Cognitive Endurance as Human Capital*

Christina Brown† Supreet Kaur Geeta Kingdon Heather Schofield

December 19, 2019

(Preliminary and Incomplete Draft - Do Not Cite or Circulate)

Abstract

Cognitive capacity—a key predictor of labor productivity—has traditionally been viewed as a fixed component of human capital. This project reexamines this view by testing whether attentional ability is endogenously shaped through one's socioeconomic environment. We focus on a specific dimension of attention: cognitive endurance, or the ability to sustain focus toward a task. We first document a novel fact: lower-income individuals exhibit larger attentional declines than more affluent ones across disparate field settings in both rich and poor countries—worker productivity, voting, and school tests—and these declines help explain performance differences among the rich and poor. Next, through a field experiment with 1,650 low-income Indian primary school students, we increase the time devoted to focused cognitive activity during the school day, using either math or non-academic content. Each of these interventions improves cognitive endurance across a variety of unrelated domains—–academic performance, listening retention, and IQ, as well as on traditional attentional ability measures—–indicating that our interventions affected an underlying core resource. These findings suggest that worse schooling environments may disadvantage the poor by hampering the development of cognitive capacity.

*We gratefully acknowledge generous funding and support by USAID DIV, the Global Engagement Fund at the University of Pennyslvania, and The Weiss Family Program Fund for Research in Development Economics. We thank seminar audiences at NBER/BREAD, The University of Pennsylvania, and UC Berkeley. We thank Rolly Kapoor, Lubna Anantakrishnan, Simranjeet Dhir, and Deepika Ghosh for excellent research assistance. All remaining errors are our own. We received IRB approval from the University of California, Berkeley (protocol number 2017-08-10250) as well as the Institute for Financial Management and Research (IFMR) in India.

†Brown: University of California, Berkeley (christinabrown@berkeley.edu); Kaur: University of California, Berkeley and NBER (University of California, Berkeley); Kingdon: University College London (g.kingdon@ucl.ac.uk); Schofield: University of Pennsylvania (hschof@wharton.upenn.edu)

1 Introduction

Cognitive ability is a fundamental input into labor productivity and decision-making. Traditionally, cognition has been viewed as a fixed component of "ability". In this paper, we reexamine this view. We postulate that cognitive capacity is not fixed, but rather is (at least partly) shaped through one's socioeconomic environment. Moreover, we argue that this may help explain why the poor exhibit worse cognition than the rich in a wide variety of settings.

Literatures in psychology and sociology have long discussed the possibility that environmental factors, including schooling, may have the potential to alter cognitive capacity (e.g., Scribner and Cole (1973).[1,2] The fundamental nature of basic cognition suggests that—because it influences performance across disparate daily activities—even small gains could be highly valuable when aggregated. Consequently, this hypothesis has potentially important implications for understanding human capital development and how it may mediate inequality.

We focus our study on a core cognitive resource, attention, which is thought to underlie all conscious activity—both cognitive processes such as solving a math problem and also “non-cognitive” activities such as exerting self-control (Chun et al., 2011). A growing literature in behavioral economics has begun to explore the central role of attention in economic life—for example, how limited attention can constrain decision-making and behavior in a wide variety of domains (Chetty et al., 2009; Hanna et al., 2014; Gabaix, 2019; Gabaix and Laibson, 2006; Gabaix et al., 2006; Angeletos and Sastry, 2019). Yet, despite its potential importance to long run educational attainment, decision-making, and labor market outcomes, little is known about whether attention is malleable or a part of "innate ability." And if the ability to sustain attention is malleable, can this fundamental capacity be developed though traditional methods of human capital accumulation such as schooling?

In this paper we first document a novel fact: lower-income individuals are less able to sustain cognitive performance over time than more affluent ones across disparate field settings—worker productivity, voting, and school tests. For example, the international TIMSS test is used to assess academic achievement and preparedness of students in the U.S. and other countries. Question order is randomized, so that some questions appear earlier in the test for some students and later for others. Students show marked declines in performance over the course of the test—consistent with cognitive fatigue. An average student is about 7.5% less likely to answer a question correctly if it occurs at the end of the test rather than the beginning. This pattern is robust to controlling for question fixed effects, and is not explained by students simply not finishing the test. Moreover,

[1] More recently, development economists have documented the impact of poverty and one's environment on temporary shifts in cognitive function (Mani et al., 2013; Mullainathan and Shafir, 2013; Bessone et al., 2019). While related, our focus is distinct: we focus on long-run shifts in capacity rather than temporary deviations from that capacity.

[2] There is evidence from laboratory studies that other executive functions – such as fluid intelligence – are malleable (Klingberg et al., 2005; Jaeggi et al., 2008). Typically, these studies find little evidence of training transferring beyond those directly trained (Bergman Nutley et al., 2011; Holmes et al., 2009; Klingberg et al., 2005; Diamond, 2013). However, as is typical of many laboratory based studies, sample sizes are typically quite small (typically 15-40 individuals per arm), such that these studies may simply be underpowered to detect effects.

low income students show substantively greater performance declines over time than richer ones. This is true when comparing rich vs. poor students within the US, and also students in rich vs. poor countries globally. These differences in decline are meaningful—explaining, for example, 10% of the performance gap between higher and lower income US students. While only suggestive, this indicates that cognitive fatigue is a determinant of performance, and varies systematically with socioeconomic status across settings. These declines are not unique to academic tests; we document similar robust empirical patterns in various observational datasets from disparate domains. More generally, such cognitive declines have implications for productivity in many domains, including learning in school and various occupations–from judges making decisions throughout the day, to doctors treating patients over the course of a long shift, to TSA screeners examining packages at the airport.

These empirical patterns provide the impetus for our hypothesis that attentional capacity may be a malleable component of human capital, helping to explain why it may correlate systematically with income. We posit that this pattern indicates that schooling may do more than traditionally thought: it might not only build skills/knowledge, it may also shape basic cognition. This view may help explain puzzling findings in the labor literature that indicate that schooling interventions boost long-term outcomes for students through channels that seem to operate beyond test scores or academic skills (e.g. Chetty et al. (2011)).

To test this hypothesis, we conduct a randomized controlled trial with 1,650 low-income primary-school students in India to examine whether attentional capacity is malleable. The goal of our interventions is to increase the amount of sustained time students spend on cognitively-challenging activity. This marks a deviation from the status quo in under-resourced schooling environments, where we document that students spend little time exerting focused attention at school or at home. We develop two interventions. In the first treatment, students practice mathematics problems for sustained periods using an individualized adaptive tablet-based math platform. This treatment mimics what good schooling does–providing cognitively challenging focused activity within the context of academic learning. In the second treatment, students play cognitively demanding games for sustained periods of time. This enables us to increase time spent on focused activity, which is free of any academic content. The control group continues to receive a status-quo math study hall period, where students do a small number of math problems copied from the chalkboard (as is standard in this setting). For each of the two treatment groups, the study hall period is replaced with these cognitive practice periods 1-3 times per week. In total, each treatment resulted in 20 hours of focused practice over the course of a 6 month period during the academic year. Note that while the interventions are delivered via tablet, this is not consequential for testing our hypothesis—it was simply a convenient implementation approach. Moreover, none of our outcome measures test students on any tablet-based medium.

