Attention as Human Capital[*]

Christina Brown[†] Supreet Kaur Geeta Kingdon Heather Schofield

February 1, 2021

(Preliminary and Incomplete Draft - Do Not Cite or Circulate)

Abstract

Schooling may shape cognitive performance not only by teaching academic content, but by altering individuals' underlying capacity for attention. We first document a novel fact: across various domains, the poor exhibit worse sustained attention than the rich—as measured by performance declines over time during a task—including in academic tests, worker productivity, and voting. These stark differences suggest that attentional capacity may be malleable, endogenously shaped through one's environment. Consistent with this hypothesis, we document that the schooling environments of the poor systematically afford less time to practice sustained attention. We test this hypothesis using a field experiment with 1,650 low-income Indian primary school students. We "train" sustained attention through increased time in independent focused activity within the school day, using either math content (mimicking good schooling) or non-academic content (providing a pure test of our mechanism). Each approach leads to substantive changes in attentional capacity across broad domains unrelated to the treatment content: academic subjects, listening, IQ tests, and traditional psychology measures of sustained attention. These gains persist three months after the intervention ends, and result in meaningful improvements in school performance of about 0.1 standard deviations. Our findings indicate a broader view of how schooling shapes human capital, and suggest that worse environments may disadvantage poor children by hampering the development of underlying cognitive capacity.

[*]We thank David Autor, Stefano DellaVigna, Ernst Fehr, Xavier Gine, Lawrence Katz, Patrick Kline, David Laibson, and Imran Rasul for helpful comments and discussions. We gratefully acknowledge generous funding and support from USAID DIV, the Global Engagement Fund at the University of Pennsylvania, and The Weiss Family Program Fund for Research in Development Economics. For operational support, we thank the Institute for Financial Management and Research (IFMR) and Pixatel. Rolly Kapoor, Lubna Anantakrishnan, Simranjeet Dhir, and Deepika Ghosh provided excellent research assistance. All remaining errors are our own. We received IRB approval from the University of California, Berkeley and 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 (supreet@berkeley.edu); Kingdon: University College London (g.kingdon@ucl.ac.uk); Schofield: University of Pennsylvania (hschof@wharton.upenn.edu)

1 Introduction

Studies consistently document that the poor fare worse in cognitive activities than the rich—in academic performance, worker productivity, IQ tests, decision-making and executive-functioning (Broer et al., 2019; Hopfenbeck et al., 2018; Hackman and Farah, 2009), and that such differences arise at a young age. One interpretation of these patterns is that some feature of poverty leads to lower human capital levels. Consistent with this, schooling interventions with low-income children tend to show both immediate and long-term benefits for a broad set of outcomes—class grades and earnings, but also disciplinary behavior, health, and IQ scores (Knudsen et al., 2006; Krueger and Whitmore, 2001; Chetty et al., 2011; Heckman et al., 2013; Baird et al., 2016; Duflo, 2001).

Why (improved) schooling should generate such broad-based benefits has not been fully resolved (e.g. Acemoglu et al., 2012). The economics literature has largely focused on two potential channels. First, schooling directly improves hard skills, e.g. math and reading, which can explain some, but perhaps not all, of the above outcomes. For example, studies indicate that the persistent effects on earnings or health do not appear to be mediated solely through learning academic content (e.g. Chetty et al., 2011). In response, a complementary literature posits that schooling may improve non-cognitive skills, such as socioemotional regulation or self-control (e.g. Heckman et al., 2013; Jackson, 2018). A more recent, emergent body of work posits a third potential pathway: schooling may shape underlying cognitive capacities themselves (e.g. Berger et al., 2020; Jaeggi et al., 2008).

We argue for a broadening of our view of how schooling shapes human capital, in line with this third pathway. We examine a novel potential channel: schooling may expand underlying attentional capacity. We focus on a particular dimension of attention: sustained attention, or the ability to direct focus toward a task over time (e.g., Chun et al. (2011)). If schooling expands the mind's capacity to focus, this would have relevance for the various outcomes described above—complementing the two channels discussed in the literature, and potentially even bolstering them.[1] This hypothesis rests on the premise that attentional capacity is malleable, and that schooling can alter it. We test these ideas using a randomized field experiment with low income primary school students in India.

Before turning to our experiment, to motivate the empirical relevance of these ideas, we first look for markers of sustained attention in field behavior. Psychologists traditionally construct measures from laboratory tests, but these measures are not always portable and typically unavailable in existing data. This relates more generally to the difficulty in measuring attention in field behavior (Gabaix, 2019). We examine a potential proxy that can be widely constructed from available data. Specifically, if individuals lose focus over time, then performance should decline over time when engaged in any task that requires cognitive resources. We hypothesize that if sustained attention truly matters, then we should see such cognitive fatigue manifest itself in a variety of field activities.

We look for this suggestive proxy in three distinct domains. First, we examine outcomes for the

[1] For example, research in psychology argues that sustained attention not only increases performance on cognitive activities, but is also an input into exerting self-control (Chun et al., 2011).

TIMSS test, a standardized 36-minute math exam given to elementary and middle school children in the U.S. and other countries. The average student is about 12% less likely to answer a question correctly if it occurs at the end of the test rather than the beginning. While this pattern is observable in many kinds of tests, observing it in TIMSS is particularly suggestive. This is because question order is randomized and students are given enough time to finish—ruling out that declines are driven by questions becoming harder or being skipped toward the end of the test. Second, we examine the hourly performance of full-time data entry workers, using data from Kaur et al. (2015). Workers are paid a piece rate per accurate field entered; when a worker makes an accuracy mistake, this means he has exerted effort a field but earns no money for it. Consistent with attentional fatigue, we document a marked decline in entry accuracy over the workday, with error rates increasing by 12% from 10 am to 4 pm. Third, we examine the domain of decision-making, using voting in elections. We replicate Augenblick and Nicholson (2015), who show that ballot order predicts whether individuals make an active choice or pick the default candidate—with magnitudes large enough to change whether 6% of the propositions in their data become law. More broadly, such performance declines have been documented by researchers in many other domains as well, such as fatigue in paramedic shifts, rulings by judges, and other professions such as sentries, air traffic controllers, and anesthetists (Brachet et al., 2012; Danziger et al., 2011; Warm and Dember, 1998). While the patterns in each particular example may have multiple interpretations, the consistency of this finding across domains motivates our view that sustained attention is an important dimension of cognition. We use our randomized field experiment, below, to elucidate the mechanism behind such declines more cleanly.

Next, building on this approach, we document a new empirical finding: the poor exhibit worse attentional declines than the rich. In each of the three examples above, we find marked differences in decline rates for individuals with higher vs. lower socioeconomic status (SES).[2] Within the TIMSS math achievement test, this is true when comparing high SES vs. low SES 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 black and white students in the US. Among data entry workers, lower SES workers have a 21% decline rate in data entry accuracy—twice that of higher SES workers. The difference in declines accounts for 10% of the productivity gap between these two groups. Finally, compared to voters in richer neighborhoods, poorer voters exhibit 50% greater decline in active choice when they move down-ballot. These findings suggest that differences in sustained attention may help explain overall differences in cognition and performance between the poor and rich in a variety of domains.

Why should the poor exhibit greater performance declines than the rich across different tasks?

[2]In each case, we use the proxy for socioeconomic status available in the dataset we have: in TIMSS, white vs. non-white individuals in the US, and GDP per capita across countries; in Kaur et al. (2015), high school vs. no high school; and in Augenblick and Nicholson (2015) above vs. below median fraction white population in each polling precinct. It would be interesting to examine SES heterogeneity in other examples as well. The three examples we use were driven by data availability (requiring data across time, and proxies for SES) and design (requiring no changes in average task difficulty over time, and no attrition in completion over time.

The literature has discussed various potential channels that disadvantage the poor: worse academic training, physiological factors like sleep or nutrition, motivational differences, or mental worries that distract attention (Bessone et al., 2019; Schofield, 2015; Gneezy et al., 2019; Mani et al., 2013; Kaur et al., 2021). While research on these channels has largely focused on average differences in performance, it is of course possible that they may also mediate declines—for example, through some interaction with how costly it is to exert focus.

We instead posit that differential declines may reflect differences in the capacity for sustained attention itself. If attentional capacity is malleable, then the ability to maintain focus will be shaped by one's cognitive training and background. This naturally points to the potential role of schooling in shaping attentional capacity.

This view is supported by the school time-use data in the TIMSS dataset. Consistent with our hypothesis, schools attended by richer students employ pedagogies in which students spend more time on focused independent practice. Moreover, in schools with such pedagogy, students exhibit less attentional declines when engaged in cognitive tasks—even after controlling for wealth. While only a correlation, this suggests that practicing exerting focus in school may help expand underlying attentional capacity. This rests on the premise that attentional capacity is indeed malleable, an idea for which there has been scant evidence in the psychology or economics literatures to date.

We test whether attentional capacity can be expanded using a field experiment with 1,650 low-income Indian primary school students. The schools in our sample exhibit many features common to low-income schooling environments. Schools are noisy, with many distractions and high student to teacher ratios. Students rarely undertake much individual practice, and homework is either not regularly assigned or not completed. Overall, it is rare for students to be required to sit and undertake a specific focused activity for 10-20 minutes at a time in silence, aside from exams. Consequently, students have very little opportunity to practice exerting attention in a sustained manner.

We "train" sustained attention through increased time in focused activity within the school day using two different approaches. In the first treatment (Math), students practice mathematics problems over 25 minute sessions using an individualized adaptive tablet-based math platform.[3] This treatment mimics what good schooling does: providing cognitively challenging focused activity within the context of academic learning. However, under our hypothesis, practicing any cognitively challenging task should improve sustained attention—regardless of whether students learn anything from it. Consequently, in our second treatment arm, students play cognitively demanding games over 25 minute sessions. This enables us to increase time spent on focused activity free of any academic content, providing a cleaner test of our mechanism. 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), resulting in little engagement or sustained practice.[4] For each of the two treatment groups, the study hall period is replaced with cognitive practice periods

[3] We use imagineMath, developed by Pixatel.

[4] The questions written on the chalkboard are drawn from the same database used in the Math treatment arm.

1-3 times per week. In total, each treatment resulted in 6-15 hours of focused practice over the course of a 6 month period during the academic year, with differences in hours due to different starting times across different schools. Note that while the interventions are delivered via tablet, this is not consequential for testing our hypothesis—it was simply a convenient implementation approach.[5] Moreover, our core outcome measures are based on regular paper and pencil tests.

To construct outcomes, we start with the premise that sustained attention is relevant for any cognitive activity over time, and its effects should therefore manifest broadly. We consequently design outcomes to test for "far transfer" effects: impacts on attentional capacity in domains that are completely unrelated to the content that was practiced in the treatment arms. We examine other academic subjects such as Hindi and English, domains such as listening and IQ tests, and also examine traditional laboratory tests of sustained attention. Finding such broad-based effects would support the view that schooling can shape core underlying cognitive capacity.

