	SE
Constant	-0.162	0.194
Mid-school GPA	-0.017	0.070
Pre quake Study Time	0.068	0.032
Resilience Index	-0.136	0.066
Academic Input Index	0.123	0.108
Female	0.092	0.130
Household Income	0.063	0.039

Notes: Coefficients are from a binary logit where outcome variable is relocation status. Standard errors are HC1.

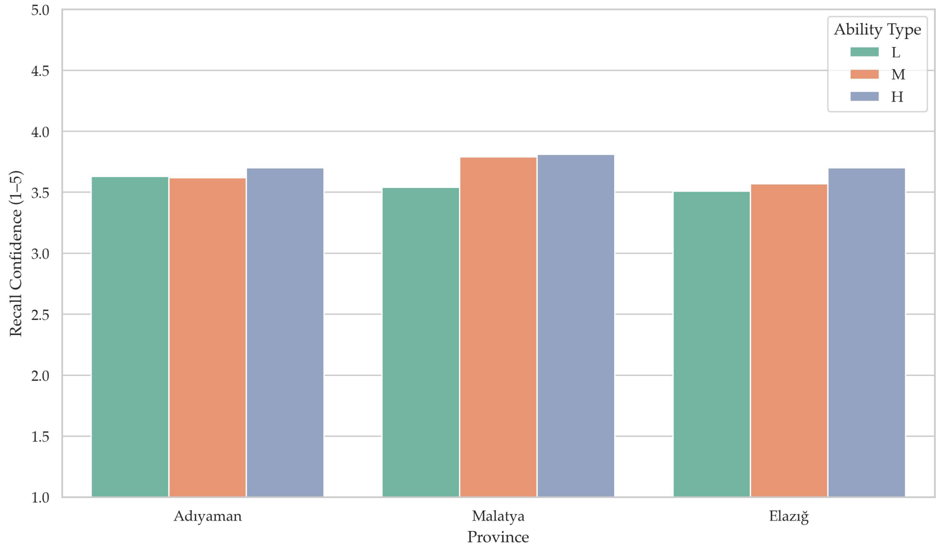
Table A17: IPW-Weighted Coefficients by Subject

		Math		Turkish	
Province	Term	Coef	s.e.	Coef	s.e.
HIA	ΔST×Prov	0.159	0.068	0.071	0.068
HIA	Main	-0.024	0.075	0.047	0.074
MIA	ΔST×Prov	0.186	0.070	0.117	0.068
MIA	Main	-0.174	0.070	-0.020	0.073

Notes: IPW-weighted WLS. Robust HC1 standard errors are reported.

G.2 Recall Bias

Figure A15: Recall Confidence by Province and Ability Group



Notes: This barplot shows the recall confidence by province and ability group. The ability group is defined as the average of the self-reported math and Turkish grades.

Table A18: Recall–Confidence Measures by Province

	Student			Parent (self)			Parent about Child		
	Mean	SD	N	Mean	SD	N	Mean	SD	N
Control	3.59	0.98	443	4.25	0.76	252	3.00	1.44	241
HIA	3.65	0.87	391	4.18	0.79	282	2.88	1.36	263
MIA	3.70	0.94	646	4.29	0.75	498	3.09	1.41	470

Notes: Each statistic uses all non-missing observations for that variable, so sample sizes differ across columns.

Table A19: Robustness to Recall Confidence

	Math		Turkish	
	Coef.	s.e.	Coef.	s.e.
HIA	-0.024	0.074	0.047	0.074
ΔST × HIA	0.180	0.070	0.112	0.068
MIA	-0.177	0.069	-0.022	0.073
ΔST × MIA	0.158	0.068	0.071	0.067
High Recall	0.150	0.059	0.119	0.064

Notes: Estimates are from IPW-weighted WLS regressions of exam scores on province dummies, resilience, recall confidence, and their interactions. Robust HC1 standard errors reported.

Table A20: Regressions Using Parent-Reported Study Time Residuals

Province	Term	Math (s.e.)	Turkish (s.e.)
HIA	$\Delta ST \times$Prov	0.103 (0.073)	0.097 (0.067)
HIA	Main	-0.025 (0.078)	0.028 (0.077)
MIA	$\Delta ST \times$Prov	0.021 (0.074)	0.043 (0.073)
MIA	Main	-0.156 (0.073)	0.006 (0.075)

Notes: Estimates are from IPW-weighted WLS regressions of exam scores using parent reported study time changes. Robust HC1 standard errors reported.

G.3 Data Collection Robustness

G.3.1 Only Attentive Respondents We check the robustness of our main results to student attentiveness and effort by applying a set of strict filters, following Chinoy et al. (2023) (who use four; we use five). Students failing any of these criteria are flagged and excluded from the sample.

Specifically, inattentive respondents are flagged as: (i) those who fail a subtle attention check,[51] (ii) those who self-report effort below "3" on a 1–5 scale, (iii) those who select extreme response options more than 50% of the time, (iv) those who select the middle option on Likert scales more than 50% of the time, and (v) those with response times at the distributional extremes.

Overall, 35% of students are flagged by at least one filter and are dropped.[52] When we restrict the sample to only attentive respondents, the main results actually get stronger: the interaction term for high-effect provinces rises from 0.20 to 0.25 for math, and from 0.12 to 0.19 for Turkish. These results are shown in column 2 of Table A21.

G.3.2 Only In-Person Surveys We also check robustness to survey mode. As described in Section B.2.2, some schools completed the survey fully online, which was less controlled than the in-person implementation. Here, we re-estimate the main results using only the in-person sample. Again, the results are somewhat stronger, as shown in column 3 of Table A21.