We examine students' attentional endurance via three tests across diverse domains: math, listening retention, and IQ.[3] We develop a simple and generalizable field approach to measure cognitive endurance in each domain. For each test, we randomize the order in which questions appear—so that a question appears early in the test for some students and later for others. We then examine declines in performance over time, controlling for question fixed effects. These tests are all administered using paper and pencil rather than any technology based administration. In addition, students are provided ample time to finish each test, so that declines cannot be driven by unanswered questions (we also directly verify this in the data).

We find that our interventions improve cognitive endurance across each of these unrelated domains: mathematics, listening retention, and IQ. In addition, the effects are relatively large––ranging from 20 to 60% of the total decline experienced by the Control group—despite the fact the students engaged in only 20 hours of training. If these results are applied to data from an international achievement test, it would cut the gap between high and low income countries in performance decline in half.

A potential concern with interpreting our findings is that our treatments could simply have made students better at each given test type, e.g. math. We overcome this concern in two ways. First, the domains over which we test attentional declines are markedly different from at least one if not both of the interventions. For example, neither intervention trains students on listening. Similarly, IQ as measured by Ravens Matrices test is considered to be an immutable characteristic. While the math treatment arm did train students on math, the games treatment arm did not. Second, if our interventions had inadvertently "taught" students how to perform better on a particular domain, then we would expect to see level shifters in performance even at the start of the test (e.g. the first decile of questions). However, we find no evidence for level effects in the beginning of each test. This indicates that our interventions did not train students on content. Rather, they made them more able to maintain cognitive focus in carrying out these tasks. The lack of level effects in the beginning also help rule out other potential confounds such as confidence or motivation.[4]

Treated students also improved on an index of traditional cognitive psychology measures of the ability to sustain attention as well as an index of classroom behaviors adapted from the Vanderbilt ADHD diagnostic teacher rating scale.[5]

Taken together, these results suggest schooling content and quality can meaningfully shape fundamental cognitive skills that are not directly taught. Because this particular skill is an input to a wide variety of activities in school and the labor market—for example, both garment factory workers and surgeons rely heavily on this skill—these differences in instruction may be a key channel through which achievement gaps across socio-economic groups persist.

[3]The math test covers the standard math curriculum, making it directly policy relevant. The listening test mimics the skills necessary to listen to a teacher and absorb the information by asking students to listen to series of short stories and respond to short factual questions. The IQ test is frequently used measure of fluid intelligence, Raven's Matrices.

[4]In addition, we rule out impacts on grit by documenting that the treatments did not affect students' performance on questions after they (randomly) encounter a particularly difficult question.

[5]Observers rating student behavior were blind to treatment condition.

This paper makes four contributions. First, we demonstrate that a fundamental element of cognition—the ability to sustain and direct attention—which has traditionally been considered a fixed element of "ability" is malleable. These changes are not limited to the domain that is directly trained. Rather they are broad-based and generalizable across domains as divergent as completing a math problem and listening to and understanding information.[6]

Second, we extend our understanding of the role of schooling in human capital accumulation. Schooling's potential influence on basic cognition may provide an alternative explanation for the observed education, wage, and health gains observed from interventions which improve schooling quality (Chetty et al., 2009; Heckman et al., 2006; Alan and Ertac, 2018; Kautz et al., 2014). Further, differences in pedagogy and the quality of the schooling environment by income have the potential to widen disparities in such skills. More affluent students naturally obtain practice throughout their school day, inputs that many low-income students often fail to receive (for the Evaluation of Educational Achievement , IEA).[7]

Third, this research speaks to our interpretation of the correlation between socio-economic status and cognitive performance (Banerjee and Mullainathan, 2008; Balart et al., 2018; Lawson et al., 2014; Hackman et al., 2015). To understand whether this gradient is a cause or a consequence of poverty it is crucial to determine whether basic cognitive skills are malleable and how they are formed. Our research suggests that growing up in poverty may limit cognitive development through low-quality schooling and limited opportunities to practice such basic skills. If individuals do not have the opportunity to practice exerting sustained cognitive focus, this capacity will be under-developed and may generate broad negative consequences for cognitive functioning, labor productivity, and economic life.

Finally, our study offers a micro-foundation for the persistence of achievement gaps between the rich and the poor. Because attentional capacity is an input for so much of human activity, environments which fail to promote the development of this skill are likely to generate long run differences in both schooling and labor market outcomes. These gaps may create feedback loops (e.g. via enrollment in lower quality schools) that span generations and decrease ones chance of escaping poverty.

2 Cognitive Endurance Across Domains

Attention is a core cognitive resource which underlies all activity. Acting as a constraint on processing power, attentional limits are likely to influence economic decision-making in myriad ways (Gabaix, 2019). The ability to sustain and direct attention toward a task over time is particularly

[6]Notably, the methodology we develop to measure cognitive endurance in this study can also be applied broadly, allowing other researchers to evaluate impacts on this skill via other tests they are conducting simply be randomizing the order of the questions asked.

[7]The specific elements of schooling which promote or hinder active engagement in different environments is an important area for future work.

relevant to a wide variety of settings and professions. To demonstrate the broad applicability of this element of cognition, we highlight motivational examples of cognitive declines and their economic importance below.

Academic Test Performance. The focus of this study, education, is an area where the ability to sustain focus is likely to be important both to learning and to performance on exams. The Trends in International Mathematics and Science Study (TIMSS) is administered to thousands of 4th, 8th, and 12th grade students every four years across the United States and a selection of other countries which vary substantially in income (e.g. Thailand, Armenia, and Singapore). Importantly the order in which blocks of questions are administered is randomized within the test, generating a consistent average difficulty across the exam and allowing us to estimate declines in attention across the exam.[8]

As shown in Figure 1, performance declines across the test are substantial. A student is roughly 7% more likely to get a question right at the beginning of the test than if they answered the same question at the end of the exam.

Voting Behavior. Importantly, the effects of cognitive endurance may also generalize well beyond the more readily apparent domains discussed above. For example, many everyday decisions and actions such as planning a party or planning for one's retirement require sustained focus. Although less obvious, these effects are likely to be pervasive. For example, Augenblick and Nicholson (2015) provide evidence of similar declines in attention in voting behavior (results are reproduced in Figure 2). Using quasi-random variation in the order of ballot initiatives, the authors find that individuals are substantially more likely to vote the default option when items that are further down-ballot. These effects are substantial enough to alter the outcome of 6% of the propositions in their data set.

Other contexts. Attentional declines are prevalent in a wide variety of other contexts and populations as well. For example, Brachet et al. (2012) find that fatigue during long paramedic shifts "result in a 0.76 percent increase in 30-day mortality". Danziger et al. (2011) find that judges become significantly harsher in their judgements as their shifts progress, but leniency returns following a break. More broadly, Warm et al. (2018), documents the crucial role of attentional capacity in a wide variety of professions such as sentries, truck drivers, air traffic control operators, and industrial quality control.

Correlation with Socio-economic Status. Although declines in attention over time are a fairly universal phenomena, attentional capacity does vary substantially across individuals. Further, this

[8] Differences in motivation could also influence the rate of decline. While Zamarro et al. (2019) do find significant differences in effort exerted, these differences largely influence the initial level of performance rather than the rate of decline. Similarly, long tests which do not allow students to finish may also drive declines in some contexts. However, completion rates are over 95% for all of the exams in this study and results are similar when considering only completed questions.

capacity is correlated with one's socio-economic status and current income (Hackman et al., 2015; Farah et al., 2006; Lawson et al., 2014; Balart et al., 2018; Mani et al., 2013; Banerjee and Mullainathan, 2008; Clearfield and Jedd, 2013). The differences are often substantial. For example, during the TIMSS test, the rate of decline in performance among students in low-income countries is roughly twice the rate of decline among students in high-income countries (Figure 3). Similar differences are found among high-income and low-income students within the United States (Figure 4). These gaps are meaningful: the difference in the rate of decline accounts for 7% of between country and 10% of within country test score differences.