We first examine whether simply practicing focusing can raise overall academic performance in students' regular school exams, administered outside of the experiment. Each of the two interventions improves exam performance across the three core academic subjects taken by all students—English, Hindi, and Math—by 0.083-0.096 standard deviations. Using variation in treatment intensity across schools, our results suggest that each additional hour of cognitive practice increased students' administrative exam scores by 0.012 standard deviations.[6] These effects are present for each of the three test subjects. They are also similar for each of the Math and Games treatment arms, and we cannot reject that both treatments had the same impact on outcomes.[7]

This effect size of 0.08-0.10 SD is large, especially when compared to much more substantive structural interventions—for example, substantially reducing class sizes in the US (0.1 SD), tracking students by ability in Kenya (0.14 SD) or remedial education with an additional teacher in India (0.14 SD) (Krueger and Whitmore, 2001; Duflo et al., 2011; Banerjee et al., 2007). These prior studies provided continuous exposure to the intervention throughout the entire year, and specifically targeted academic learning in the subjects tested. In contrast, for example, neither of our treatment arms provided any exposure to Hindi, and the Games arm—whose overall treatment effects are as big or bigger than the Math arm—arguably provided no academic training across subjects.

These results alone do not enable us to disentangle mechanisms in two respects. First, there is a potential concern that while we designed our interventions to increase practice focusing, they may have also affected some other channel such as motivation or confidence. Second, if these effects do reflect improved attentional focus, they may arise both because students learned more (e.g., were

[5] Importantly, the exact design of interventions to increase time spent in sustained focus are likely to vary across contexts. While these interventions are appropriate to this context, we do not take a stance on whether they are likely to apply broadly (e.g. to high income settings).

[6] While all students within a school (and therefore across treatment arms) received the same exact number of class sessions, treatment hours was not explicitly randomly assigned across schools, and could therefore be correlated with other features of schools. We therefore view the analysis using treatment intensity as suggestive.

[7] While the coefficient on the Math treatment arm is slightly larger than the Games arm for the math test, we cannot reject that they are the same.

better at paying attention to the teacher in class), and because they were better at taking the end of term tests (i.e., had less attentional decline during the tests)—both of which are important and policy relevant channels. To elucidate mechanisms, we design additional tests that enable us to cleanly identify effects on performance declines—not because we view this as the more important of the channels, but because it offers a targeted test of sustained attention as a mechanism. We also use this to help rule out confounds for our main treatment effects, such as changes in motivation. We supplement this with traditional measures of sustained attention to provide further positive evidence for our proposed channel. Overall, we undertake two additional sets of tests.

First, we look for positive evidence for changes in attentional capacity using traditional measures from the psychology literature: the Sustained Attention to Response Task (SART), which measures response time to notice and respond to the appearance of a stimulus, and a symbol matching task. Each of the two treatment arms improves performance on these measures by 0.08 standard deviations.

In addition, we examine effects on classroom behavior. We have observers, who are blind to treatment status, record measures of classroom behaviors adapted from the Vanderbilt ADHD diagnostic teacher rating scale. This includes, for example, rating attentiveness to the teacher's lecture and response to stimuli. Treatment improves performance on this index by 0.17 standard deviations.

Second, we design and administer additional tests in three domains—listening, Raven's Matrices, and mathematics—to enable us to examine performance declines over time. We randomize question order on each test, so that the same question appears early in the test for some students and late in the test for others. By including question fixed effects in our analyses, we can cleanly identify whether our treatments mitigate declines in performance over the course of the test—our marker for sustained attention in field settings.[8] We administered these tests during the school day, with ample time so that declines are not confounded by non-response. From students' perspective, these were required school tests, providing natural stakes.

We also use this data to distinguish improvements in attentional capacity from other confounding channels, such as motivation or confidence. To test for this, we examine differences in performance across treatment and control groups for the questions in the beginning of each test—before cognitive fatigue has set in.[9] If the treatments improved confidence or motivation, or alternate cognitive channels such as memory, we would expect level effects even in the beginning of the test. In contrast, if only sustained attention has improved, then treatment effects should emerge later on, once cognitive fatigue has set in. In addition, we supplement this with additional direct tests for motivation and grit (see below).

We first document that in each of our three tests, among control group students, we see significant declines in the probability of getting the question correct over time within each test. This matches

[8]These tests are all administered using paper and pencil. In addition, students are provided ample time to finish each test, so that declines cannot be driven by unanswered questions (we also directly verify that students complete the tests in the data).

[9]For this, we focus on the listening and ravens tests. Level effects for the math test are difficult to interpret, since both Math arm and control group received practice in math problems.

the patterns in our initial motivational evidence above, and creates scope for us to find treatment mitigating these declines. Consistent with our prediction, in each test, the treatment group shows less decline over time than the control group. In addition, the effects are meaningful in magnitude—ranging from 13% to 25% of the total decline experienced by the Control group. If these results are applied to data from an international achievement test, it would cut the gap in performance decline between high and low-income countries by 40%. As was the case above, each of the two treatment arms significantly mitigates declines, providing further reassurance for the robustness of our results.

Further consistent with sustained attention as the mechanism, we find no evidence for treatment effects at the start of each of these tests (e.g. the first decile of questions).[10] Rather, treatment effects only emerge later on, and grow over time. This indicates that effects are not simply coming from increases in confidence or motivation to do well.[11] Rather, they made students more able to maintain cognitive focus in carrying out these tasks. In addition, we use supplementary tests to rule out additional competing explanations. For example, we introduce incentives for a random subset of students to do well on the tests, and find that this has no differential impact on performance levels or treatment effects on declines. This further helps rule out motivation as a channel, and is consistent with the idea that in our specific cultural context and experimental setting, students take the tests seriously. Similarly, we find no treatment effects on whether performance drops after a particularly hard problem, helping rule out changes in grit as a channel.

Together, our results support the view that the sizable impacts we observe on students' school performance are at least partly driven by improvements in sustained attention. The broad transfer we see, and the multiple pathways through which it can affect learning, indicate that even small improvements in declines can aggregate up to meaningful overall impacts. Our intervention serves as a proof of concept—indicating that schooling itself can shape attentional capacity. This provides causal evidence that is consistent with our motivational facts above.

To test for persistence in effects, we conducted a three-month follow-up, in which we again administered tests in math, listening, and Ravens Matrices. We find that treated students continue to show less decline over time during these tests, and cannot reject that the change in treatment effects is zero. This provides suggestive evidence for persistence, though of course does not speak to persistence over longer horizons.[12] However, note that to the extent that the occupations of richer and more well-educated individuals continue to reinforce focused cognitive practice—e.g. a white collar job of a computer programming versus working loading boxes on a factory floor—effects on attention capacity may continue to be reinforced differentially even after individuals leave school.

[10]The only exception is that in the beginning of the math test, the students in the Games arm appear to do worse than those in the Math and Control arms (statistically insignificant). This pattern is what one would expect since the Math and Control arms explicitly practiced math problems but the Games arm did not. However, we cannot reject that the treatment effect on *declines* is the same across both the Math and Games arms for the math test.

[11]On average students get questions in the first decile correct about 52% of the time; consequently, there is large scope for increases in performance at the start of the test if students are more motivated or confident.

[12]Our ability to collect data for further follow-up was halted by the Covid pandemic, which led schools to stop operating and led to the shut-down of some of the schools in our sample.

This paper makes several 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 in childhood.[13] These changes are not limited to the domain that is directly trained. Rather we find "far transfer" in effects: impacts in a broad set of domains unrelated to the content practiced in the treatment arms, indicating broad-based and generalizeable impacts. This constitutes, to our knowledge, the first evidence of far transfer effects in attentional training in any setting. Moreover, our methodological approach of detecting sustained attention using performance declines can be applied broadly, adding a new measurement tool for use by psychologists and economists.

Second, relatedly, we advance rapidly growing work in behavioral economics, which examines the implications of limited attention for economic behavior (e.g. Gabaix, 2019). This work takes as given that individuals have finite attentional capacity, and examines the subsequent implications for decision-making and behavior—for example, if individuals fail to attend to certain dimensions of the environment (e.g. Chetty et al., 2009; Hanna et al., 2014; Gagnon-Bartsch et al., 2018), or how involuntary allocation of attention can affect outcomes for the poor (e.g. Banerjee and Mullainathan, 2008; Mullainathan and Shafir, 2013; Mani et al., 2013; Kaur et al., 2021). We expand on this work by documenting that this capacity itself can be expanded, with sizable implications for field outcomes. Further, our findings suggest that these considerations may be especially relevant for the poor.

Third, we extend our understanding of the role of schooling in human capital accumulation (Acemoglu et al., 2012). 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 (Association For The Evaluation Of Educational Achievement, 2013).

More speculatively, our findings speak to the 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). Our research suggests that growing up in poverty may limit cognitive development through low-quality schooling. If individuals do not have the opportunity to practice exerting sustained cognitive focus, this capacity may be under-developed and generate broad negative consequences for cognitive functioning, educational attainment, labor productivity, and economic life. Because attentional capacity is an input for so much of human activity, even small differences in this skill could aggregate up to have meaningful impacts. Consequently, our

[13] Laboratory studies document that some executive functions, such as fluid intelligence, are malleable (Klingberg et al., 2005; Jaeggi et al., 2008). However, they typically find little evidence of effects transfer beyond the specific domain that is directly trained (Bergman Nutley et al., 2011; Holmes et al., 2009; Klingberg et al., 2005; Diamond, 2013)—perhaps because sample sizes are typically quite small (e.g., 15-40 individuals per arm), limiting statistical power. A notable exception is Berger et al. (2020), who find far transfer effects in training working memory.

findings suggest an additional potential micro-foundation for the persistence of achievement gaps between the rich and the poor. These gaps may create feedback loops (e.g. via jobs that do not reinforce cognitive focus, or enrollment of children in lower quality schools) that could contribute to the intergenerational transmission of poverty. While only suggestive, our findings provide impetus for further research testing for the role of cognitive training in schooling, and its potential impacts on cognition and productivity.

2 Declines in Attention

2.1 Attentional Declines are Common Across Many 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 relevant to a wide variety of settings and professions. As motivational evidence, we document cognitive declines across a range of settings below. This empirical regularity suggests the broad applicability of this element of cognition. We then examine heterogeneity in this skill by SES and demonstrate a consistent pattern of greater declines among low-SES populations across multiple settings.

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.[14]

As shown in Figure 1, performance declines across the test are substantial. Within the US, a student is 6% (3 percentage points) less likely to get a question right at the end of the test than at the beginning. At 12% (over 6 percentage points), these declines are even more substantial in the international data (see Figure 2).

