FICTIONAL MONEY, REAL COSTS: IMPACTS OF FINANCIAL SALIENCE ON DISADVANTAGED STUDENTS *

Claire Duquennois

Abstract

Disadvantaged students perform differentially worse when randomly given a financially salient mathematics exam. For students with socio-economic indicators below the national median, a 10 percentage point increase in the share of monetary themed questions depresses exam performance by 0.026 standard deviations, about 6% of their performance gap. Using question-level data, I confirm the role of financial salience by comparing performance on monetary and highly similar non-monetary questions. Leveraging the randomized ordering of questions, I identify an effect on subsequent questions, providing evidence that the attention capture effects of poverty affect policy relevant outcomes outside of experimental settings. JEL Codes: D91, I24, I25, O15.

*Assistant Professor at the Department of Economics, University of Pittsburgh. Email: ced87@pitt.edu. Mailing Address: 4700 Wesley W. Posver Hall, 230 South Bouquet Street, Pittsburgh, PA, 15260. I would like to thank the ASSISTments team and in particular Korinn Ostrow for their help with the ASSISTments data. I would also like to thank Andrew Dustan for his assistance with the COMIPEMS data. I thank Lise Vesterlund, Aprajit Mahajan, Ethan Ligon, Marco Gonzalez-Navarro, Ted Miguel, Fred Finan, Jeff Perloff and David Sunding for their advice, as well as Megan Lang and the numerous attendees at seminars and workshops, whose suggestions have guided this project. Finally I would like to thank Jeremy Magruder, Supreet Kaur, Alain de Janvry and Elisabeth Sadoulet for their incredible advice and support throughout. All errors are my own.

1 Introduction

Performance on examinations matters. Test results are commonly used for assessment of students and schools, as an allocation criterion or admission requirement, and for licensing and certification. Student performance on examinations can thus have significant economic implications and determine future educational and economic opportunities.

Examinations are an efficient mechanism to benchmark and rank a population based on a specific set of skills. The notion that they are fair, however, has increasingly been questioned. A significant concern is that performance differences reflect inequities in the testing process itself, rather than differences in underlying skills, and thus may contribute to the intergenerational transmission of existing inequality. Students of the same ability, but with different backgrounds, are known to respond differently to questions, though there is limited understanding as to why (Freedle (2010), Editors (2010)).[1] This paper explores one possible reason.

I investigate whether differential performance may be generated by the frequent use of monetary themed questions on mathematics examinations.[2] Open any first grade mathematics workbook and you will undoubtedly see simple algebra problems centered around the buying and selling of various items. These types of monetized scenarios are frequently used in early mathematical education the world over and are commonly featured on tests. Recent work looking at how poverty impacts cognition has proposed that, for low income individuals, attention can become focused on scarcity and lead to stress and inattention, particularly when choices about money and finances are being considered (Mani et al. (2013), Mullainathan and Shafir (2013)). The use of these questions could bring money to the top-of-mind for low income test takers, distracting them and impacting their performance.

I begin by documenting the differential performance of disadvantaged students on mathematics exams when they feature a larger share of monetary themed questions. Using data from an international cross-country standardized exam, I exploit the natural variation in the financial salience of the mathematics exam that is generated from the random assignment of the examination booklets. I find that a 10 percentage point increase in the

[1] Differential performance by different ethnic and socio-economic groups has been documented on the SAT for instance (Freedle (2003), Santelices and Wilson (2010)). There is also evidence that women are more likely to skip multiple choice format questions and underperform when exams are more competitive (Ors et al. (2013), Jurajda and Münich (2011), Riener and Wagner (2017), Akyol et al. (2016), Ben-Shakhar and Sinai (1991), Pekkarinen (2015), Baldiga (2014)).

[2] I define monetary themed questions as questions that involve topics such as buying, selling, making payments, saving and spending money or calculations using currency. Examples of monetary themed questions for the three datasets are presented in figure 3, and A2, A7, A8 and A4 in the appendix.

financial salience of the exam depresses the performance of students with socio-economic status (SES) indicators below the national median by 0.026 standard deviations.[3] This is a non-negligible effect representing about 6% of the overall performance gap for below median SES students. This effect manifests as early as in the fourth grade and is largest for the most disadvantaged. Using the question-level data, I show that low SES students underperform on monetary questions as compared to their performance on other questions. This pattern is also present in the question-level data of a US homework platform and holds even when comparing performance on highly similar questions that differ only in their monetary theme.