3 Experimental Design

3.1 Background

Sample. To test these ideas, we conducted a randomized field experiment with 1,650 students in 6 Indian primary schools in and around Lucknow, India. These schools serve students in low to middle-income households, with per capita incomes between \$1.50 and \$5 per person per day (a common range for private schools in India). All students in grades one through five of these schools (ages five to eleven) were enrolled into the program and randomized at the individual level, stratified by class section and baseline math test scores.

Context. Developing countries have made enormous gains in boosting school enrollment; primary school completion is now 96% in India (World Bank 2014). Yet, despite the growth in enrollment, the quality of education remains dismal. For example, 53% of third to fifth graders cannot do basic subtraction (first grade math) (Pratham, 2011). Weaker students are promoted through grades, but fall so far behind, they are unable to engage in class material. Classrooms of such diverse achievement are difficult for teachers to manage, leading to a disruptive environment and poor instruction. Pedagogy which promotes rote memorization is common. Parents do not expect children to do work outside of school, leading to little focused academic at home. Consequently, students seldom have the opportunity to engage in focused cognitive activity for sustained periods of time either inside or outside the classroom. Such conditions are typical of many developing countries (Bank, 2004).

3.2 Experimental Arms

We implement two distinct sets of interventions, each of which is randomized at the student level. The goal of each intervention is to increase the time spent continuously engaged in attention-heavy tasks. The first intervention mimics the way in which students traditionally practice exerting directed attention over sustained periods in schooling settings—academic practice, in our case by solving math problems. Students in this arm are provided with adaptive software on a tablet which is customized for each student's baseline achievement level—and also makes it impossible to cheat from one's neighbors. Consequently, each student is individually engaged in solving problems for the entire period, a feature which has been shown to be important in developing attentional skills (Diamond, 2013; Klingberg et al., 2005).[9]

Math Treatment. The Math Treatment condition substantially increases the likelihood that students engage in sustained focus during study hall periods. However, it also potentially boosts academic learning. In contrast, our hypothesized mechanism suggests that any sustained cognitive engagement should deliver attentional benefits. Consequently, we include an additional treatment arm, which requires students to engage in cognitive activities that do not entail any academic learning or practice.

Games Treatment. This second intervention — the Games Treatment — requires students to engage in cognitive activities such as attention-oriented games for sustained periods, without any academic learning. Students spend the study hall periods engaged in non-instructional cognitive games which require extended focus. For example, students play games of attentional focus such as tangrams and N-back[10]. Students cycled through seven such games, in order to promote variation and continued engagement. As in the Math condition, the content is designed to generate consistent engagement through increasing difficulty of each of the activities. The games were delivered via tablets to allow students to move at their own pace. Both the Math and Games treatments resulted in roughly 20 to 25 minutes of sustained practice for the average student in a typical 30 minute session.

Control. These two treatment arms are compared to a Control arm which dedicates the same amount of time to a traditional study hall period. During a typical study hall, teachers write a small number (typically approximately five) of problems on the board, and the students are asked to solve these problems in their notebooks. The questions used in these study halls were drawn from the same question bank as was used in the Math Treatment. Engagement levels during these study halls vary substantially by school, classroom size, and the difficulty of the questions. However, students typically finish well before the period is over and then talk to neighbors, or (among weaker students who are not at grade level) do not attempt the questions at all. On average based on monitoring conducted by a program staff member, we estimate that students in this arm undertake five the ten minutes of sustained practice during the 30 minute period.

[9]Notably, this approach to increasing practice of sustained focus is likely to be context specific. While some features of the interventions such as their adaptive nature are likely to be important broadly, other features such as the technology-based solution, may be less relevant in contexts where tablets are less novel and hence less engaging.

[10]Tangrams tasks students with rotating and moving objects to generate a given shape. N-back presents the children with an ordered series of stimuli. They indicate whether the current stimulus is the same as the one N-previous to it.

3.3 Timing

These activities were scheduled to take place two to three times per week over four months. However, given school closures and holidays the average number of sessions undertaken was XXX, resulting in 15 to 20 hours of practice in sustaining focus among treated students.[11] These periods replaced study halls or "activities" classes such as an art period. The above design ensures that the physical time spent in the different treatment conditions (i.e. the number of study hall periods) is exactly the same across all three groups. All other aspects of students' schedules and curricula remained unchanged and consistent across experimental arms.

4 Outcomes

4.1 Overview

An important feature of basic cognition is its broad applicability; basic cognitive processes are used in nearly all activities. In order to determine whether the treatments are improving cognitive endurance rather than simply training a given task, we measure the effect of each of these interventions on cognitive endurance using three primary tests across diverse domains of performance: math, listening , and an IQ test. These field measures rest on a novel approach which enables us to directly test for whether attentional skills are observed broadly across a variety of domains/contexts.

Specifically, to measure cognitive endurance across a wide variety of domains we examine students' performance over the length of an exam and compare how much less likely a student is to get a question correct at the end of the exam than the beginning of the exam. In order to construct an identified test for these "decline" effects, we randomize the order in which questions appear across students. This means, for example, the same question item could occur as question 1, 10, etc. in a students' test packet. Test packets were randomized across students.[12]

Our key prediction is that students with better sustained attention will show less decline in performance over time, controlling for the difficulty of the question. Thus, while students in the control and treatment conditions may perform similarly well in the initial test questions, we expect to see a gap emerging over the length of the exam, where the control students lose focus and performance declines in the latter part of the exam. This prediction is a result of the training provided, which targets the ability to sustain focus rather than the test domain itself (e.g. listening to a story). Correspondingly, the prediction of reduced decline in performance is a very specific to the mechanism of sustained attention. This feature, along with the richness generated by the random ordering of questions, helps us distinguish our proposed mechanism from confounding explanations.

[11] Given that some practice was undertaken in the Control, this difference is likely to be an upper bound on the additional time spent on focused activity.

[12] The test packets were well randomized with the number of imbalances across experimental arms no more than would be expected by chance.

4.2 Primary Outcomes

This general design is applied to practical tests in three diverse and important domains:

(1) **Math.** A standard paper-and-pencil test, which focuses on the content in the math curriculum for each student's given grade level. Students' performance on the test was counted toward their final mathematics grade, giving the test natural stakes. Students are given roughly 25 minutes to complete this exam.

(2) **Ravens Matrices.** This is a non-verbal multiple-choice test of reasoning in which the participant is asked to identify the element that completes a pattern in a figure (Raven, 1936, 2000). This test is often said to capture "ability" or "IQ". Students took a shortened paper-and-pencil version of the test, adapted for appropriateness for each grade level.[13]

(3) **Listening.** This task measures students' ability to listen to a passage without losing focus, as is required in nearly all typical classroom settings. Using headphones, each student listened to a pre-recorded set of short simple stories. After each story, the student was asked questions about the content of the story, for example, "what color was the dolphin?"? In order to avoid any concerns about literacy, answers were multiple choice and visual (e.g. in the above example, green, blue, black, and grey squares to denote the color of the dolphin). After answering the three questions, the students listened to the next passage, again followed by simple multiple choice questions answered in a paper-booklet. Both the order of the passages and the order of questions within passages was randomized across students.