Worker Productivity. Virtually any work task requires the direction of attention, which can diminish over the course of the workday. Using data from Kaur et al. (2015), we examine hourly

[14] 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.

performance of full-time data entry workers over nine months. Workers' earnings were comprised of a piece rate for each accurate field entered. A data entry error is therefore costly: the worker expends time and effort to transcribe a word, but receives no payments. Figure 3 documents a marked decline in accuracy rates over the course of the day. On average, error rates increase roughly 12% between 10am and 4pm. To make up for the loss in productivity from the attentional decline, worker piece rates would need to increase by 2.4% by the end of the workday to achieve the same output as in the beginning of the day.[15]

Attentional declines appear prevalent and consequential in a wide variety of other professions 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. In each of these domains, error rates and attentional lapses increase over time with economically meaningful impacts on productivity and outcomes with long-run consequences such as criminal conviction and mortality.

Voting Behavior. The effects of attentional capacity 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 4). 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.

2.2 Attentional Declines are Correlated with Socio-economic Status

Many elements of cognition are known to to be correlated with 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). We find a similar correlation between socio-economic status and the ability to sustain focus. The differences occur in important domains and are often substantial. The three examples below were driven by data availability (requiring data across time, and proxies for SES) and design (requiring no changes in average task difficulty over time, and no attrition in completion over time. It would be interesting to examine

[15] Analysis conducted using 10 am - 4 pm to avoid compositional effects of workers arriving and departing. Piece rate calculations are based on the effort elasticity of 0.33 estimated in Kaur et al. (2015).

SES heterogeneity in other examples as well as additional data is available.[16]

Academic Test Performance. 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 6). Similar differences are found among high-income and low-income students within the United States (Figure 5). 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.

Worker Productivity. Similarly, when comparing more educated workers to less educated workers engaged in data entry (i.e. those with above or below high school education), there is clear evidence of a differential pattern of increase in production errors over the course of the day. [17] As shown in Figure 7, less educated workers experience an estimated increase in their error rate of 1.3 percentage points, or 21% from the beginning to the end of the workday. Their decline in accuracy that is twice as large as that of more educated workers by 4 pm. Overall, the magnitude of the decline among less educated workers amounts to 10% of the productivity gap between higher and lower educated workers in the sample.

Voting Behavior. Finally, voting data follows a similar pattern of declines (Figure 8). Using racial composition of a neighborhood as a proxy for socio-economic status, we find that individuals in less affluent neighborhoods are increasingly more likely to rely on the default option when items are further down-ballot. These differences are substantial. In Proposition 35 – the initiative provided as an example in the original paper – the decline associated with a 10 item shift is position is roughly 50% greater in less affluent neighborhoods.

3 Experimental Design

The broad pattern of attentional declines across many domains and differences in the rate of decline by socio-economic status motivate our randomized controlled trial. The RCT explores whether low-income students in India — who typically receive very little practice in sustaining attention over time — can improve their ability to sustain focus in a broadly generalizable fashion using a simple school-based intervention.

[16] Further, building such tests into future work should be relatively easy. In either lab or field studies with reasonable variation in socio-economic status, simply randomizing task order would be sufficient to allow such comparisons.

[17] In (Kaur et al., 2015), the only available SES measure is education. Educational levels are divided using a median split, which corresponds to whether the worker has above a 12th grade education.

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.[18]

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 academic work outside of school, leading to little time spent on homework. 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 a cognitively challenging task.

Math Treatment. The first intervention mimics the way in which students traditionally practice exerting directed attention over sustained periods in good 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).[19]

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.

[18] We also included income tercile in constructed strata in the subset of schools where parental income was available.

[19] 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.

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 non-instructional cognitive games for sustained periods, without any academic learning. For example, students play games of attentional focus such as tangrams and N-back.[20] 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 study hall period (i.e. 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.

Treatment Intensity. These activities were scheduled to take place two to three times per week over six months, with pauses for weeks with extra holidays and festivals, and during regular school testing periods. Overall, the average number of sessions undertaken in an academic year was 28. The mean number of hours in the year spent in class sessions was 9.2, and ranged from 4.5 to 15 hours across schools depending on when we started the intervention in the schools and the number of breaks in implementation.[21]

Implementation. We conducted these activities during the school day, in 30 minute periods. These periods replaced study halls or elective classes such as an art period 2-3 times per week, based on an ex ante schedule created in conjunction with each school. During that period, students in a given class section were split into one of three classrooms, based on their respective experimental arm,

[20] 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.

[21] 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.

where they undertook their assigned activity. Periods were usually about 30 minutes, leading to about 20-25 minutes of time when treatment students could actually engage in practice once they were settled. This 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. Consequently, in measuring outcomes, we take a broad approach and study changes across a variety of topical domains and through a variety of testing strategies. We begin by examining effects on students' regular school administered tests to measure impacts on educational outcomes. Next, to test for impacts on sustained attention as an underlying mechanism, we use three approaches. First, we measure impacts on traditional psychology measures of sustained attention. Second, we examine treatment-blind observer measures of students' classroom behavior, adapted from an ADHD teacher rating scale. Third, and most substantively, we utilize the decline approach developed in Section 2 to test whether treated students exhibit smaller performance declines over time. This third set of tests also builds in ancillary features that enable us to rule out confounds such as motivation or confidence. Throughout all these measures, we test students in domains that were unrelated to the content they practiced as part of the treatment arms—enabling us to draw conclusions about whether our results capture a change in core cognitive capacity.

4.2 School Administered Exams

To measure overall improvements in school performance, we first examine students' test scores on the end of term exams administered by schools—the standard field measure for educational impacts. We look at the three core subjects taught and tested by all schools in our sample: Math, Hindi, and English. Note that seeing effects across these subjects would in itself be indicative of broad "transfer" effects, consistent with the idea that our treatment affected an underlying cognitive resource. This is because, for example, neither of our treatment arms provided any exposure to Hindi, and the Games arm arguably provided no academic training across subjects.

We complement this measure with additional outcomes that are designed to provide positive tests for sustained attention changes specifically, and to help rule out alternate mechanisms for our effects.

4.3 Measures of Attention from Psychology

Psychologists measure sustained attention using laboratory measures that capture whether individuals can sustain focus. We used two such measures to test for positive evidence of attentional improvements. We follow the convention in the psychology literature and measure mean performance on each task, as defined below.

(1) **Symbol matching.** Students are given a paper-based workbook, each page of which contains a grid of randomly ordered pictorial symbols. A specific set of 2-3 target symbols is displayed at the top of the sheet above the grid. Students are asked to go through the grid, crossing out any of the target symbols they encounter. Scores 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.

4.4 Classroom Behaviors

Students' behavior was 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 roll number in a specific location on a paper and turning it in 5 minutes later – following a class activity. Failure to complete the tasks in line with the instructions indicates a failure to attend to the instructions.

(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).

4.5 Testing for Performance Declines across Domains

The school administered exams described above are the standard metric of the total gain in human capital achieved by the interventions to enhance focus. However, these gains can be achieved via multiple channels—for example, improving learning in the classroom or less cognitive fatigue while taking the test improving test performance. Both these channels are important and policy relevant in their own right. We focus on the latter channel in constructing our tests—in line with the approach in Section 2—simply because this isolates focus over time, providing a positive test of our mechanism. Consequently, we design and administer tests to students that enable us to cleanly examine decline effects. These tests aim to accomplish three goals.

First, these tests aim to provide positive evidence for improvements in sustained attention. To accomplish this, we randomize the order of the questions in each exam and estimate the rate of decline over time on task.[22] Our key prediction is that students with better sustained attention will show less decline in performance over time. Thus, while students in the control and treatment conditions may perform similarly well in the initial test questions for content that was not directly trained, 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.

Second, the hypothesized channel of improved ability to sustain focus suggests effects should be broad-based. Effects should not only be present in tasks "similar" to those trained (e.g. math), but also in tasks unrelated to the tasks undertaken in the intervention (e.g. one's ability to listen and retain information). To determine whether such effects are present, we conduct multiple tests including ones which are far from the tasks trained. Each of these individual tests is described in greater detail below:

(1) **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.

(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

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

version of the test, adapted for appropriateness for each grade level.[23]

(3) **Math.** A standard paper-and-pencil test, which focuses on the content in the math curriculum for each student's given grade level. This test was chosen because of its direct policy relevance. Students' performance on the test was counted toward their final mathematics grade, giving the test natural stakes.

Third, we use specific design features of these tests — in combination with the prediction that we should observe differences in the rate of decline, but not the initial level of performance — to rule out other competing mechanisms. For example, if the treatment groups were more motivated to work hard and try, then we might expect to see them do better in a listening test even in the first questions at the start of the test. In contrast, if our treatments only improved sustained attention, then all groups should do equally well at the start of the test, but treatment students should show less decline in performance over time.[24] Finally, we directly test for some confounds through incorporating ancilliary tests. For example, we randomized incentives on a subset of tests (with students receiving candy if they did well) in order to examine the potential role of motivation. A list of potential confounds and the design features and tests used to examine them are described more fully in Section 7.

All tests are conducted during the school day, either in program class time or during additional study hall periods. Test varied in length by grade, with a minimum of 15 minutes and a maximum of 30 minutes. Note that students interpreted these tests as being regular school tests.

4.6 Timing

The intervention was conducted between September and January. School administered tests were given mid-year (December) and at the end of the year (March). Study administered tests (listening, ravens, math) and the traditional psychology measures were generally 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. These logistical constraints impacted all arms of the study equally and did not result in any imbalances in measurement of outcomes. Classroom behaviors were measured at endline only.

[23]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.

[24]In addition, 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 1). Students were instructed to take the tests in the order provided. Test monitors report these instructions were nearly always followed. Beyond these general design features, we also included some features to isolate channels are specific to individual tests. For example, one test (listening) operates on a defined timeline, ruling out the use of test taking strategies.

5 Empirical Approach

5.1 Intention to Treat and Treatment Intensity Analyses

Analysis of the school administered tests, traditional psychology measures of sustained attention, and classroom measures of behavior rely primarily on a simple intention to treat analysis. Specifically, we estimate:

$$z\_score_s = \beta_0 + \beta_1 Treated_s + \beta_2 Baseline_s + \gamma_1' X_i + \epsilon_s \tag{1}$$

Where z_score_s denotes the normalized score for the student.[25] $Treated_s$ represents a treatment indicator (or indicators for each treatment arm, when treatment arms are disaggregated). The regression also includes a baseline measure of the outcome when available – $Baseline_s$ – as well as a vector of relevant controls X_i as noted in each table. When more than one observation per student is available, standard errors are clustered by student.

This analysis is supplemented by a treatment intensity analysis in which we estimate performance on of these outcomes as a function of the number of hours of treatment received. These regressions follow equation 1 where $Treated_s$ is replaced with the number of hours of treatment received, where control students are coded as receiving zero hours. Recall that variaton in treatment hours was largely due to when we began working in a given school, and the school's preset calendar of holidays, festivals, and exams. Consequently, while all students within a school (and therefore across treatment arms) received the same exact number of class sessions, treatment hours was not explicitly randomly assigned across schools, and could therefore be correlated with other features of schools. We therefore view the analysis using treatment intensity as suggestive, and rely primarily on our Intent to Treat estimates using dummies for assignment.