To investigate whether an attention capture mechanism contributes to these performance differences, I look at performance on questions that are randomly positioned after a monetary themed question. In both itemized question-level datasets, consistent with an attention capture effect, I observe a temporary pattern of differential underperformance by low SES students on questions that are placed subsequent to a monetary themed question. Though the effects are temporary, the impacts are economically meaningful since exam scores are used to determine important economic opportunities such as eligibility for further education, placement in schools or access to scholarships. To show this, I take my empirical estimates to real data from a high school entrance exam in Mexico City, and simulate impacts on high school access.

This relationship between poverty and cognition has primarily been observed in experimental settings using psychological tests on adults. By investigating student performance under normal exam and homework conditions and showing that these effects impact exam scores, a policy relevant outcome, these findings address an important gap in the cognitive functioning literature.

The impact of poverty on economic productivity is of central importance to theories of economic development and has generated a large applied microeconomic literature exploring many different explanations and mechanisms (e.g. Gertler et al. (2012), Banerjee et al. (2015), Kim et al. (2006), Schofield (2014), Bandiera et al. (2017)). Understanding how behavioral biases impact this relationship between poverty, inequality and economic productivity is a growing field of research (e.g. Bernheim et al. (2015), Karlan et al. (2014), Bertrand and Morse (2011), Breza et al. (2018), Bartoš et al. (2021)). Within this line of inquiry, the literature on the psychology of poverty has investigated the causal impacts of financial scarcity on stress, affect and cognition and their consequences for the decisions

[3]This pattern is also identified in data from a national educational assessment exam in Mexico discussed in the appendix.

of low income individuals (Kremer et al. (2019), Haushofer and Fehr (2014), Schilbach et al. (2016), Ridley et al. (2020), Haushofer and Shapiro (2016), Haushofer and Shapiro (2018), Mani et al. (2013), Shah et al. (2018a), Carvalho et al. (2016), Shah et al. (2012), Bertrand et al. (2004), Shah et al. (2018b), Zhao and Tomm (2017)). This body of work has focused on adults, though some of the mechanisms proposed could also impact children in the educational environment. The impact of poverty on educational outcomes has been a central topic of much research in economics, and other disciplines in the social sciences (e.g. Bowles et al. (2009), Dobbie and Fryer Jr (2011), Akee et al. (2010), McLoyd (1998)). While behavioral mechanisms have been researched in the educational environment (e.g. Levitt et al. (2016), Lavecchia et al. (2016), Koch et al. (2015), Deckers et al. (2015), Bursztyn and Jensen (2015)), the causal cognitive impacts of poverty have not. School work and examinations are exceptionally cognitively demanding activities. They are frequently used to allocate access to educational and economic opportunities and thus have important policy implications. Because they shape human capital accumulation and productivity, performance on school work and exams is an important mechanism in the intergenerational transmission of socio-economic status. The educational environment is thus a particularly relevant setting in which to evaluate how low SES students are affected by the cognitive effects of poverty.

The attention capture mechanism I investigate draws heavily on the evidence from the psychology of poverty literature regarding the relationship between cognitive functioning and poverty. This literature suggests that poverty captures attention, generates intrusive and distracting thoughts that can reduce an individual's cognitive resources (Mullainathan and Shafir (2013), Mani et al. (2013), Shah et al. (2018a), Carvalho et al. (2016), Kaur et al. (2019)). Several mechanisms have been investigated as outlined in Spears (2011). The limited cognition mechanism posits that economic decisions are more difficult for the poor as they face more difficult trade-offs which deplete their cognitive resources, leaving them with less cognitive control. Thus while scarcity generates a 'focus dividend' leading to a more efficient allocation of scarce resources this comes at the cost of cognitive depletion (Mullainathan and Shafir (2013), Shah et al. (2012)). The limited attention mechanism differs from the limited cognition mechanism in that it does not require a cognitively taxing economic decision. Rather, it simply suggests that, under conditions of poverty, attention becomes focused on scarcity, leading to stress and inattention to other issues. These mechanisms have been tested in a number of lab and field experiments (Mani et al. (2013), Shah et al. (2012), Carvalho et al. (2016), Kaur et al. (2019)). The mechanism I propose, while drawing heavily on the limited attention mechanism, adds the caveat that something must capture attention to activate temporary inattention

and errors. Even if the effects are temporary, they may carry significant consequences. Decisions would be impacted precisely when a low income individual makes cognitively demanding decisions about financial resources, making any such effect important to understand for scholars who study decision making in the context of poverty.