All tests are conducted during the school day, either in program class time or during additional study hall periods. Importantly, the tests are designed to give students sufficient time to complete the full test. This goal was met with over 95% of all students reaching the final question on the test (Table 2).[14]

4.3 Secondary Outcomes

In addition to these novel measurements which allow us to examine cognitive endurance in more "natural" settings, students also engaged to two more traditional laboratory measures of the ability to sustain focus:[15]

(1) **Symbol matching.** Students were given a paper-based workbook in which they were asked to locate and cross out specific symbols in a large matrix of randomly ordered symbols. Scores

[13] While this exam typically proceeds from the easiest to most difficult questions, with the exception of a short set of easy practice questions which are not included in the analysis, the order is randomized in this case as well.

[14] Students were instructed to take the tests in the order provided. Test monitors report these instructions nearly always followed. A more through discussion of test taking strategies as a potential confound, and our approach to rule out such strategies, are discussed in section ????

[15] These tests are explicitly designed to capture one's ability to focus attention over time, hence we anticipate improved performance throughout the test.

are a positive function of the number of symbols correctly crossed out and a negative function of the number of symbols incorrectly crossed out.

(1) **Sustained Attention to Response Task (SART).** Students look at a computer screen for ten minutes, during which time various shapes (i.e. stimuli) randomly appear and then quickly disappear from the screen. The student is tasked with simply pressing the space bar as quickly as possible each time a particular shape (i.e. a bell) appears to show that she has seen it (Peebles and Bothell, 2004). Overall performance is measured as a mixture of speed and accuracy common to the literature.

Finally, students behaviors were observed in their classrooms by individuals who were blind to treatment status. Specifically, we adapted three measures from the Vanderbilt ADHD Diagnostic Teacher Rating Scale.

(1) **Following instructions.** Students were asked to complete two activities – moving supplies from one part of the classroom to another and writing their role number on a paper and hand it to a staff member – following a class activity.

(2) **Response to auditory stimuli.** Whether students are able to notice and respond to an auditory stimuli outside the classroom.

(3) **Physical symptoms of inattention.** Whether the student shows physical symptoms of inattention (e.g. fidgeting, looking out a window, pestering their seat-mate).

The intervention was conducted between September and January. Tests were administered at four times: Baseline (September), Mid-line (December), Endline (February), and Follow-up (April). However, certain tests were randomly sub-sampled or not administered in all rounds due to logistical constraints on test administration. However, these logistical constraints impacted all arms of the study equally and did not result in any imbalances in measurement of outcomes.

5 Empirical Approach

5.1 Overview

The goal of our approach is to capture declines in attention over performance on a task in a generalizble manner. Psychologists have a variety of measures of a similar element of cognition typically referred to as "vigilance". These measures rely on attention-heavy repetitive tasks. For example, the ability to react quickly to a visual stimulus presented on a computer screen. While we include these standard measures in our analysis, this approach has limited ability to examine field measures as it requires a repetitive stimulus to ensure equal difficulty across time. Hence, we develop a new and widely-applicable approach to measure of cognitive endurance. We randomize task order and compare performance on an item earlier versus later in time, controlling for the difficulty of the specific task (e.g. a question fixed effect on a math test). Conceptually, this approach allows us to capture how someone does on a task when it comes sooner (when attention is not depleted) vs late (when attention is depleted).

5.2 Estimating Equation

We estimate:

$$Correct_{ijk} = \beta_0 + \beta_1 Treated_j + \sum_{k=1}^{10} \delta_k ItemLocation_k + \sum_{k=1}^{10} \gamma_k Treated_j * ItemLocation_k$$
$$+ \chi_i + \epsilon_{ijk}$$

Where $Correct_{ijk}$ denotes whether a question item, i, for child, j, located in decile, k, was answered correctly. β_1 captures the difference in performance at the beginning of the test. We predict $\beta_1 = 0$, with the possible exception of the math test as the groups received a differential amount of math practice. This prediction is important to ruling out potential confounds, as described further below. δ_2 - δ_{10} are decile bins which flexibly capture declines in control group.[16] While performance on a wide variety of tests declines across time, the exact pattern of decline in performance varies significantly across tests. These variable patterns of decline motivate our non-parametric empirical approach. γ_6 - γ_{10} are the coefficients of interest, indicating whether there is differential fatigue among treated students. We hypothesize these coefficients will be positive (they will ameliorate the rate of decline) and report p-values for the linear combination of coefficients in the second half of the test. χ_i are question fixed effects, controlling for the difficulty of the test item.

6 Results

Experimental arms were well-balanced at baseline on all testing outcomes. In addition, attrition was low (5%) and balanced across experimental arms (Table 1).

6.1 Primary outcomes: Math, Listening, and Ravens Matrices

Initial results support of our predictions. Relative to the control group, students in each treatment arm show an improved ability to maintain focus in a range of disparate activities: Mathematics, listening tests, and Raven's Progressive Matrices — providing evidence for the underlying generalizability of the cognitive effects. Consistent with our proposed mechanism, treatment and control

[16] To account for varied test lengths, we use question item as a proxy for elapsed time and normalize the length of all tests to 100%.

students generally perform similarly in the beginning of each test; however, control students' performance declines over time as they become cognitively fatigued, while treatment students are able to maintain attentional focus longer. This effect is first demonstrated on the Math test by plotting a local polynomial by treatment. As seen in Figure 5, students in the Math Treatment arm show similar performance to Control students in the first 40% of the exam. However, a distinct gap in performance emerges in the remaining 60% of the exam, with Treated students experiencing slower declines and greater overall performance. Although the Games Treatment students begin at a lower level of performance – likely due to the Control's additional study hall practice in math – the rate of decline among the treated students is once again, significantly shallower. Notably, the performance of the Games Treatment students catches up with that of the Control students, despite significantly less overall math practice.

We see similar overall patterns in the listening test as shown in Figure 6.[17] Initial performance is, again, not statistically different across Treated and Control students, but a gap emerges fairly rapidly. The magnitude of the gap is largest in the periods of greatest decline among the Control students. Notably, this test is one which has fixed timing (e.g. one can not skip ahead), ruling out any confounds due to test-taking strategies. In addition, neither of the treatments involved any additional time listening to an instructor, suggesting that gains can not be due to additional training on the task.

Performance on Raven's Matrices, often taken as an IQ test, is also improved by the Games Treatment, although not by the Math Treatment (Figure 7). Performance in this Games treatment arm was nearly flat over the course of the exam, while control students declined by roughly three percentage points over the 20 minute exam.

Drawing on the empirical approach described above, we also assess the magnitudes and statistical significance of these effects. As shown in Table 3, there are no significant differences in initial levels of performance across any of the tests. Pooled across all three exams, the effects are both meaningful in terms of magnitude — roughly one-third of the total decline is ameliorated — and highly statistically significant. Although not statistically significant, we also see suggestive evidence of effects earlier in the test, with a coefficient magnitude roughly one-half as large as in the second half of the test.

Similarly, we see a statistically significant reduction in declines on both the math and listening tests in the second half of the test. While the estimated coefficient on the IQ test does not reach statistical significance when pooled due to the null effect among Math treatment students, the estimated pooled coefficient implies a net reduction in declines of roughly 25% overall.