5.2 Performance Declines in Tests

Consistent with the predictions of limited ability to sustain attention, we expect that students' performance will decline over time during any cognitive challenging activity. Because question order is randomized in all the classroom tests we administer to students, such a decline captures cognitive fatigue rather than changes in question difficulty. Specifically, we 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). To test whether our interventions improved students' ability to retain focus, we examine treatment effects on the extent of decline as students reach later parts of the test.

[25] For all indices, scores on individual outcomes were multiplied by -1 when needed such that all outcomes have the feature that "more is better".

We estimate:

$$\begin{aligned}&Correct_{ils} = \beta_0 + \beta_1 Treated_s + \textstyle\sum_{l=2}^{10} \lambda_l Location_{il} + \beta_2 Treated_s * 1[2 \leq Location_{il} \leq 5] \\ &+\beta_3 Treated_s * 1[6 \leq Location_{il \leq 10]} + \beta_4 Baseline_s + \chi_i + \epsilon_{ils}\end{aligned} \tag{2}$$

Where $Correct_{ids}$ denotes whether a question item, i, in location (decile), l, for child, s, 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. λ_2 - λ_{10} are location (decile) bins which flexibly capture declines in control group.[26] 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. β_3 is primary coefficient of interest, indicating whether there is differential fatigue among treated students. We hypothesize that β_3 will be positive (the treatments will ameliorate the rate of decline). $Baseline_s$ controls for the child's baseline test scores. χ_i are question fixed effects, controlling for the difficulty of the test item. For inference, we cluster standard errors by student—the unit of randomization—throughout the analysis.

One potential concern in estimating Equation 2 is that the estimation (inflexibly) takes a stance on the rate of decline. However, it is not clear ex-ante when such declines should occur. For example, the rate of decline may be a function of the difficulty of the content, the length of the exam, or baseline attentional capacity of students. Hence, we supplement our first specification with an additional higher-power specification which takes a two-step approach. We first flexibly estimate the rate of decline according to Equation 3 and then use this variable as a proxy for the "potential" decline at each point in the endline tests in estimating Equation 4. This approach draws on the intuition that there is no scope for a treatment effect unless there is a decline (improving power) and allows us to estimate the fraction of the decline mitigated by the treatment. The decline variable is estimated for each school separately to account for variation in skill across schools, and is derived from the *baseline* data.[27]

$$BaselineDecline_l = \tfrac{1}{SI} \textstyle\sum_{s=1}^{S} \sum_{i=1}^{I} [(Correct_{ils} | quintile = 1) - (Correct_{ils} | quintile = l)] \tag{3}$$

$$\begin{aligned}Correct_{ils} = \beta_0 + \beta_1 Treated_s + \beta_2 BaselineDecline_l + \beta_3 Treated_s * BaselineDecline_l \\ + \beta_4 Baseline_s + \chi_i + \epsilon_{ils}\end{aligned} \tag{4}$$

[26] 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%.

[27] This approach relies on the assumption that baseline decline rates are predictive of endline decline rates over time—an assumption we can directly verify through coefficient β_2 in Equation 4 below.

In this specification, β_3 captures the extent to which the treatments mitigate the rate of decline. As with Equation 2, we predict that β_3 will be positive. To account for the fact that $BaselineDecline_l$ is estimated from the data, we bootstrap standard errors.

6 Results

The randomization was smooth, with no more than the expected level of imbalance across all testing outcomes at baseline (Table A.I). In addition, attrition was low (5%) and very well balanced across experimental arms (Table A.II).

6.1 School Administered Exams

Treated students show noticeable improvements on the exams administered by the schools. Pooling across the core subjects of English, Hindi, and Math we find a 0.09 SD improvement in performance (Table 2). These level of improvement is notable given the short duration of the intervention – an average of approximately 6-15 hours of focused practice integrated into schooling across four months. When scaled spent for time in the intervention, the effects are similar in magnitude to other computer assisted learning interventions in India (Muralidharan et al., 2019). The magnitude of the increase is similar to that found in Project Star, which substantially reduced classroom sizes for an entire year (Krueger and Whitmore, 2001). Similarly, the effects are also only slightly smaller than those found for tracking students by ability (0.14 SD) or a remedial education program with an additional teacher (0.14 SD) with continuous exposure to the interventions over an entire school year in each of these programs (Duflo et al., 2011; Banerjee et al., 2007).

Beyond the intent to treat analysis we also leverage the variation in time spent on the intervention. In Table 2 columns 6 we regress performance on the classroom exams as a function of hours spent on task (where control students receive 0 hours). These results reinforce the findings in Column 1 and further indicate that the improvement scale with the time spent training.

Notably, as seen in columns 2 through 4, these results are not simply driven by improved performance by students in the Math arm improving performance markedly on the math exam. Rather, performance improvements are consistent across each of the subjects despite the fact that neither treatment arm trained in English or Hindi.[28] Further, as seen in columns 5 and 7, the point estimates between the two treatment arms are similar in magnitude (0.08 SD (0.012 SD) for Math and 0.10 SD (0.013) for Games in the intent to treat analysis (treatment intensity) analysis) and are statistically indistinguishable, suggesting both treatments were effective.[29]

[28] A subset of questions for the Math arm did include English text (e.g. Add 1 and 4). However, the questions for the Control arm study hall were drawn from the same question bank. Notably neither the Math nor the Games arm involved any additional exposure to Hindi.

[29] The effect of each treatment is also similar when disaggregating by test. See Table A.III).

6.2 Measures of Sustained Attention from Psychology and Classroom Behaviors

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 0.08 SD ($p < 0.05$) (Table 3). These results are not only similar across treatment arms (column 4: 0.09 SD for Math and 0.07 for Games), but are increasing with treatment intensity as demonstrated in columns 5 and 6.

In addition, the effects also appear to generalize to observable classroom behaviors. As seen in Table 4, treated students improve on an index of three measures of classroom attention adapted from a teacher rating scale used to measure ADHD by 0.17 SD, with the effects driven by improved ability to attend to and follow instructions and improved responses to stimuli. Similar to both the classroom tests and the sustained attention measures drawn from the psychology literature, these results are indistinguishable across treatment arms (column 5) and increasing in exposure to the treatment (columns 6 and 7).

6.3 Experimental Exams: Listening, Ravens Matrices, and Math

6.3.1 Overview

As described previously, we also administer a set of three exams – listening, Raven's Matrices, and math – via the study to further elucidate mechanisms. In these tests, we randomize the order in which the questions are administered and estimate the rate of decline in performance over time. The results of these tests also support the hypothesis that the treatments generated a broad-based improvement in the ability to sustain focus. Relative to the control group, students in each treatment arm show an improved ability to maintain focus in a range of disparate activities: Listening tests, Raven's Progressive Matrices, and Math tests — providing evidence for the underlying generalizability of the cognitive effects. Specifically, treated students experience 15% to 25% less decline in the second half of the test (Table 5). Also consistent with our proposed mechanism, treatment and control students generally perform similarly in the beginning of each test (Table 5 columns 2 - 4).[30]

6.3.2 Non-parametric Plots of Performance Over Time

We begin by providing non-parametric local polynomial plots of the treatment effects in Figures 9, 10, and 11. Because initial performance is quite similar across tests and arms (Table 5), to more clearly visualize declines initial levels are normalized to 0.[31] Each of these plots shows two consis-

[30]One exception to the fact that performance is nearly identical at the beginning of the test is that students in the Games treatment perform slightly worse at the beginning of the Math test. This is consistent with the fact that both the Math and Control arms gained additional practice in math while the Games arm did not.

[31]Plotting the raw data in the listening test, there are clear "reset" effects between passages. Hence, we examine declines within each passage which are substantial.

tent patterns. First, similar to the declines observed in the TIMSS data, students perform worse on a given question item if it occurs later in the test and students are cognitively fatigued. This is true despite very high test completion rates and the fact that question order is randomized and we residualize on question fixed effects to control for question difficulty. Second, consistent with the hypothesis, treated students' performance declines more slowly across the course of the exam, with improvements of roughly 15% to 25%. We discuss each plot in more detail below.

Listening. Initial performance is, again, not statistically different across Treated and Control students, but a gap emerges fairly rapidly and continues to grow over throughout the exam. In the final decile, treated students exhibit 24% less decline than 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.

Ravens. Performance on Raven's Matrices, often taken as an IQ test, shows a similar overall pattern. Drawing on regression results in Table 5, Treated students decline 13% less in the second half of the exam. However, overall impacts are even larger as Treated students are able to better maintain attention from fairly early in the exam, with an increasing gap over time (Figure 10). These variations in the exact pattern of decline may be related to variations in the difficulty of the tasks and underscore the importance of the flexible functional form used to estimate these effects.

Math. Finally, we see similar overall patterns in the math test. Similar to other low income countries in the TIMSS data, declines in performance over time are substantial among control students. Control students are roughly 10 percentage points more likely to answer a question correctly if it occurs at the beginning of the test rather than at the end of the test (Figure 11). The roughly 15 to 20 hours of treatment reduces this decline by 14% in the second half of the exam.

6.3.3 Pooled Results

Drawing on the empirical approach described above, we also assess the statistical significance of these effects. As shown in Table 5, there are no significant differences in initial levels of performance across any of the tests. Overall, the level differnce between the Treatment and Control students is -0.001, or .1%. The magnitude of this difference is similar across each of the three tests (column 2). Also consistent with our hypothesis, pooled across all three exams, the reductions in the rate of decline among treated students are both meaningful in terms of magnitude — roughly one-quarter 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.

These effects are even more notable given the relatively limited training in this program and the diverse subject matter tested. 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 impact on the ability to sustain attention over time. Further, the school administered tests suggest that gains in attention may aggregate further through direct learning in the classroom for those skills taught in the classroom.

6.3.4 Math and Games Treatments Have Similar Effects

Each of the treatments was designed to increase the time spent sustaining focus on a task. As such, we would also expect that both treatments should be effective at mitigating declines in performance in the second half of the test. Consistent with this hypothesis, the results in Table 5 columns 3 and 4 show very similar impacts of the two treatments, with coefficients on the effect in the second half of the test of 0.0127 and 0.0132 for Math and Games, respectively.

6.3.5 Effects are also Similar by Test

Drawing on Equation 4 to improve precision, we also examine the effects of the interventions on each exam individually in Table 6. The treatment effects are relatively similar in magnitude across each of the tests, ranging from 0.08 for Raven's and Listening to 0.10 for Math.

6.4 Persistence

To enable a test for persistence in effects, we conducted a three-month follow-up, in which we again administered tests in math, listening, and Ravens Matrices. These tests were conducted after a one-month break between when students progress from one grade to the next; during this break, students do not attend school—providing a stronger test for persistence.