G.3.3 Implementation-Induced Variation Finally, we check robustness to implementation-induced variation. As discussed in B.2.2, field data collection is subject to challenges like technical issues, school administration delays, and varying levels of support. Column 4 of Table A21 presents results using a more selective sample, based on implementation-stage variables. Note that for schools not visited in person, some variables (e.g., classroom noise) are unavailable, reducing the sample size. While the main coefficients lose significance, their direction remains unchanged. The variables included are managerial

[51] This refers to missing the instruction to select both "strongly agree" and "strongly disagree" on a particular item.

[52] Most commonly for extreme response selection (27%), followed by extreme response time (20%).

cooperation, teacher support, implementer recall confidence, student noise, school being surveyed online, and school visit order.

Table A21: Data Collection Robustness: Main Results

	Full Sample	**Only Attentive**	**Only In-Person**	**Full + Impl. Controls**
Panel A: Math				
MIA × Study	0.14** (0.06)	0.17** (0.06)	0.15*** (0.05)	0.03 (0.06)
HIA × Study	0.20*** (0.06)	0.25*** (0.07)	0.24*** (0.07)	0.06 (0.08)
R^2	0.140	0.156	0.176	0.400
Observations	1403	961	1157	764
Panel B: Turkish				
MIA × Study	0.10* (0.06)	0.12** (0.06)	0.11** (0.05)	-0.01 (0.06)
HIA × Study	0.12* (0.06)	0.19*** (0.07)	0.19*** (0.07)	0.05 (0.07)
R^2	0.100	0.109	0.124	0.304
Observations	1403	961	1157	764

Notes: This table shows coefficients on province × study time interactions for Math and Turkish across samples. Standard errors are robust to heteroskedasticity. Stars: *** p<0.01, ** p<0.05, * p<0.1.

Figure A4: Survey Flow

Background Characteristics

Common Qs: Age, gender, neighborhood, school
Parents: Self and spouse education, employment status, occupation, and income.

↓

Earthquake Exposure

Common Qs: Quake-time location, post-quake movement history, physical and personal loss.
Students: Schooling disruption.
Parents: Self and spouse post-quake employment status change, property loss, government support, educational spending.

↓

Mental Health

Students: CAPSAW and GAD scales.
Parents: Same scales to measure both the self and child mental health.

↓

Noncognitive Skills

Students: Duckworth Grit Scale and Connor-Davidson Resilience Scale.
Parents: Same scales to measure both the self and child noncognitive skills.

↓

Parental Inputs

Common Qs: Communication with parents/child and parents' academic involvement.

↓

Education and Test Outcomes

Students: High school entrance exam outcomes, GPA, tutoring support, college expectations.

↓

Time Use

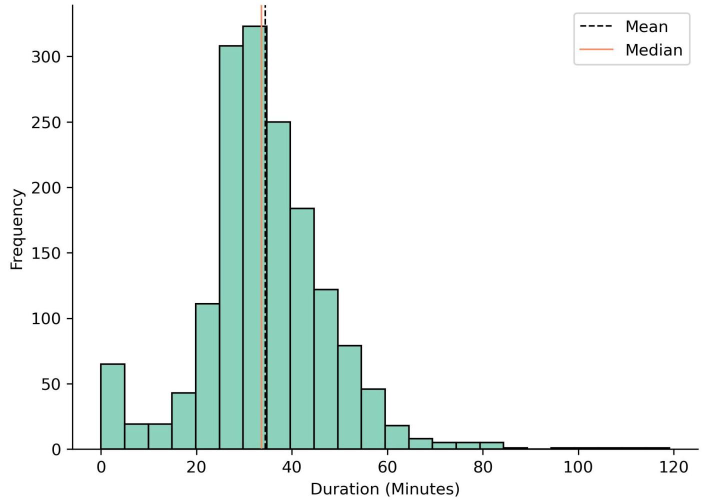
Common Qs: Students' time use on different activities.

↓

Social Network and Study Environment

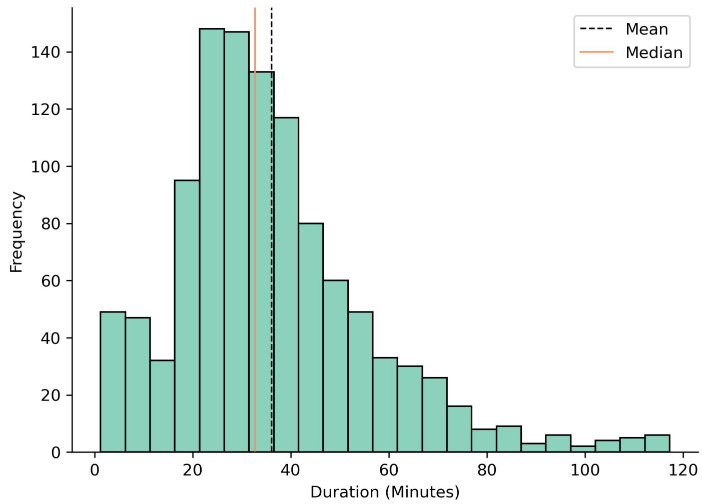
Common Qs: Post-quake social network loss, student's new study environment.

Figure A5: Students Survey Duration



Notes: This figure illustrates the distribution of time spent by students. The median time is indicated by an orange line, while the mean time is represented by a dashed black line. Responses exceeding 120 minutes are excluded from this figure.

Figure A6: Parents Survey Duration



Notes: This figure illustrates the distribution of time spent by parents. The median time is indicated by an orange line, while the mean time is represented by a dashed black line. Responses exceeding 120 minutes are excluded from this figure.