These effects are even more notable given the relatively limited training in this program. Students spend fewer than 20 hours in this program, yet spend roughly 800 to 1,000 hours per year in instruction and practice at school. While the training effects may not be linearly additive over time, they do suggest that even small differences in the instructional quality could have a substantial

[17] Plotting the raw data, there are clear "reset" effects between passages. Hence, We examine declines within each passage which are substantial.

impact on cognitive endurance over time.

6.2 Measures of Sustained Attention from Psychology

We find similar patterns in an index of sustained attention measures traditionally used in the psychology literature(Johnson et al., 2007; Oades, 2000). Our treatments improve an index of the Sustained Attention to Response Task (SART) and a Symbol Matching test by 0.05 SD ($p = 0.86$) (Table 4). In addition, the effects also appear to generalize to classroom behaviors. As seen in Table 5, treated students improve on an index of three measures of classroom attention by 0.12 SD, with the effects driven by improved ability to attend to and follow instructions and improved responses to stimuli.

6.3 Summary

The treatments produce broad and generalizable improvements in the ability to sustain focus over time. Treated students are better able to maintain focus on tests of math, listening, and IQ. The results are supported by improvements in both cognitive psychology tasks designed to measure sustained focus as well as in observations of classroom behavior.

Taken together, these results suggest that basic cognitive function is malleable and that a simple school-based intervention can improve children's ability to sustain focus over time. These patterns also support our view that the treatments do not teach new content within the test domains, but rather improve students' ability to sustain cognitive effort over time. Because question order is randomized across students, our design enables us to attribute this pattern to a change in cognitive endurance (rather than other explanations such as differences in question difficulty over time).

7 Potential Confounds

The training provided to treated participants targets the ability to sustain focus rather than the test domain itself. Hence, a core prediction of our study is that the treatments will mitigate declines in performance over time without systematically impacting the initial level of performance.[18] Correspondingly, the prediction of reduced decline in performance relative to the Control over the course of the exam is a very specific to the mechanism of sustained attention. This feature, along with the richness generated by the random ordering of questions, diverse tests, and multiple treatment arms helps us distinguish our proposed mechanism from confounding explanations. Potential confounds, and our approach to rule each out — both through design features and additional analyses — and specifically test the proposed mechanism, are detailed below.

[18] With the exception of the math test in which the Math and Control students receive additional practice relative to the Games students.

Improved Math Aptitude or Reduced Cost of Effort. The Math Treatment arm (Games treatment arm) received more (less) math practice than the Control students. The differential math practice may serve to directly alter math skills or change the cognitive costs of completing math questions through a variety of mechanisms (e.g. solving math problems requires less effort with additional practice). While these mechanisms may affect math performance, it is unclear why they should affect performance on other tests, such as listening. This mechanism also fails to explain why improvements would be observed on tests which are unrelated to the training (e.g. train in math and improve in listening).

Confidence or Motivation. The treatments could improve confidence or motivation levels. However, such mechanisms will be level shifters throughout the exam: they should improve performance in both the early parts of the test and the later parts of the test. In contrast, our mechanism predicts that the treatments will ameliorate declines rather than generate uniform improvements. In addition, in order to test for motivational impacts, we incentivize a sub-set of the tests via enticing prizes (e.g. toys, colored pencil sets, etc).[19] These additional incentives do not alter the results.

Grit. Grit or other explanations related to ability to overcome challenges are another potential confound (Duckworth and Duckworth, 2016). We leverage the random question ordering to test for this competing explanation. Specifically, by chance, some students received a version of the test which began with easier questions and some received a version which began with more difficult questions. We test if the appearance of a difficult question — either early in the test, or generically throughout — lowers later performance. We find no such impacts across many different definitions of "difficult" questions.

Improved Technology Skills. Because the training occurs on tablets, which are a novel technology for some of the students, it is possible that the treatment students will simply become more familiar with the technology. To rule out this potential confound, all of the primary outcome measures and all but one of the secondary outcome measures are paper-and-pencil-based.[20]

Differential Attendance. If treated students are more likely to attend (because they enjoy the activities), this could improve academic performance. We test this hypothesis directly but find no differences in attendance.

[19] We ensured that the prizes were appealing throughout the distribution of performance by offering increasing prizes by place in the score distribution. Students could choose a specific prize among a set designated for their quartile of performance.

[20] SART, which must be electronic to accurately measure reaction times, is computer-based. In addition, we specifically administered the task on a computer with a large keyboard to make it as distinct as possible from the tablet-based interventions.

Test-taking Strategies. The Math treatment may help students intuit better test taking strategies, such as skipping hard questions. First, this skill is not trained in the Games arm, where the games do not permit strategies of these types. However, we address this concern by designing the tests to ensure sufficient time and high completion rates. Over 95% of students reach the final question on all exams (Table 2).[21] In addition, a subset of our tests (e.g. listening) mechanically do not permit students to skip around or move faster through the tests. Finally, results are qualitatively similar if we restrict to attempted questions.

Summary. Although many potential confounds exist, many are ruled out by design features such as paper-and-pencil tests and the breadth of the testing battery. A key prediction that differentiates the remaining potential confounds from an attentional capacity channel is whether we observe improved performance among treated students at the beginning of the test. However, as can be seen both in the decline figures and the first row of Table 3, no such level differences exist, isolating an attentional channel.

8 Conclusion

Cognitive capacity—a key predictor of labor productivity—has traditionally been viewed as a fixed component of human capital. This project reexamines this view by testing whether attentional ability is endogenously shaped through one's socioeconomic environment. We focus on a specific dimension of attention: cognitive endurance, or the ability to sustain focus toward a task. We document that lower-income individuals exhibit worse cognitive endurance than more affluent ones across disparate field settings in both rich and poor countries–worker productivity, voting, and school tests. Through a field experiment with 1,650 low-income Indian primary school students, we increase the time devoted to focused cognitive activity during the school day, using either math or non-academic content. Each of these interventions improves cognitive endurance across a variety of unrelated domains—–academic performance, listening retention, and IQ, as well as on traditional attentional ability measures—–indicating that our interventions affected an underlying core resource. These findings suggest that worse schooling environments may disadvantage the poor by hampering the development of cognitive capacity.

[21] Test monitors were also instructed to look for such "skipping" behavior, but it was only very rarely noted given the young age of the students.

References

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

Angeletos, George-Marios and Karthik Sastry, "Inattentive Economies," *NBER Working paper 26413*, 2019.

Augenblick, Ned and Scott Nicholson, "Ballot position, choice fatigue, and voter behaviour," *The Review of Economic Studies*, 2015, *83* (2), 460–480.

Balart, Pau, Matthijs Oosterveen, and Dinand Webbink, "Test scores, noncognitive skills and economic growth," *Economics of Education Review*, April 2018, *63*, 134–153.

Banerjee, Abhijit V and Sendhil Mullainathan, "Limited Attention and Income Distribution," *American Economic Review*, April 2008, *98* (2), 489–493.

Bank, World, "World Development Report," 2004.

Bessone, Pedro, Gautam Rao, Frank Schilbach, Heather Schofield, and Mattie Toma, "The Economic Consequences of Increasing Sleep Among the Urban Poor," *mimeo*, 2019.

Brachet, Tanguy, Guy David, and Andrea M Drechsler, "The effect of shift structure on performance," *American Economic Journal: Applied Economics*, 2012, *4* (2), 219–46.