The prior analysis pooled these follow-up tests with the main midline and endline tests for power. In Table 7, we separately estimate the treatment effects in the 3-month follow-up by adding an interaction term for the follow-up tests. We show these results for each of our two estimation strategies (Equations 2 and 4) in Cols. (1) and (2) respectively. In each specification, the interaction term is essentially zero and insignificant. In addition, in the higher-powered specification in Col. (2), treated students show significantly less decline than control students (p-value 0.0528, reported at the bottom of the table).

6.5 Summary

The treatments produce broad and generalizable improvements in the ability to sustain focus over time. Treated students improve performance on a range of classroom test by 0.1 SD and experience

15% to 25% less decline on unrelated tasks using tests to cleanly isolate attention as a key channel driving these effects. The results are further 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. Further, the improvement across a wide range of domains (e.g. Hindi, English, Math, listening, IQ) as well as test types ranging from traditional measures of human capital accumulation such as school administered tests to standard cognitive psychology measures suggests that these effects are very widely applicable and are likely to be large when aggregated across the many contexts in which they apply.

7 Potential Confounds

Overview. As noted above, the study administered tests were explicitly designed to isolate the mechanism of sustained attention while ruling out potential confounds. One core element of this approach is that the training provided to treated participants targets the ability to sustain focus rather than the test domain itself. Hence, a key prediction of our study is that the treatments will mitigate declines in performance over time without systematically impacting the initial level of performance for tests where no direct training occurs.[32] 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.

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).

[32] The math test in which the Math and Control students receive additional practice relative to the Games students is the notable exception.

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).[33] These additional incentives do not alter the results (A.IV).

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.[34]

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.

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 1).[35] 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

[33]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.

[34]SART, which must be electronic to accurately measure reaction times, is computer-based. However, the task does not require any knowlege of technology – one simply presses the spacebar when a stimulus appears on a screen – and we specifically administered the task on a computer with a large independent keyboard to make it as distinct as possible from the tablet-based interventions.

[35]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.

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 5, no such level differences exist, isolating an attentional channel.

8 Discussion: Schooling Approaches in High vs. Low Income Schools

These findings suggest that the pedagogy used in schools, as well as other school based practices, have the potential to influence not only the formation of "hard skills" and "soft skills", but also one's ability to sustain focus – a fundamental element of cognition. One common feature of schools catering to lower SES students, is that students often spend more time engaged in activities that require less sustained focus. For example, rote memorization such as chanting "1 plus 1 is two" or having students to occupy themselves unproductively while a teacher grades papers is quite common in these environments. In contrast, schools catering to higher SES students typically spend more time in student-focused independent practice.

We validate this pattern using the classroom time use data in the TIMSS dataset. Table 8, Column (1) shows that schools in high-income countries spend much more class time on giving students opportunities to practice skills independently, relative to teacher-centered lecture.

Our experimental findings suggest that children who are more exposed to such "attention heavy" pedagogy, would experience less rapid declines. Drawing once more on the TIMSS data, we find that this is indeed true (Table 8, Column (2)). Even more notably, this relationship holds even when we control for income differences that are correlated with pedagogy (Column (3)). In other words, pedagogy that is more similar to our intervention is strongly predictive of a slower rate of decline holding constant income. Although only correlational, this evidence is consistent with our experimental results—suggesting that the results could be much more widely applicable.

9 Conclusion

It is well known that schooling influences both "hard skills" such as literacy and numeracy and "soft skills" such as disciplinary behavior and socioemotional regulation. This paper suggests it may go further and shape a basic element of human capital essential to many aspects of life: attentional capacity.

Providing roughly 20 hours of focused practice to low-income students in India improves overall performance in a variety of subject areas – ranging from Hindi to Math – and reduces the rate of decline in performance across a wide range of unrelated tasks – such as listening retention and Raven's matrices. These gains were complimented by improvements in traditional cognitive psychology measures of sustained attention as well as changes in classroom behavior.[36]

While these results are encouraging, viewed from a different angle, this result also suggests differences in the quality of schooling or reliance on pedagogy that does not promote practice of this skill may widen long-run economic disparities. The foundational nature of this skill means that even small differences in attentional declines may have broad and economically meaningful consequences when aggregated. For example, we provide evidence of declines in attention in areas as diverse as schooling, errors at work, and voting. Further, we document the novel fact that the rate of decline in attention varies substantially with socio-economic status. Reducing these declines by the average change among treated students in the RCT could result in broad and economically meaningful improvements in these outcomes.

Yet, these are just a few of the many domains that would be affected. Further research to flesh out the scope and consequences of these declines in additional domains would allow us to more fully assess the downstream consequences of the improvements in the ability to sustain focus. For example, are high rates of traffic accidents in developing countries also potentially influenced by rapid attentional declines, for example during long shifts worked by truck and auto drivers?

Further, these results raise the question of whether schooling has the potential to influence other elements of cognition. Overall, our study provides an initial proof of concept that one important component of basic cognition—attentional capacity—is malleable. Examining the extent to which our cognitive capacity is shaped by the environment is an important direction for future research, with implications for understanding the the role of schooling in productivity as well as the underpinnings of economic mobility and inequality.

[36]It is important to note that one challenge in promoting improvements in attentional capacity is that it is unlikely that a single activity can be used to train this skill across diverse populations. What holds a child's attention in a low-income school in India may not be what holds a child's attention in an affluent suburb in the United States. Given this challenge, we view this intervention as a proof of concept. While there is some intuition about what activities are likely to train this skill (e.g. active participation rather than rote memorization, dynamically adaptive tasks rather than static ones), much more needs to be done to understand what features of an activity generate sustained engagement over time.

References

Acemoglu, Daron et al., "What does human capital do? A review of Goldin and Katz's the race between education and technology," *Journal of Economic Literature*, 2012, *50* (2), 426–63.

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.

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

Baird, Sarah, Joan Hamory Hicks, Michael Kremer, and Edward Miguel, "Worms at Work: Long-run Impacts of a Child Health Investment," *The Quarterly Journal of Economics*, 07 2016, *131* (4), 1637–1680.

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.

_ , Shawn Cole, Esther Duflo, and Leigh Linden, "Remedying education: Evidence from two randomized experiments in India," *The Quarterly Journal of Economics*, 2007, *122* (3), 1235–1264.

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

Berger, Eva M, Ernst Fehr, Henning Hermes, Daniel Schunk, and Kirsten Winkel, "The Impact of Working Memory Training on Children's Cognitive and Noncognitive Skills," *Working Paper*, 2020.

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.

Broer, Markus, Yifan Bai, and Frank Fonseca, "A review of the literature on socioeconomic status and educational achievement," in "Socioeconomic Inequality and Educational Outcomes," Springer, 2019, pp. 7–17.

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.

Duflo, Esther, "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *THE AMERICAN ECONOMIC REVIEW*, 2001, *91* (4), 49.

_ , Pascaline Dupas, and Michael Kremer, "Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya," *American Economic Review*, 2011, *101* (5), 1739–74.

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.

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.

Gagnon-Bartsch, Tristan, Matthew Rabin, and Joshua Schwartzstein, *Channeled attention and stable errors*, Harvard Business School, 2018.

Gneezy, Uri, John A List, Jeffrey A Livingston, Xiangdong Qin, Sally Sadoff, and Yang Xu, "Measuring success in education: the role of effort on the test itself," *American Economic Review: Insights*, 2019, *1* (3), 291–308.

Hackman, Daniel A. and Martha J. Farah, "Socioeconomic status and the developing brain," *Trends in Cognitive Sciences*, February 2009, *13* (2), 65–73.

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.

Heckman, James, Rodrigo Pinto, and Peter Savelyev, "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review*, October 2013, *103* (6), 2052–2086.

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.

Hopfenbeck, Therese N, Jenny Lenkeit, Yasmine El Masri, Kate Cantrell, Jeanne Ryan, and Jo-Anne Baird, "Lessons learned from PISA: A systematic review of peer-reviewed articles on the programme for international student assessment," *Scandinavian Journal of Educational Research*, 2018, *62* (3), 333–353.

Jackson, C. Kirabo, "What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes," *Journal of Political Economy*, October 2018, *126* (5), 2072–2107. Publisher: The University of Chicago Press.

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.

Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan, "Self-control at work," *Journal of Political Economy*, 2015, *123* (6), 1227–1277.

_ , Sendhil Mullainathan, Suanna Oh, and Frank Schilbach, "Do Financial Concerns Make Workers Less Productive?," Technical Report, National Bureau of Economic Research 2021.

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.

Knudsen, Eric I, James J Heckman, Judy L Cameron, and Jack P Shonkoff, "Economic, neurobiological, and behavioral perspectives on building America's future workforce," *Proceedings of the national Academy of Sciences*, 2006, *103* (27), 10155–10162.

Krueger, Alan B and Diane M Whitmore, "The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR," *The Economic Journal*, 2001, *111* (468), 1–28.

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.

Muralidharan, Karthik, Abhijeet Singh, and Alejandro J Ganimian, "Disrupting education? Experimental evidence on technology-aided instruction in India," *American Economic Review*, 2019, *109* (4), 1426–60.

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.

Association For The Evaluation Of Educational Achievement, "TIMSS 2011 Assessment," Technical Report, TIMSS PIRLS International Study Center, Lynch School of Education, Boston College, Cambridge, MA 2013.

Schofield, Heather, "The Economic Costs of Low Caloric Intake: Evidence from India," *Working Paper*, 2015.

Warm, J. S. and W. N Dember, "Tests of vigilance taxonomy," in "Viewing psychology as a whole: The integrative science of William N. Dember," American Psychological Association, 1998, p. 87–112.

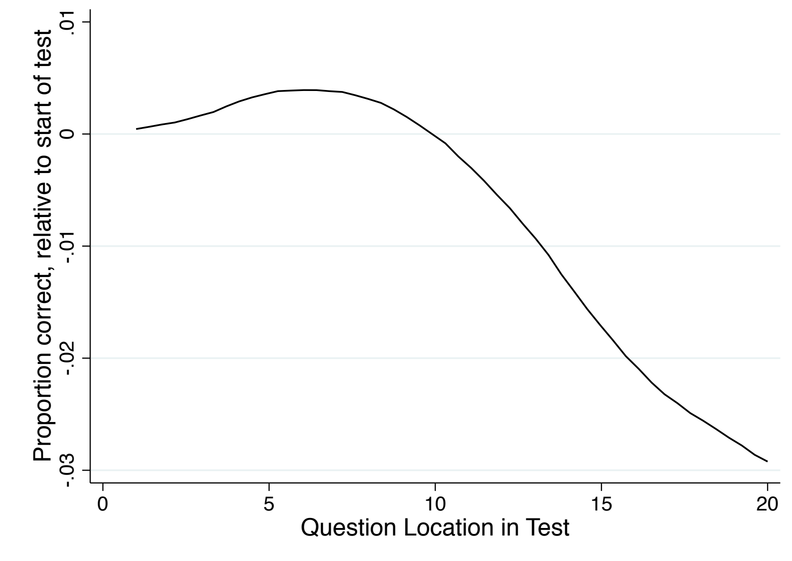
Warm, Joel S, Gerald Matthews, and Victor S Finomore Jr, "Vigilance, workload, and stress," in "Performance under stress," CRC Press, 2018, pp. 131–158.