This work is related to the work of Vohs et al. (2006) and others in psychology that have looked at how a number of behaviors are impacted by financial priming (Zhou et al. (2009), Vohs (2015), Lodder et al. (2019), Gasiorowska et al. (2016)). The mechanism, and the setting in which I study it, is also reminiscent of the stereotype threat effect first posited by Steele and Aronson (1995). They suggested that an individual's performance on an examination is sensitive to priming about a stereotype of their group, though to my knowledge, the impacts of financial priming by test questions on performance has not been explored. The stereotype threat hypothesis has generated a significant amount of research, though almost exclusively in lab and field-lab settings (Spencer et al. (2016), Fryer Jr et al. (2008), Rohrer et al. (2015)).[4] Similarly, empirical challenges have also led most existing research on poverty's effects on cognitive functioning to be done in lab or lab-in-the field settings with impacts measured using laboratory measures of cognition.[5] Confirming that behavioral mechanisms impact economic decisions in the field is important for validating experimental results (Levitt and List (2007), DellaVigna (2009)). By utilizing real examination data, I address this gap in the literature and alleviate concerns of experimenter demand effects and sensitivity to specifically designed wording of priming statements that may not be reflective of typical examination conditions. By using secondary sources for my examination and homework data, I show that the experimental results on the cognitive effects of poverty impact a policy relevant outcome and have external validity beyond the experimental setting.

The rest of the paper is organized as follows: Section 2 describes the primary datasets used in my analysis. Section 3 presents estimation methods and results on exam performance, using exam-level variation in financial salience. Section 4 identifies the attention capture effect. First, I show that low SES students underperform on the monetary questions and on questions positioned subsequent to the monetary questions in the TIMSS exams. Second, I confirm that the same pattern is present in the data from the ASSISTments homework platform. Low SES students underperform on ASSISTments monetary

[4] A few researchers have experimented with placement of demographic questions around actual AP exams (Stricker and Ward (2004), Danaher and Crandall (2008)), while Wei (2012) exploits natural variation in pretest background questions to detect a stereotype reactance effect in the NAEP math test.

[5] Kaur et al. (2019) is the exception here as the impacts of their field experiment are measured on worker productivity, a particularly important policy variable.

questions, even when compared to highly similar non-monetary themed questions. They also underperform on non-monetary questions when they are randomly positioned after a monetary themed question. Finally, the contribution of a cognitive fatigue mechanism is considered. Section 5 discusses implications for high stakes examinations and simulates the effects on exam performance and high school placement using data from a high school entrance exam in Mexico City. Finally, section 6 concludes.

2 Data

Evidence that the share of monetary questions featured on an exam or assignment differentially affects the performance of low socio-economic status (SES) students is presented for the TIMSS international cross-country standardized tests and the US based ASSISTments homework platform. Evidence from Mexico's ENLACE exams is presented in appendix A2. In addition to confirming the replicability of this result, each of these datasets has distinct attributes allowing for a more thorough understanding of the mechanisms behind the general result. Table 1 summarizes the key attributes of each of these datasets.

2.1 Cross-Country Examination Data: TIMSS

The Trends in International Mathematics and Science Study (TIMSS) is an international standardized test in math and science administered to a random sample of 4th and 8th graders in participating countries.[6] These examinations have been taking place every 4 years since 1995. The TIMSS tests are one of the main sources reported by the World Bank for international learning outcomes data.

The TIMSS data features question-level responses and, since 2011, most countries also administered a parental questionnaire for the 4th grade exam which reports basic occupational and educational categories of the parents. Most of these countries are middle to high income.[7] I use the highest reported parental education category as my primary SES indicator.[8] I also generate an indicator variable for whether the highest parental ed-

[6]Sampling follows a stratified two-stage cluster sample design. First, a probability weighted stratified random sample of schools is selected and then a random sample of classes is selected from within each school. This procedure generally results in the selection of approximately 150 schools and 4000 students per country.

[7]For brevity I use the term 'country', though exam administrators also work with regional authorities that wish to benchmark their performance. For the 4th grade exams, 53 countries participated in 2015 and 60 in 2011, though parental questionnaires were only administered in 50 and 37 respectively. The full list of countries is available in appendix figure A1.