Chetty, Raj, Adam Looney, and Kory Kroft, "Salience and taxation: Theory and evidence," *American economic review*, 2009, *99* (4), 1145–77.

_ , John N Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan, "How does your kindergarten classroom affect your earnings? Evidence from Project STAR," *The Quarterly journal of economics*, 2011, *126* (4), 1593–1660.

Chun, Marvin M., Julie D. Golomb, and Nicholas B. Turk-Browne, "A Taxonomy of External and Internal Attention," *Annual Review of Psychology*, January 2011, *62* (1), 73–101.

Clearfield, Melissa W and Kelly E Jedd, "The effects of socio-economic status on infant attention," *Infant and Child Development*, 2013, *22* (1), 53–67.

Danziger, Shai, Jonathan Levav, and Liora Avnaim-Pesso, "Extraneous factors in judicial decisions," *Proceedings of the National Academy of Sciences*, 2011, *108* (17), 6889–6892.

Diamond, Adele, "Executive Functions," *Annual Review of Psychology*, January 2013, *64* (1), 135–168.

Duckworth, Angela and Angela Duckworth, *Grit: The power of passion and perseverance*, Vol. 234, Scribner New York, NY, 2016.

Farah, Martha J., David M. Shera, Jessica H. Savage, Laura Betancourt, Joan M. Giannetta, Nancy L. Brodsky, Elsa K. Malmud, and Hallam Hurt, "Childhood poverty: Specific associations with neurocognitive development," *Brain Research*, September 2006, *1110* (1), 166–174.

for the Evaluation of Educational Achievement (IEA), International Association, "TIMSS 2011 Assessment," Technical Report, TIMSS PIRLS International Study Center, Lynch School of Education, Boston College, Cambridge, MA 2013.

Gabaix, Xavier, "Behavioral Inattention," in B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., *Handbook of Behavioral Economics: Foundations and Applications*, Vol. 2, Amsterdam: Elsevier/North-Holland, 2019.

_ and David Laibson, "Shrouded attributes, consumer myopia, and information suppression in competitive markets," *The Quarterly Journal of Economics*, 2006, *121* (2), 505–540.

_ , _ , Guillermo Moloche, and Stephen Weinberg, "Costly information acquisition: Experimental analysis of a boundedly rational model," *American Economic Review*, 2006, *96* (4), 1043–1068.

Hackman, Daniel A, Robert Gallop, Gary W Evans, and Martha J Farah, "Socioeconomic status and executive function: Developmental trajectories and mediation," *Developmental science*, 2015, *18* (5), 686–702.

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein, "Learning through noticing: Theory and evidence from a field experiment," *The Quarterly Journal of Economics*, 2014, *129* (3), 1311–1353.

Heckman, James J, Jora Stixrud, and Sergio Urzua, "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior," *Journal of Labor economics*, 2006, *24* (3), 411–482.

Holmes, Joni, Susan E. Gathercole, and Darren L. Dunning, "Adaptive training leads to sustained enhancement of poor working memory in children," *Developmental Science*, July 2009, *12* (4), F9–F15.

Jaeggi, S. M., M. Buschkuehl, J. Jonides, and W. J. Perrig, "Improving fluid intelligence with training on working memory," *Proceedings of the National Academy of Sciences*, May 2008, *105* (19), 6829–6833.

Johnson, Katherine A., Simon P. Kelly, Mark A. Bellgrove, Edwina Barry, Marie Cox, Michael Gill, and Ian H. Robertson, "Response variability in Attention Deficit Hyperactivity Disorder: Evidence for neuropsychological heterogeneity," *Neuropsychologia*, January 2007, *45* (4), 630–638.

Kautz, Tim, James J Heckman, Ron Diris, Bas Ter Weel, and Lex Borghans, "Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success," Technical Report, National Bureau of Economic Research 2014.

Klingberg, Torkel, Mats Johnson, Christopher G Gillberg, and Helena Westerberg, "Computerized Training of Working Memory in Children With ADHD-A Randomized, Controlled Trial," *J. AM. ACAD. CHILD ADOLESC. PSYCHIATRY*, 2005, p. 11.

Lawson, Gwendolyn M, Cayce J Hook, Daniel A Hackman, Martha J Farah, James A Griffin, Lisa S Freund, and Peggy McCardle, "Socioeconomic status and neurocognitive development: Executive function," *Executive Function in Preschool Children: Integrating Measurement, Neurodevelopment, and Translational Research. American Psychological Association*, 2014, pp. 1–28.

Mani, A., S. Mullainathan, E. Shafir, and J. Zhao, "Poverty Impedes Cognitive Function," *Science*, August 2013, *341* (6149), 976–980.

Mullainathan, Sendhil and Eldar Shafir, *Scarcity: Why having too little means so much*, Macmillan, 2013.

Nutley, Sissela Bergman, Stina Söderqvist, Sara Bryde, Lisa B. Thorell, Keith Humphreys, and Torkel Klingberg, "Gains in fluid intelligence after training non-verbal reasoning in 4-year-old children: a controlled, randomized study: Fluid intelligence gains after training non-verbal reasoning," *Developmental Science*, May 2011, *14* (3), 591–601.

Oades, Robert D, "Differential measures of 'sustained attention' in children with attention-deficit/hyperactivity or tic disorders: relations to monoamine metabolism," *Psychiatry Research*, March 2000, *93* (2), 165–178.

Peebles, David and Daniel Bothell, "Modelling Performance in the Sustained Attention to Response Task.," in "ICCM" 2004, pp. 231–236.

Pratham, "Annual Survey of Education (ASER)," 2011.

Raven, John, "The Raven's Progressive Matrices: Change and Stability over Culture and Time," *Cognitive Psychology*, August 2000, *41* (1), 1–48.

Raven, John C., "Raven. Mental Tests Used in Genetic Studies: The Performances of Related Individuals in Tests Mainly Educative and Mainly Reproductive," *Unpublished master's thesis, University of London*, 1936.

Scribner, Sylvia and Michael Cole, "Cognitive consequences of formal and informal education," *Science*, 1973, *182* (4112), 553–559.

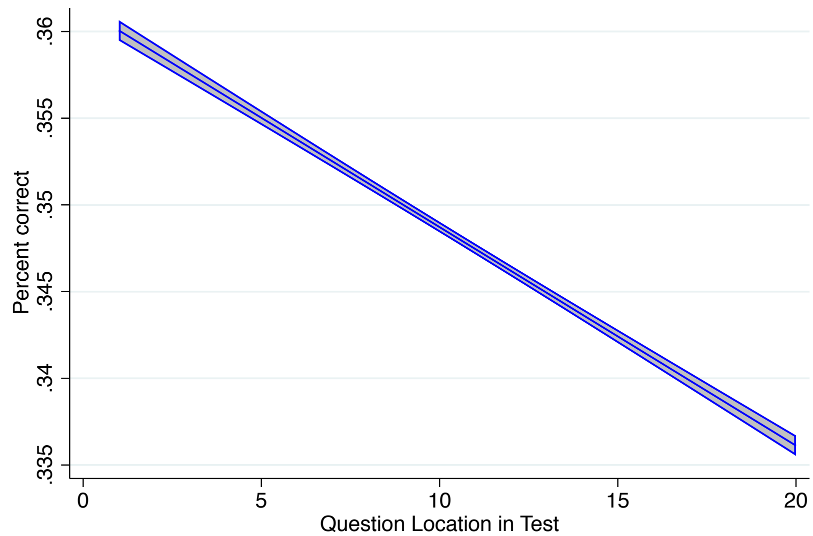
Warm, Joel S, Gerald Matthews, and Victor S Finomore Jr, "Vigilance, workload, and stress," in "Performance under stress," CRC Press, 2018, pp. 131–158.