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.

10 Figures

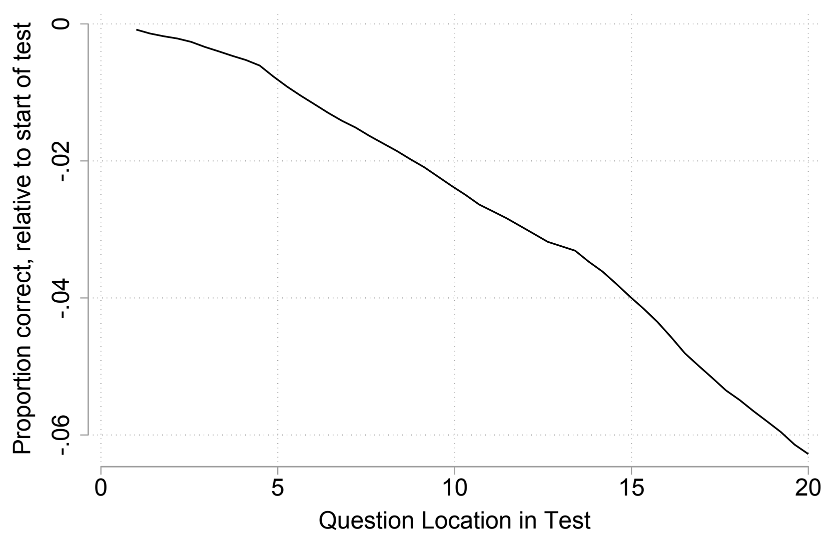
Motivation: Attentional declines are common and important

Figure 1: US Education - Declines in performance over time on achievement tests



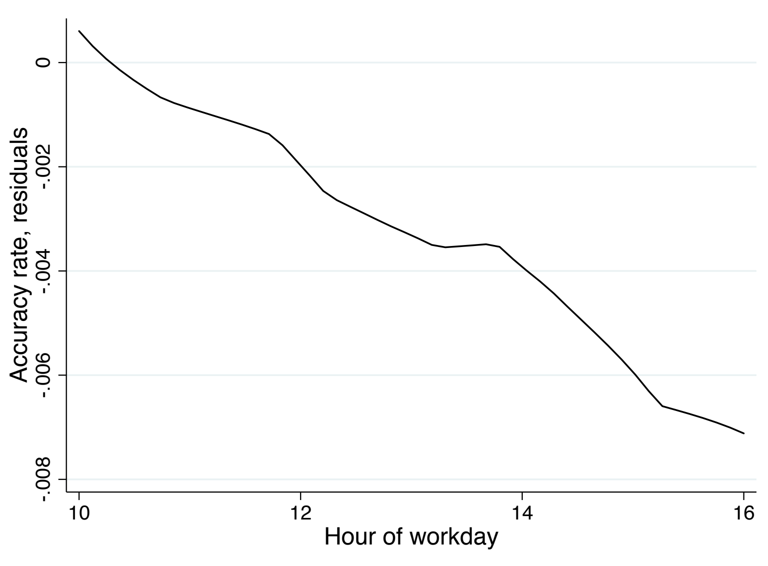
Notes: TIMSS data within a low-SES US population, authors' calculations. Question order is block randomized.

Figure 2: International Education - Declines in performance over time on achievement tests



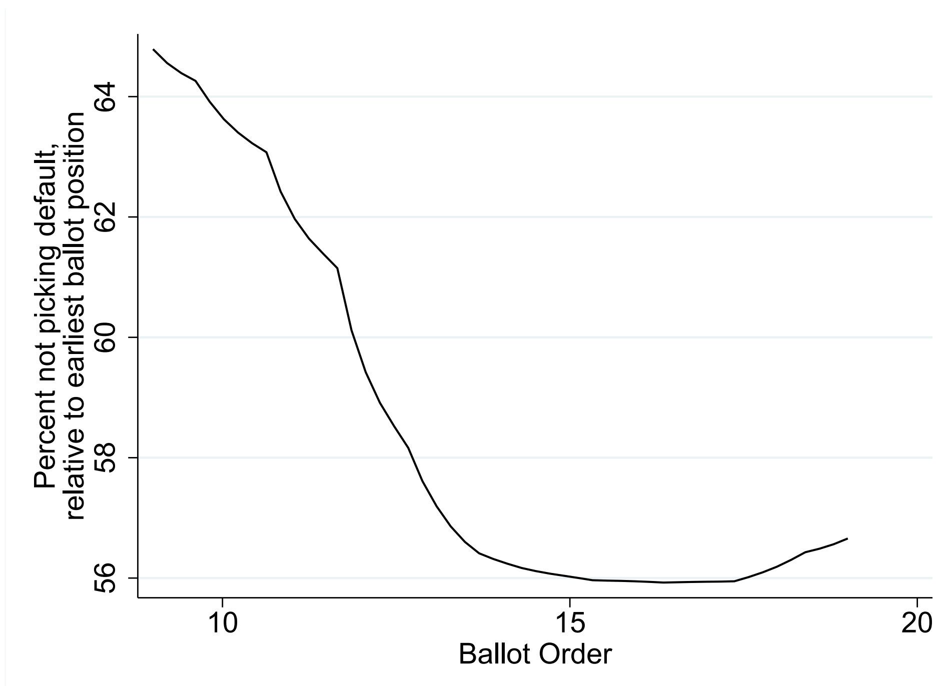
Notes: TIMSS data, low-SES countries around the world, authors' calculations. Question order is block randomized.

Figure 3: Worker Productivity - Declines in Data Entry Accuracy over the Day



Note: Data from Kaur et al. (2015). Accuracy rate is the proportion of fields entered with no errors. Data are residualized accounting for worker fixed effects. The sample is 8,382 worker-hours of data entry. The sample is restricted to paydays (when attendance is high to mitigate selection concerns), workers who were present from 10am-4pm on a given day (so that the composition of workers is constant within a worker-day during these hours), and below-median education levels. Patterns are similar without the first two restrictions.

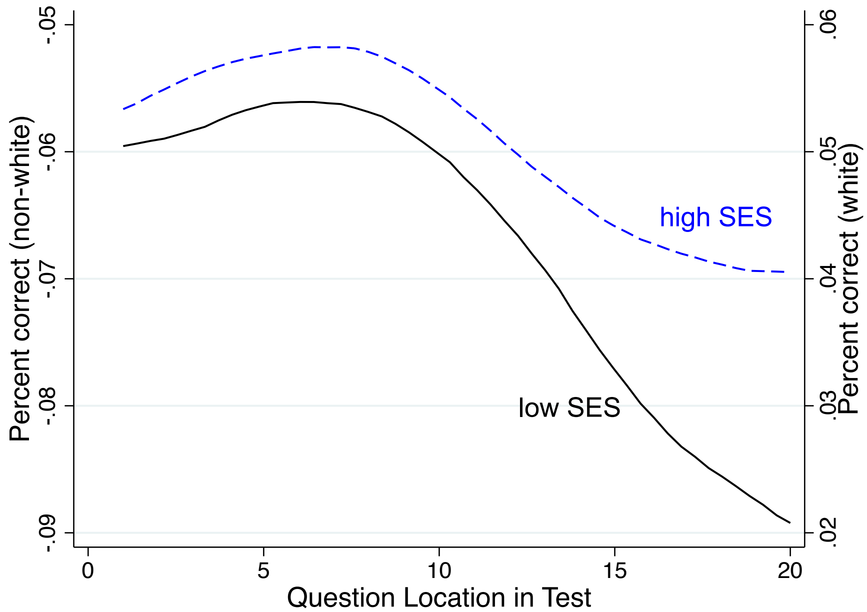
Figure 4: Voting - Declines in Cognitive Effort in the Voting Booth



Note: Source is Augenblick and Nicholson (2015). Item order is quasi-random. Low-SES voters become more likely to rely on defaults as an initiative is further down-ballot.

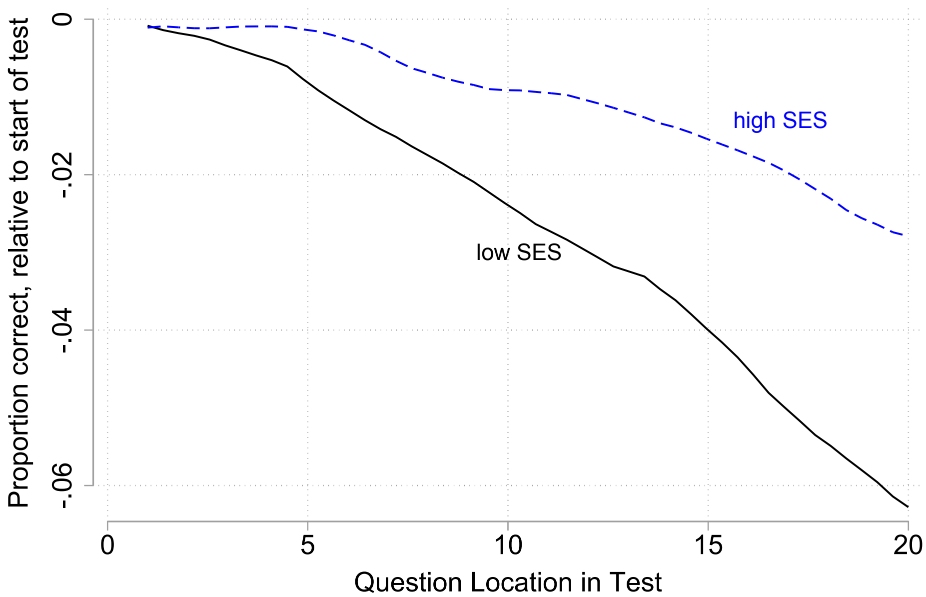
Motivation: Correlation between SES and Attentional Decline

Figure 5: Declines in Achievement Tests - Heterogeneity by SES within the United States



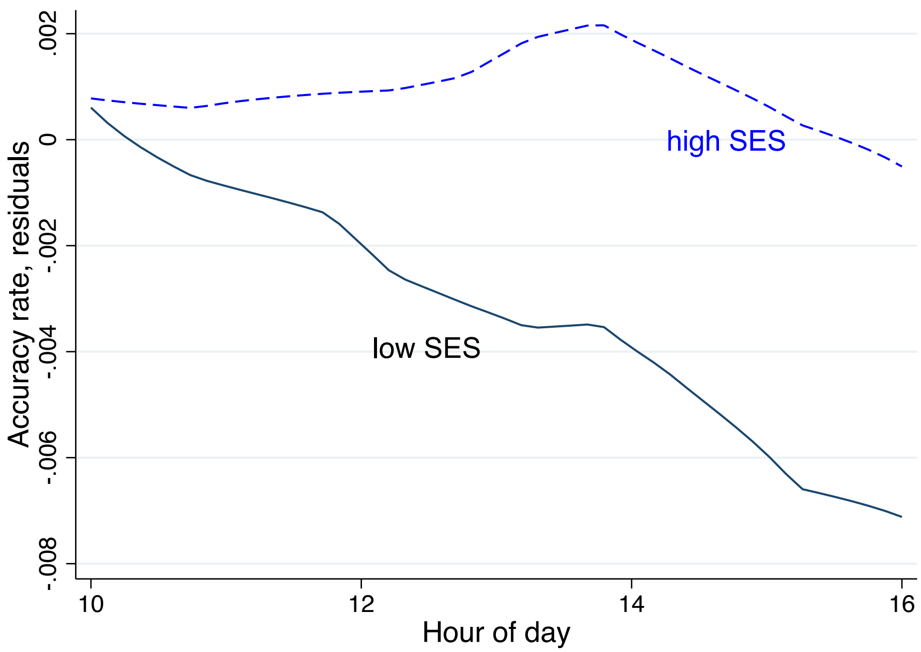
Notes: TIMSS data, authors' calculations. Question order is block randomized. SES is proxied by race (white, non white).