[8]The TIMSS surveys categorize parents into 5 different educational categories: parents who have completed university, who have some post-secondary education, who have completed their secondary educa-

ucation reported for a student falls below the national median as observed in the TIMSS data.[9] As SES may be associated with very different education levels in different national contexts, estimates using this indicator will be more reflective of relative SES within the national context. My main sample consists of the 379,468 students for whom parental education is available.[10]

Each year, a student taking a TIMSS exam is randomly assigned one of 14 possible booklets. The same booklets and questions are used in all schools. Each booklet consists of three components: a mathematics and a science section, followed by the student survey, all of which are separated by short breaks. Question order within a booklet does not vary. For my estimations I focus exclusively on the mathematics section of the exams.[11]

Information is available on each prompt, including prompt characteristics such as the answer type (completed response or multiple choice), topic area and cognitive domain and a brief thematic descriptor.[12] I flag as 'monetary' any question whose prompt or prompt descriptor contains terms such as 'money', 'buy', 'sell', 'cost', 'pay' or 'zeds' (the fictional currency used for this international exam).[13]

Pooling the 2015 and 2011 4th grade data gives me 28 different exam booklets. On average students face 25.32 different math questions, making each question worth approximately 4% of the math exam score, which I calculate as simply a student's mean performance on all of the mathematics questions in their booklet.[14] Out of 708 questions,

tion, who have some lower secondary education and who have a primary education or less. Occupational categories are more difficult to compare and interpret given the cross-country nature of this data. Results using highest family occupational category are similar.

[9] As the national median in Honduras is for parents to have primary or no education, I set this indicator to one for Honduran students who are at the national median in order to have a comparison group.

[10] 445,342 students have taken the examination and were asked to complete the parental survey over the two rounds of 2011 and 2015. There is selection into the main sample due to non-random parental non-response which is discussed in appendix A1.

[11] Each mathematics section consists of two blocks of prompts that permutate throughout the 14 booklets so that each block of prompts is featured in two different booklets. For clarity throughout the remainder of this paper, I use the term 'prompt' to refer to a unique query, while 'question' will refer to a prompt in a specific booklet and year. To measure time trends in learning outcomes, eight blocks of mathematics questions get readministered between 2011 and 2015. Thus, a unique prompt is either featured in two questions if in a non-readministered block or four questions if in a readministered block. Table A2 in the appendix illustrates the structure of the 14 TIMSS booklets that could be handed to a student in a given year.

[12] TIMSS readministers prompts across examination waves and thus does not release all examination prompts. TIMSS exams feature both multiple choice and completed response questions. Most of the questions only allow for a single correct answer, but occasionally multiple answers are considered correct and some questions allow for partially correct answers. As a quantitative measure of 'correctness' is not given I do not count partially correct answers as correct.

[13] An example of a monetary themed question from a TIMSS exam is provided in figure A2 of the appendix.

[14] TIMSS exams are designed to measure the distribution of proficiency in a population rather than accurately measure the proficiency of a single individual, thus the exam mean differs from the official TIMSS

44 are flagged as monetary questions and feature 14 unique prompts. For identification, I exploit the random assignment of booklets to students as there is variation in the proportion of monetary questions featured in the booklets that, through random assignment, is uncorrelated to parental education categories.[15]

2.2 Homework Platform Data: ASSISTments

ASSISTments is a free online homework platform in the US.[16] Teachers use the platform to assign problem sets to their students. Problem sets consist of a large pool of questions meant to practice a specific skill. When assigned a problem set, students respond to questions until they answer three in a row correctly. Several hints are attached to each question; the students can consult the hints and can make several attempts at answering each question. Importantly, there is no set order to the questions a student will face, as questions are randomly drawn without replacement from the question pool of the assigned problem set.

Student-level user data includes the sequence of questions a student faced, the amount of time spent on each question, the number of attempts made, the number of hints requested, and whether they completed the problem set by answering three questions in a row correctly.[17] Upon request, ASSISTments generated SES indicators by matching participating schools to National Center for Educational Statistics (NCES) giving me school enrollment and enrollment in free and reduced price lunch programs. With this I construct a school-level SES indicator, the share of students in a school enrolled in free or reduced lunch programs. Individual SES indicators are not available for this dataset.

The main sample consists of 17,428 different student submissions for 577 assignments covering eleven different problem sets that include both monetary and non-monetary questions. These feature 1448 questions, of which 450 are coded as monetary themed.[18]

achievement measure, which is generated using a complex parameterized imputation procedure.

[15]Panel a of figure A3 shows the variation in the proportion of monetary questions featured in the booklets as well as variation in the reported parental education categories. Table A3 in the appendix confirms that booklets are randomly distributed to students and that there is no correlation between the SES indicator and the share of monetary questions a student faces on their exam.

[16]ASSISTments is operated by the Worcester Polytechnic Institute's Computer Science Department. Though ASSISTments is not a widely used homework aid, it is partially funded by the NSF as a research platform and assignment data is available for research purposes.