Zamarro, Gema, Collin Hitt, and Ildefonso Mendez, "When Students Don't Care: Reexamining International Differences in Achievement and Student Effort," *Journal of Human Capital*, 2019, *13* (4), 000–000.

9 Figures

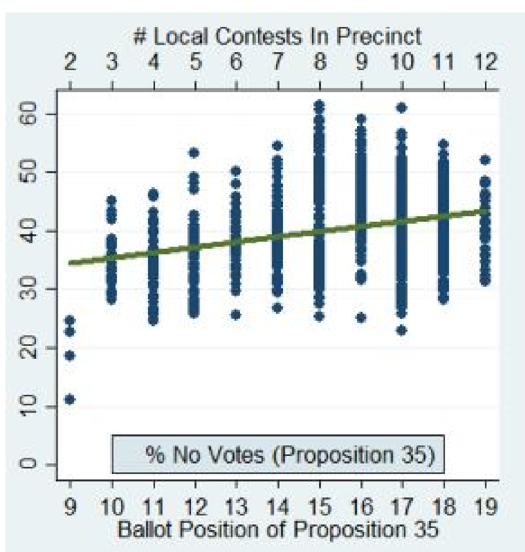
Motivation: Attentional declines are common and important

Figure 1: TIMSS: Declines in performance over time



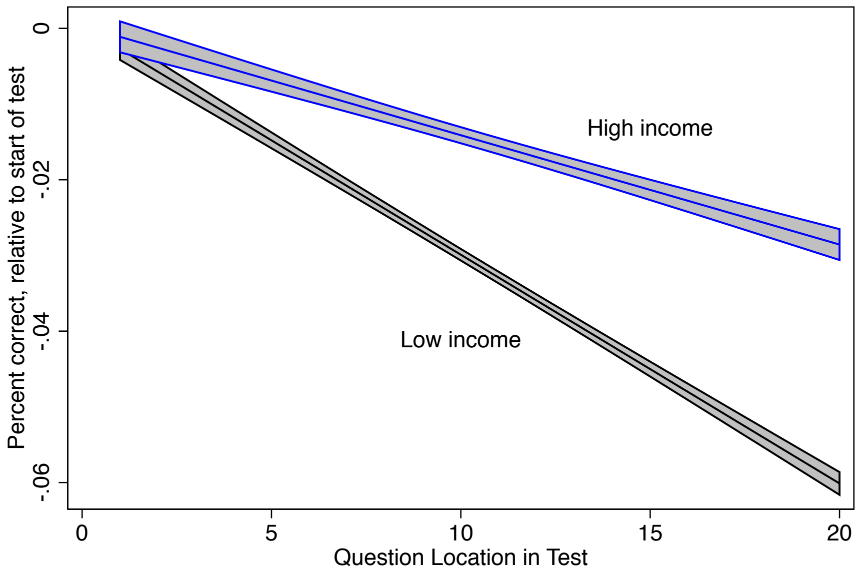
Notes: Authors' calculations. Question order is block randomized.

Figure 2: Voting: Declines in cognitive effort over time



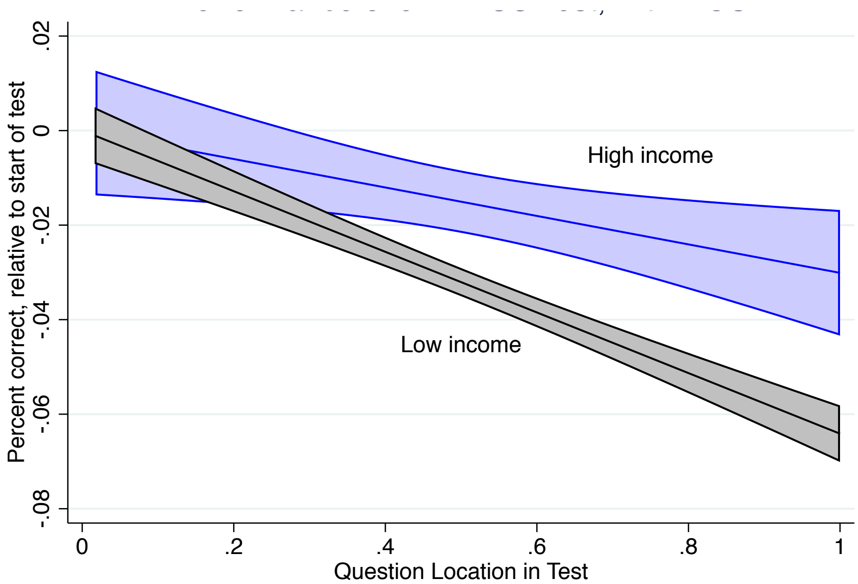
Note: Source is Augenblick and Nicholson 2015. Item order is quasi-random

Figure 3: Heterogeneity in Declines by National GDP per Capita



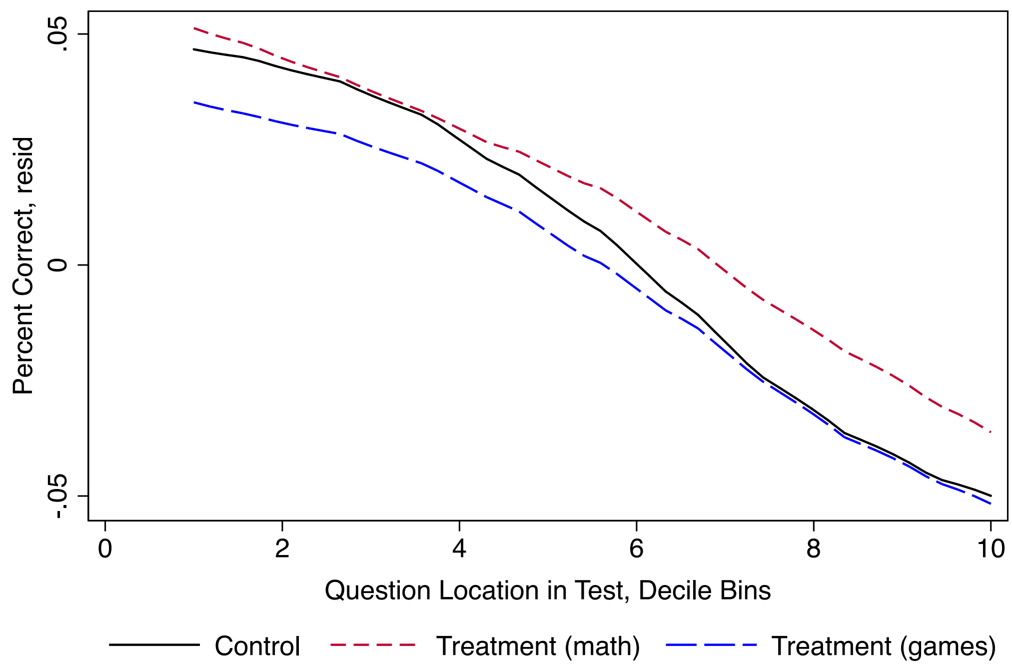
Notes: Authors' calculations. Question order is block randomized.