Figure 6: Declines in Achievement Tests - Heterogeneity by GDP per Capita across Countries



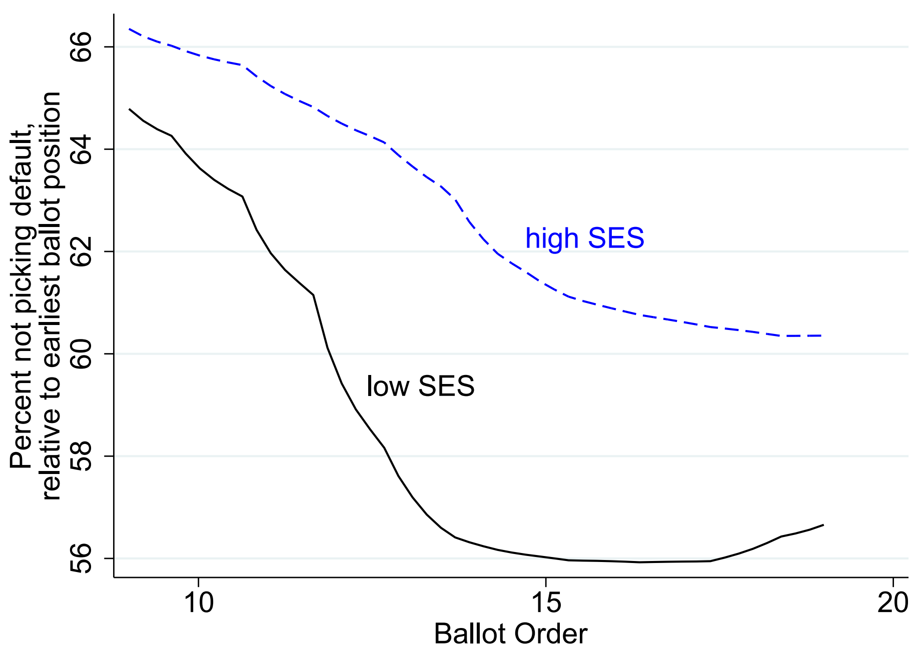
Notes: TIMSS data, authors' calculations. Question order is block randomized. Initial level differences are normalized to zero for comparison. High (low) SES countries are proxied by the top (bottom) decile of GDP/capita.

Figure 7: Declines in Worker Accuracy - Heterogeneity by Education



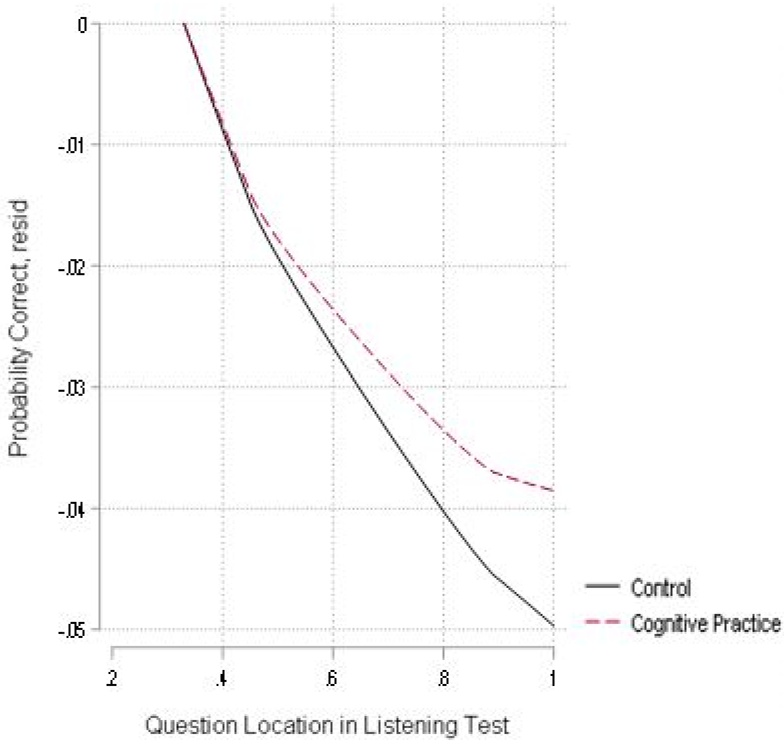
Note: Data from Kaur et al. (2015). Accuracy rate is the proportion of fields entered with no errors. Data are residualized accounting for worker fixed effects. High SES is defined as 1 if the worker has above high school education (corresponding to the median split of the sample). The sample is 8,382 worker-hours of data entry (90 workers). The sample is restricted to paydays (when attendance is high to mitigate selection concerns) and workers who were present from 10am-4pm on a given day (so that the composition of workers is constant within a worker-day during these hours). Patterns are similar without these restrictions.

Figure 8: Declines in Cognitive Effort in the Voting Booth - Heterogeneity by SES



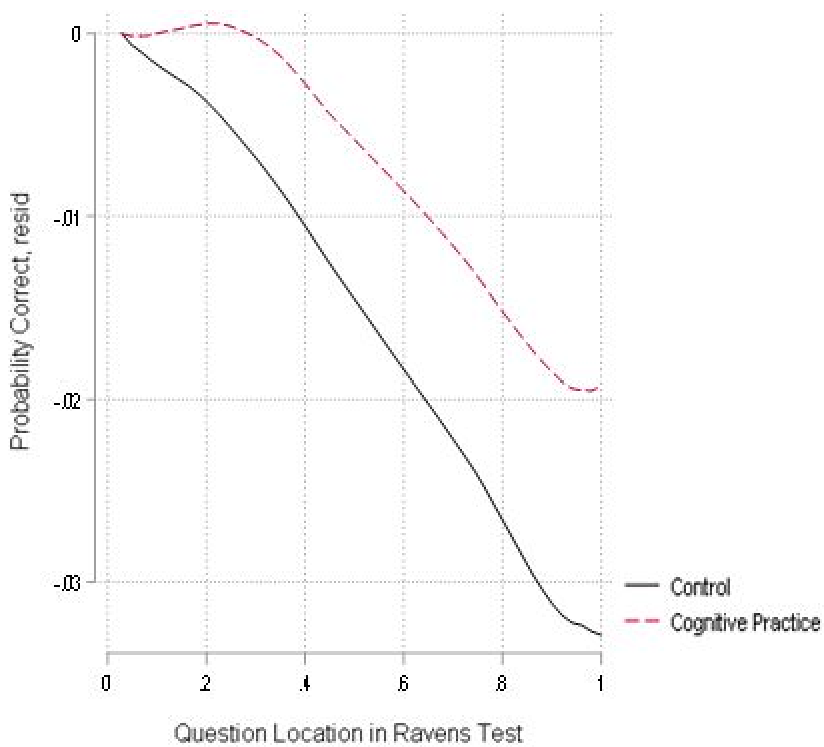
Note: Data from Augenblick and Nicholson (2015) and the United States census. Item order in the voting data is quasi-random. Voters become more likely to rely on defaults as an initiative is further down-ballot and these differences are larger for less affluent neighborhoods as proxied by a lower fraction of non-Hispanic white individuals in the precinct using a median split.

Figure 9: Training Slows Decline in Listening Comprehension Test



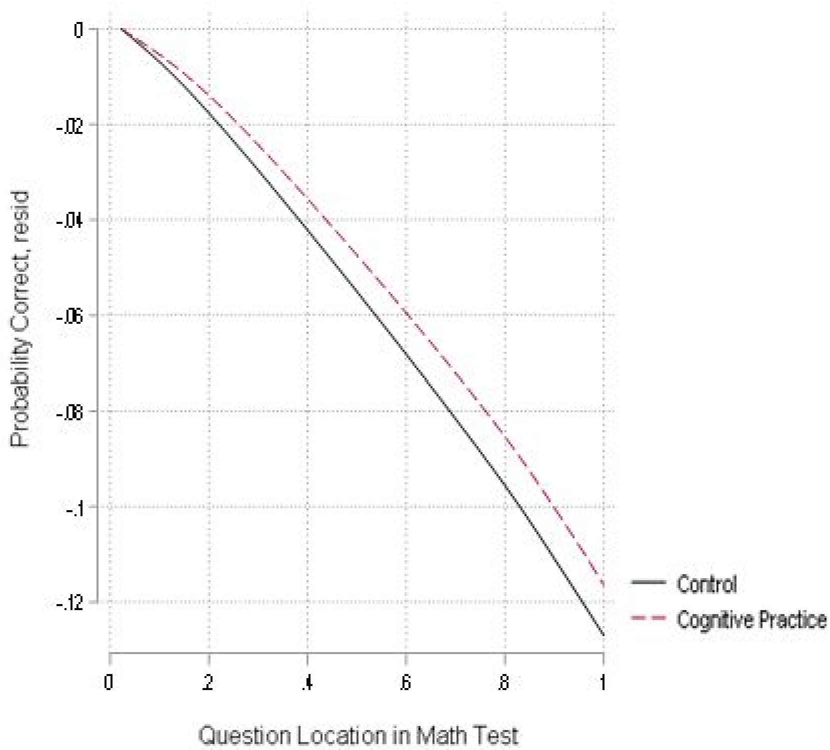
Treated students exhibit **21% less decline** in the second half of the exam (Table 5).

Figure 10: Training Slows Decline in Ravens Matrices (IQ) Test



Treated students exhibit **13% less decline** in the second half of the exam (Table 5).