[17]I observe students once they log on to the platform and do not observe students who might fail to log on. Any selection into platform use due to problems with computer access likely makes these estimates conservative.

[18]Data was requested for problem sets that featured both monetary and non-monetary themed questions. Problem sets that included multiple part questions were then dropped. To focus on students who are actually engaged in completing the assignment and haven't walked away from the computer, time spent on a question is coded as NA if the student spends more than 15 minutes (about the 93rd percentile) on a

For identification, I exploit the within problem set random assignment of questions to students which is uncorrelated with the share of schoolmates receiving free or reduced price lunch.[19]

The ASSISTments data provides insight into the learning process. Furthermore, the ASSISTments data has the distinct advantage of featuring both a large number of questions and random question ordering. These attributes are key to understanding the mechanisms behind the general result, which will be the focus of section 4.

3 Impacts of Financial Salience on Performance

I begin by using the TIMSS data to present student-level estimations that look at how the variation in the proportion of monetary questions featured on an exam impacts exam-level performance. I find that more financially salient exams differentially depress the exam scores of lower SES students.

This approach using the TIMSS data exploits the random assignment of test booklets to students and the variation in the number of monetary questions between booklets. The effect on a low SES student of receiving a financially salient booklet is estimated as follows,

$$E_{ib} = \Theta_1 + \Theta_2 LowP_i + \Theta_3 LowP_i * PM_b + \kappa_b + c_c + \epsilon_{bi}, \tag{1}$$

$$E_{ib} = \theta_1 + \sum_{p=2}^{5} \theta_{2p} P_i + \sum_{p=2}^{5} \theta_{3p} P_i * PM_b + \kappa_b + c_c + \epsilon_{bi}. \tag{2}$$

question. To omit random guessing behavior, time is coded as NA if a student spends less than 5 seconds on a question (1% of observations). To focus on 'active' question-level observations, all TIMSS regressions condition on time being between 5 seconds and 15 minutes. Similarly, if attempt counts are above 5 attempts (the 95th percentile) attempts are coded as NA. The sample was then limited to submissions for which the SES indicator is observed for the student and where the monetary indicator is defined for all questions. Finally, assignments with fewer than 5 student submissions are dropped.

Though ASSISTments suggests this is rarely exercised, teachers have the option of fixing the ordering of questions, or the ordering of question types, on an assignment, which is not directly observable in the data. To be certain that I use only randomly ordered assignments, I conducted two assignment-level tests. I calculate the share of monetary questions that appear in the first three questions of student submissions and compare this to the expected value based on the share of monetary questions in the problem set, dropping any assignments with a p-value <0.1 (175 of 809 assignments). I also check that monetary questions are not correlated with a particular sequential positioning and drop assignments where this correlation is significant at the 10% level (57 additional assignments).

[19] Panel b of figure A3 shows that there is significant variation in the proportion of monetary questions featured on student submissions and that there is also substantial variation in the share of schoolmates receiving free or reduced price lunch, the two key sources of variation I exploit in this dataset. Table A4 in the appendix verifies that conditional on the assigned problem set, the proportion of monetary question a student faces on their entire assignment and the number of monetary questions they face in the first three questions of their assignment is random and uncorrelated with their SES indicator.

Each observation represents student (i) 's performance on their examination booklet (b). I regress standardized exam scores (E_{ib}) on the SES indicator and the interaction between the SES indicator and the proportion of monetary themed questions (PM_b) featured in the randomly assigned booklet.[20] In equation 1, $(LowP_i)$ indicates if the reported parental education category is lower than the national median as observed in the TIMSS data. For equation 2, (P_i) are parental education category dummies. I also include booklet fixed effects (κ_b) and country or class fixed effects (c_c) as controls.[21]

Results are reported in table 2. Estimates of Θ_3 in the sixth row of columns 1 and 3 imply that a 10 percentage point increase in the share of monetary questions featured on an exam differentially depresses the performance of students whose parental education falls below the national median, by 0.026 standard deviations. Note that on the TIMSS exams the proportion of monetary questions featured in a booklet ranges from 0 to 0.217. Columns 2 and 4 show that this effect is negatively related to parental education, with the largest effect for the most disadvantaged students. Columns 3 and 4 include classroom fixed effects.[22]

The magnitude of the effect of monetary questions is not small. It is informative to compare this effect to the general performance gap between these students as measured by the Θ_2 coefficients. The 0.026 standard deviation decrease resulting from a 10 percentage point increase in the proportion of monetary questions is equivalent to about 6% of the within-country performance differential between students whose parental education is at or above the national median and those below. This increases to about 10% when considering the within classroom performance differential.