Figure 4: Heterogeneity in Declines by SES within the United States



Notes: Authors' calculations. Question order is block randomized.

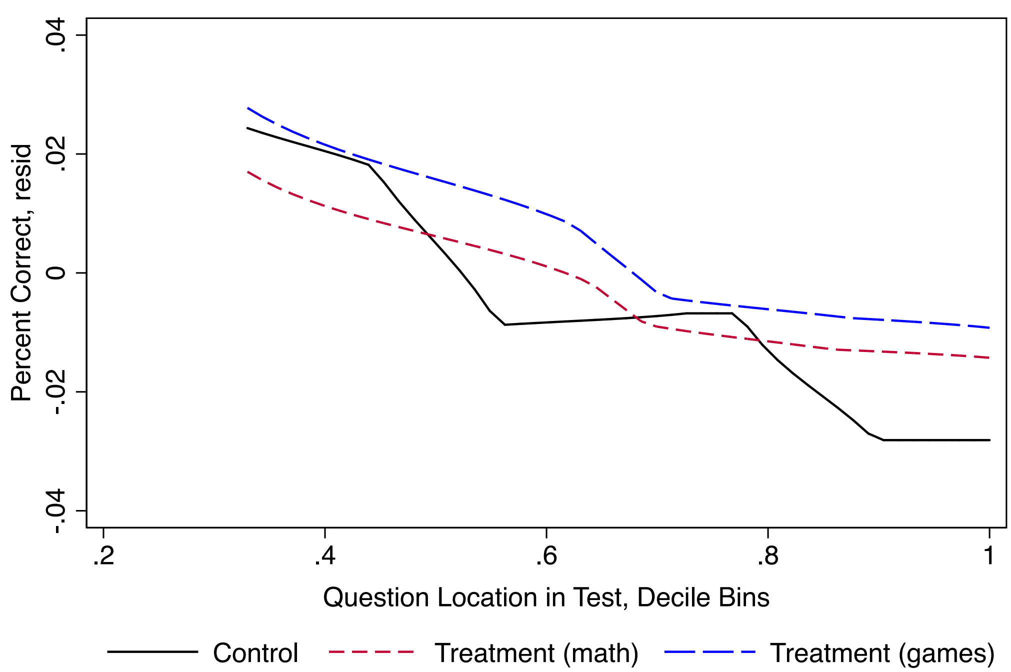
Figure 5: Training slows decline on Math test



Note: Control and Math Treatment both receive math practice but Games arm does not → expect level difference for Games Treatment

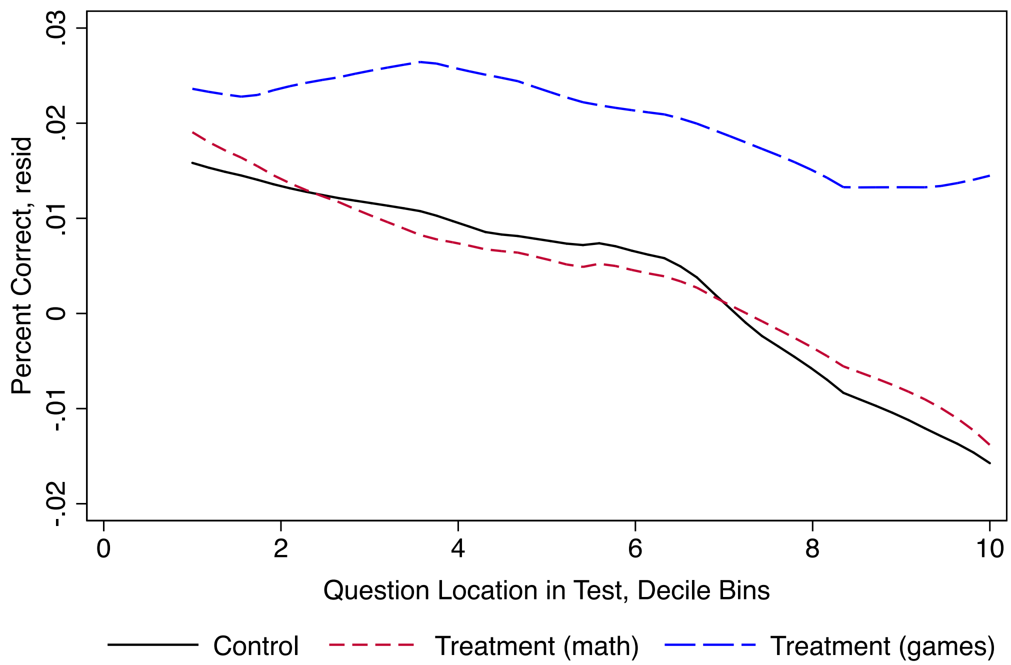
Treated students exhibit **22% less decline** in final decile.

Figure 6: Training improves Listening test performance



Treated students exhibit **31% less decline** in final decile.

Figure 7: Training improves Ravens Matrices (IQ) performance



Treated students exhibit **62% less decline** in final decile.

10 Tables

Table 1: Attrition

	(1) Attrition
Treatment (math)	0.000125
	(0.0118)
Treatment (games)	0.00337
	(0.0113)
R^2	0.000
Dependent variable mean	0.0500
Number of students	1720
Number of student-years	2365

Notes: Outcome is whether we observe at least one (non baseline) test each year.
Standard errors in parentheses.
** p<0.10, ** p<0.05, *** p<0.01.*

Table 2: Test completion

Statistic	Math	Listening	Ravens
% attempted	80	98	96
% skipped	13	1	2
% of students completing last question	71	98	89
Avg. last question % completed	95	99	97

Table 3: Math, Listening, and Raven's Test Declines

	All tests (pooled)	Math	Listening	Ravens
Treated	-0.00270	-0.00204	-0.00540	0.00124
	(0.00881)	(0.0120)	(0.0112)	(0.0193)
Treated*Decile 2 to 5	0.00785	0.0110	0.0160	0.0117
	(0.00574)	(0.00972)	(0.00986)	(0.0184)
Treated*Decile 6 to 10	0.0207***	0.0214**	0.0234**	0.0122
	(0.00670)	(0.0109)	(0.01000)	(0.0185)
R^2	0.285	0.269	0.255	0.228
Dependent variable mean	0.503	0.585	0.208	0.207
Number of students	1650	1629	1622	1633
Number of items	330949	194384	68847	67718

*Notes: Clustered standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.*

Table 4: Psychology Literature Sustained Attention Tests

	(1) Index of both	(2) SART	(3) Symbol matching
Treated	0.0448*	0.0613*	0.0372
	(0.0241)	(0.0315)	(0.0316)
R^2	0.123	0.169	0.143
Dependent variable mean	-7.35e-10	-7.35e-10	-7.35e-10
Number of students	1636	1380	1628
Number of items	9704	3897	5807

*Notes: Outcomes are measured as (True positive z-score - False positive z-score). Clustered standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.*

Table 5: Observation from Standard Lecture Class (z-score)

	(1) Index of all 3	(2) Follow instructions	(3) Response to stimuli	(4) Physical signs
Treated	0.120**	0.0977*	0.137**	-0.0267
	(0.0540)	(0.0579)	(0.0617)	(0.0345)
R^2	0.081	0.153	0.071	0.136
Dependent variable mean	0.0174	1.39e-08	-2.29e-08	0.0186
Number of students	1203	1206	1205	1204

*Notes: Clustered standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.*