Figure 11: Training Slows Decline on Math Test



Treated students exhibit **14% less decline** (0.1 SD) in the second half of the exam (Table 5).

11 Tables

Table 1: Test completion

	Math	Listening	Ravens
% attempted	0.773	0.996	0.992
% skipped	0.151	0.001	0.005
% of students completing last question item	0.793	0.996	0.983
Avg last question completed location	0.933	0.997	0.993

Table 2: Performance on School Exams

	Outcome: z-score						
	Treatment Dummy					**Treatment Hours**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Pooled	Math	Hindi	English	Pooled	Pooled	Pooled
Treat	0.0888**	0.0828**	0.0963**	0.0870**		0.0124***	
	(0.0382)	(0.0419)	(0.0457)	(0.0439)		(0.0036)	
Math					0.0748*		0.0115***
					(0.0448)		(0.0043)
Games					0.103**		0.0133***
					(0.0428)		(0.0041)
N	8125	2716	2715	2694	8125	8125	8125
R^2	0.355	0.370	0.355	0.383	0.355	0.356	0.356

*Notes: This table reports treatment effects on administrative exams administered by schools to students (pooling mid year and end of year exams). Observations are at the student-test level. The dependent variable is the student's z-score on the test. Treat denotes receiving any treatment, Math and Games denote the Math practice or Games practice treatments, respectively. Cols. (1)-(5) regress z-score on a dummy for treatment, Cols. (6)-(7) regress z-score on the number of hours of treatment the student received up until the date of the exam (with the number of hours equal to 0 for the control group). All regressions include class section (strata) fixed effects and baseline controls. Standard errors clustered by student. * p<0.10, ** p<0.05, *** p<0.01.*

Table 3: Psychology Literature Sustained Attention Tests

	Outcome: z-score					
	Treatment Dummy				**Treatment Hours**	
	(1) Pooled	(2) SART	(3) COS	(4) Pooled	(5) Pooled	(6) Pooled
Treat	0.0788**	0.110**	0.0562		0.0102***	
	(0.0361)	(0.0555)	(0.0423)		(0.00322)	
Math				0.0870**		0.0104***
				(0.0414)		(0.00374)
Games				0.0711*		0.0100***
				(0.0425)		(0.00385)
R^2	0.151	0.212	0.204	0.151	0.151	0.151
Number of obs.	9704	3897	5807	9704	9704	9704

*Notes: Outcomes are measured as (True positive z-score - False positive z-score). Clustered standard errors in parentheses. Controls for baseline test scores and the students' section. Observations are students. * p<0.10, ** p<0.05, *** p<0.01.*

Table 4: Classroom Observation Measures of Attentiveness

	Outcome: z-score						
	Treatment Dummy					**Treatment Hours**	
	(1) Index	(2) Task completion	(3) Response to stimuli	(4) Physical signs	(5) Index	(6) Index	(7) Index
Treat	0.169***	0.0971*	0.136**	-0.0452		0.0186***	
	(0.0598)	(0.0575)	(0.0614)	(0.0585)		(0.00521)	
Math					0.215***		0.0214***
					(0.0691)		(0.00602)
Games					0.123*		0.0157**
					(0.0694)		(0.00611)
R^2	0.099	0.153	0.071	0.136	0.101	0.103	0.104
Number of obs.	1203	1206	1205	1204	1203	1203	1203

*Notes: Index adapted from Vanderbilt ADHD diagnostic teacher rating scale. Classroom observers were blind to students' treatment status. Clustered standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.*

Table 5: Treatment Effects on Declines

Dependent Variable: 1[question correct]

	Definition of treat variable			
	Treat (pooled) (1)	Treat (pooled) (2)	Math (3)	Games (4)
Treat x Deciles 6-10	0.0114**	0.0129***	0.0127**	0.0132**
	(0.006)	(0.005)	(0.006)	(0.006)
Treat x Deciles 2-5	0.0062	0.0080	0.0032	0.0129**
	(0.006)	(0.005)	(0.006)	(0.006)
Treat	-0.00120			
	(0.006)			
Treat x Math test		-0.00441	0.0039	-0.0129
		(0.008)	(0.010)	(0.010)
Treat x Listening test		-0.0002	0.0012	-0.0016
		(0.007)	(0.008)	(0.008)
Treat x Ravens test		-0.0032	-0.0064	-0.0000
		(0.009)	(0.010)	(0.010)

Notes: Question item order was randomized across students. All regressions contain question and test version fixed effects, and baseline controls. The omitted category is the control arm of the experiment. This table was estimated as specified in Equation 2. Coefficients in cols. (3) - (4) are estimated from a single regression on all the data. Standard errors clustered by student. N=1,632 students; 352,847 question items. The dependent variable mean is 0.47 in the control group.

Table 6: Treatment Effects on Declines: Expected Declines

Dependent variable: 1[question correct]

	Test Subject			
	All tests (pooled)	Math	Listening	Ravens
	(1)	(2)	(3)	(4)
Treat x BaselineDecline	0.0801***	0.102**	0.0829**	0.0763*
	(0.026)	(0.043)	(0.039)	(0.038)
Treat	-0.0017	-0.0094	-0.0006	-0.003
	(0.0071)	(0.011)	(0.008)	(0.009)
BaselineDecline	-0.162***	-0.414***	-0.4196***	-0.0816***
	(0.027)	(0.045)	(0.038)	(0.04)

Notes: Question item order was randomized across students. All regressions contain question and test version fixed effects, and baseline controls. This table was estimated as specified in Equation 4. BaselineDecline is the amount of average decline in each quintile of the test location, relative to the first quintile of the test, within each given school. Coefficients in cols. (2) - (4) are estimated from a single regression on all the data. Bootstrapped standard errors, corrected for clustering by student. N=1,632 students; 352,847 question items. The dependent variable mean is 0.47 in the control group.

Table 7: Persistence of Treatment Effects

Dependent variable: 1[question correct]

	(1)	(2)
Treat x Deciles 6-10	0.0143***	
	(0.006)	
Treat x Deciles 6-10 x Follow-up	-0.00395	
	(0.012)	
Treat x BaselineDecline		0.0856***
		(0.031)
Treat x BaselineDecline x Follow-up		-0.00237
		(0.044)
F-test p-value: Sum of 2 coefficients = 0	0.3123	0.0528
R-squared	0.359	0.359
N	334500	334500

Notes: Follow-up tests were administered 3 months after the end of the intervention. Question item order was randomized across students. All regressions contain baseline controls, question fixed effects, and test version fixed effects. Standard errors corrected to allow for clustering by student. Bootstrapped standard errors presented in Col. (2).

Table 8: Pedagogy and Attentional Declines

Outcome:	Indep Prac Time	Question Item Correct	
	(1)	(2)	(3)
Question location		-0.0548***	-0.149***
		(0.00350)	(0.0178)
Use of high income pedagogy		0.0717***	0.0362***
		(0.00290)	(0.00288)
Use of high income pedagogy*Question Location		0.0120***	0.00639*
		(0.00309)	(0.00326)
Log GDP	0.0261***		0.0631***
	(0.00222)		(0.00170)
Log GDP* Question location			0.0101***
			(0.00184)
Constant	0.0368*	0.487***	-0.104***
	(0.0211)	(0.00314)	(0.0156)
Dependent variable mean	0.283	0.458	0.458
R^2	0.025	0.119	0.166
Observations	235830	333801	333801

*Notes: Source is the TIMSS dataset. Indep Prac Time is the percent of math class time students spend working on problems independently. Clustered standard errors are in parentheses. * p<0.10, ** p<0.05, *** p<0.01.*

A Supplementary Tables and Figures

Table A.I: Baseline Balance

Variable	(1) Control N	(1) Control Mean/SE	(2) Games Practice Treatment N	(2) Games Practice Treatment Mean/SE	(3) Math Practice Treatment N	(3) Math Practice Treatment Mean/SE	T-test Difference (1)-(2)	T-test Difference (1)-(3)	T-test Difference (2)-(3)
Baseline Listening (mean)	556	0.228 (0.015)	558	0.193 (0.014)	558	0.234 (0.015)	0.035*	-0.007	-0.042**
Baseline Math (mean)	556	0.344 (0.010)	558	0.353 (0.011)	558	0.358 (0.010)	-0.009	-0.014	-0.005
Baseline Ravens Matricies (mean)	556	0.330 (0.011)	558	0.319 (0.012)	558	0.350 (0.012)	0.011	-0.020	-0.031*
Baseline Symbol Matching (mean)	556	0.475 (0.009)	558	0.459 (0.009)	558	0.455 (0.009)	0.015	0.020	0.004
Baseline Listening (decline)	556	0.009 (0.010)	558	0.002 (0.010)	558	-0.005 (0.011)	0.007	0.013	0.007
Baseline Math (decline)	556	-0.099 (0.015)	558	-0.089 (0.016)	558	-0.071 (0.015)	-0.009	-0.028	-0.019
Baseline Ravens Matricies (decline)	556	-0.061 (0.016)	558	-0.041 (0.016)	558	-0.071 (0.017)	-0.020	0.010	0.030
Focused practice index	292	-0.026 (0.060)	332	0.080 (0.052)	333	-0.048 (0.056)	-0.107	0.021	0.128*
Baseline nutrition index	277	-0.029 (0.062)	311	0.011 (0.056)	315	-0.001 (0.055)	-0.040	-0.028	0.012

Notes: The value displayed for t-tests are the differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels.

Table A.II: 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 A.III: Performance on School Exams - Disaggregated by Treatment

	(1) Pooled	(2) Maths	(3) Hindi	(4) English
Math - Training Hours	0.011***	0.012***	0.012**	0.010**
	(0.004)	(0.005)	(0.005)	(0.005)
Attention - Training Hours	0.013***	0.011**	0.017***	0.012**
	(0.004)	(0.005)	(0.005)	(0.005)
R^2	0.356	0.371	0.358	0.384
N	8125	2716	2715	2694

*Notes: This table reports treatment effects on administrative exams administered by schools to students (pooling mid year and end of year exams). Observations are at the student-test level. We regress z-score on the number of hours of each treatment the student received up until the date of the exam (with the number of hours equal to 0 for the control group). All regressions include class section (strata) fixed effects and baseline controls. Standard errors are clustered by student. * p<0.10, ** p<0.05, *** p<0.01*

Table A.IV: Effect on Incentives on Test Performance

	Item Correct		
	(1)	(2)	(3)
Incentive	0.00331	-0.0227	-0.0462
	(0.0218)	(0.0304)	(0.0851)
Treat		0.0104	-0.0243
		(0.0210)	(0.0653)
Treat*Incentive		0.0408	0.0100
		(0.0291)	(0.1160)
Treat*Decile 6-10			0.0388
			(0.0639)
Incentive*Decile 6-10			0.0401
			(0.0810)
Treat*Incentive*Decile 6-10			0.0312
			(0.1133)
N	10894	10894	10894
r2	0.240	0.218	0.241

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