In appendix A2 I use a similar approach to generate estimated effects for Mexico's ENLACE exams. The results and magnitude of the effect presented in table A7 are consistent with the results presented above.

4 Identifying Attention Capture

The evidence above shows that financial salience differentially impacts the exam performance of low SES students. An attention capture effect has not yet clearly been identified. Monetary questions may differ from other mathematics questions. They may be used to

[20] Exam scores are calculated as the mean performance on the questions in the question-level data. These are then standardized to generate the z-score for each observation.

[21] Since students are randomly assigned to exams, I do not cluster these standard errors. The main results are robust to clustering at the booklet level as illustrated in appendix table A5.

[22] Results using highest family occupational category are similar and reported in table A6 of the appendix.

test different skills in which low SES students face a disadvantage. Or, the effects may be entirely driven by the fact that monetary questions are more difficult for low SES students as they may have fewer opportunities to engage in monetary transactions. For instance, these students may be less likely to receive an allowance with which they can make purchases in a shop where they must collect change. There is evidence that such applied experience can help children conceptualize math problems. When focusing on a highly selected sample of low income children working as shopkeepers in informal Indian markets, and thus with substantial experience in monetized transactions, Banerjee et al. (2017) find evidence of improved performance on algebra questions that are framed monetarily as compared to their abstract equivalent. In my context it is not clear whether low SES students would have more or less experience applying mathematical skills to monetary transactions. Nonetheless, it is possible that this experience mechanism could explain some of the estimated effect presented in table 2. An attention capture mechanism cannot be disentangled from these other mechanisms by looking only at how and exam's financial salience affects exam performance.

To clearly identify an attention capture effect, I use the question-level data available in both the TIMSS and ASSISTments datasets. If non-negligible, an attention capture effect would impact performance on both the monetary questions and on questions that happen to be positioned subsequent to a monetary question. These subsequent questions are not monetary themed but are potentially affected by attention capture due to their position after a financially salient question. This approach allows the identification of an attention capture effect, showing that the cognitive effects of poverty impact performance on examinations.

4.1 Attention Capture in the TIMSS data

Using the 19 TIMSS booklets that have a single monetary 'event', and thus clearly defined pre and post periods it is possible to estimate event study coefficients to see how the performance of low SES students is impacted on the monetary and subsequent questions. Figure 1 plots estimates for coefficients π_{2t} from the following estimating equation:

$$C_{iq} = \pi_1 + \sum_{t=-6, t\neq -1}^{10} \pi_{2t}(T_q = t) * LowP_i + \mu_q + \eta_i + X_{pq} + \epsilon_{iq}. \tag{3}$$

Each observation is a student (i)'s performance on question (q). (C_{iq}) is an indicator of a correct response. (T_q) is a question's position relative to the monetary question which is interacted with $(LowP_i)$, an indicator of parental education being below the national

median. I include student (η_i) and question (μ_q) fixed effects to control for student and question unobservables.[23] Finally, (X_{pq}) is a vector of fixed effects designed to capture differential performance by parental education categories due to question observables.[24]

Figure 1 shows the differential underperformance of students with parental education indicators that fall below the national median on monetary questions and on the first few questions that follow monetary questions.[25] The effects are temporary, concentrated in the first four questions after the monetary question. Nevertheless, though temporary, figure 1 suggests that the effect on subsequent questions could account for a substantial share of the overall effect of monetary questions.

This approach can be generalized to all of the TIMSS exam data, including exams featuring multiple monetary question, using the following specifications:

$$C_{iq} = \Lambda_1 + \Lambda_2 LowP_i * M_q + \Lambda_3 LowP_i * Post_q + \mu_q + \eta_i + X_{pq} + \epsilon_{qi} \tag{4}$$

$$C_{iq} = \lambda_1 + \sum_{p=2}^{5} \lambda_{2p} P_i * M_q + \sum_{p=2}^{5} \lambda_{3p} P_i * Post_q + \mu_q + \eta_i + X_{pq} + \epsilon_{qi}. \tag{5}$$

Each observation is a student (i)'s performance on question (q). I regress the indicator for a correct response (C_{iq}) on the interaction between the SES indicator (having a parental education level below the national median ($LowP_i$) or parental education category dummies (P_i)) and the monetary indicator (M_q) as well as the post-monetary indicator $(Post_q)$ for the four questions directly subsequent a monetary question.[26] (μ_q), the question fixed effect, and (η_i), the individual fixed effect, are included. Depending on the specification, the (X_{pq}) vector of fixed effects is included as well as fixed effects to control for class performance on the monetary and subsequent questions.[27]

[23]Note that because the sequence of TIMSS questions is fixed within a booklet, question fixed effects directly capture the effect of placement within a booklet.

[24]TIMSS disseminates some information about each prompt on the exam. Because placement is not randomized in the TIMSS data, fixed effects designed to capture differential performance due to question observables such as sequence position, difficulty, question type and topic are included in most question-level TIMSS estimations. These include parental education by sequence position fixed effects, where sequence position is a constructed categorical variable indicating if a question is featured in the first five questions of the exam, second five and so forth. Parental education by difficulty fixed effects, where difficulty is a constructed categorical variable that uses the mean performance on a question by students with university educated parents to categorize questions into 20 difficulty bins. Parental education by country by topic and parental education by country by question type address the possibility that certain education systems may differentially prepare students in different mathematical topics or use different testing methods.

[25]Students tend to proceed consecutively through their booklet though it is possible for students to skip around between questions. Non-consecutive response patterns most likely dampen the estimated effects of positioning.

[26]The lag effect is tracked on four subsequent questions as results presented in figure 1 show that this is where the effect is concentrated.

[27]Appendix table A8 estimates equation 4 with different combinations of the (X_{pq}) fixed effects. Though

Results are reported in table 3. Columns 1 and 2 estimate equations 4 and 5 using only the student and question fixed effects. Columns 3 and 4 include the (X_{pq}) fixed effects and columns 5 and 6 add fixed effects to control for class performance on the monetary and subsequent questions. Students with above median educated parents (columns 1, 3 and 5) or with university educated parents (columns 2, 4 and 6) are the omitted categories. In all specifications, lower SES students' performance is differentially depressed on monetary questions and the questions that follow them. The largest effects are found in the most disadvantaged students. Columns 5 and 6 show that this pattern is robust to controlling for class performance on these particular questions. One may be concerned that the effect stems from differences in teaching patterns between more advantaged and disadvantaged classrooms. The results in columns 5 and 6 show that parental education still has explanatory power for performance on these questions, even when controlling for class performance, though the magnitude of the effect is reduced.

My preferred specifications are those in columns 3 and 4. Students with below median parental education indicators are 1.2 percentage points (0.05 sd) less likely to answer the monetary question correctly and 0.89 percentage points (0.04 sd) less likely to answer the four questions that follow them correctly.[28] Aggregated, these estimates suggest that each additional monetary question featured on a 25 question exam would depress the score of a student whose parents have an education level below the national median by 0.191 percentage points or about 0.008 standard deviations compared to students whose parents have an education level above the national median.[29] Since each question accounts for 4% of the mathematics test, a 10 percentage point increase in the share of monetary questions (equivalent to a 2.5 question increase) would depress the performance of low SES students by 0.02 standard deviations using the estimates in table 3, only slightly less than the 0.026 effect estimated in section 3. I can thus decompose the exam-level results from section 3 into the direct effect of monetary questions explaining 19% of the 0.026 standard deviation exam-level estimate and the indirect effect on subsequent questions explaining 58% of the examination level estimate.[30]

the magnitude of the estimated effects is somewhat sensitive to the choice of fixed effects, the effect on both monetary and subsequent questions remains negative and statistically significant in all estimations.

[28]Appendix table A9 investigates whether unanswered questions, which are coded as incorrect in the estimations above, could be driving the effects. Interestingly, low SES students seem to be slightly less likely to leave a monetary question, or subsequent questions, unanswered.

[29]The standard deviation of the exam scores is 23.55 and each question is worth approximately 4% of the exam. Using the estimates from column 3, I calculate the direct effect as $-1.207 * 0.04 = -0.048$ points or -0.002 standard deviations with an additional effect on four subsequent questions of $-0.891 * 0.04 * 4 = -0.143$ points or -0.006 standard deviations for a total of a 0.191 percentage point decrease on the exam or about 0.008 standard deviations.

[30]The direct effect is calculated as the 0.002 standard deviations per question multiplied by 2.5 questions, or 0.005 standard deviations of the 0.026 estimated effect. The indirect effect is calculated as the 0.006

The results in this section show that low SES students underperform on the monetary questions and on the questions that are positioned directly subsequent to the monetary questions. The specifications in this section control for differential performance that may be generated by the question observables that are reported in the TIMSS data. It is possible that the difference in performance on the monetary questions (Λ_2) is driven by unobservable question attributes that correlate with monetary themes, as these questions do differ from other exam question.[31] For the performance differential on the subsequent questions (Λ_3), there is no indication that these questions have observable or unobservable characteristics that systematically differ from the other exam questions. Nevertheless, I cannot completely rule out this possibility in the TIMSS data. For this I turn to the ASSISTments data where I observe similar results in an environment where such confounds are not present.

4.2 Attention Capture in the ASSISTments data

Like the TIMSS data, the ASSISTments homework platform data also features itemized question-level data. I begin with a simple comparison of how question-level performance (responding correctly) and effort metrics (time spent, hints requested and attempts made) vary by the percent of students on free and reduced price lunch in a school for monetary and non-monetary questions. These comparisons are illustrated in column 1 of figure 2. These plots suggest that the monetary questions in the ASSISTments problem sets are generally easier for students. However, this effect is smaller for students in disadvantaged schools. In schools where few students receive free or reduced price lunch, students are more likely to answer monetary questions correctly and request fewer hints, make fewer attempts, and spend less time on these questions. For students in schools where most students receive free or reduced price lunches, the advantages presented by monetary questions are much smaller if not nonexistent. To better compare these differences in performance, I estimate

$$Y_{iq} = \delta_1 + \delta_2 PFR_s * M_q + \delta_3 M_q + \delta_4 Seq_{iq} + sp_{sp} + \epsilon_{iq}. \tag{6}$$

Each observation is a student (i)'s performance on question (q). I regress question-

standard deviations per question multiplied by 2.5 questions, or 0.015 standard deviations of the 0.026 estimated effect.

[31]Comparisons of observable prompt characteristics across the question indicators of interest are illustrated in appendix figure A6. Monetary questions do differ in topical content, in that they are never used to test geometry topics, which represent almost half of the other questions. The four questions that follow monetary questions have a distribution of question characteristics that is similar to the other questions in the TIMSS booklets.

level performance metrics (Y_{iq}) on the interaction between the proportion of students in the school receiving free or reduced price lunch (PFR_s) and an indicator for monetary themed questions (M_q). A control for the sequence position, (Seq_{iq}), which captures how student performance changes as they proceed through the assignment is also included. Finally, I include a school by problem set fixed effect (sp_{sp}) to control for how students in a school (s) perform on questions in problem set (p). Question-level performance metrics include whether the student answered the question correctly, how many hints were requested, how many attempts were made, and the time spent on the question.[32]

Results in table 4 are similar to the patterns displayed in column 1 of figure 2, though the added controls change these dynamics slightly. The mean value of the proportion of students in the school receiving free or reduced price lunch is 0.27 while the minimum value is 0.01 (close to $PFR_s = 0$) and the maximum is 0.93 (close to $PFR_s = 1$). Students in the most advantaged schools, where few students receive free or reduced school lunches $(PFR_s = 0)$, do indeed find these questions easier. Estimates for δ_3 reported in the second row show they are 6 percentage points (+0.13 sd) more likely to answer them correctly, request 0.22 (-0.17 sd) fewer hints and make 0.1 (-0.1 sd) fewer attempts. With the added controls, it becomes clear that the advantage these questions present to students in the most advantaged schools is not present in the most disadvantaged schools where most students are in the free or reduced lunch program $(PFR_s \approx 1)$. Summing δ_2 and δ_3 in rows 1 and 2 shows that these students request the same number of hints and make 0.05 more attempts on these questions (+0.05 sd) while spending 23 more seconds (+0.18 sd) on them. Despite the fact that they are expending more time on monetary questions they are 4 percentage points (-0.09 sd) less likely to answer these questions correctly as compared to schoolmates answering non-monetary themed questions in the same problem set.

An important concern may be that monetary questions are used to test a very different set of mathematical skills in which low SES students may face a disadvantage. For instance, these questions may be more likely to test numerical operations rather than geometric reasoning.This was addressed by the addition of the X_{pq} vector of controls in the analysis of the TIMSS data. On the ASSISTments platform, problem sets are very narrow in thematic content, as teachers use them to practice very specific mathematical skills such as 'Finding the Whole from the Percent and Part in a Word Problem' or 'Percent Increases and Decreases', thus the school by problem set fixed effect, sp_{sp}, should capture much of the differences in performance due to the mathematical skill being practiced. Neverthe-

[32] ASSISTments does occasionally report quantified partial credit; however, most of the data is either a 0 or 100.

less, one may still be concerned that monetary questions require a different skill set. These questions may involve more reading than, for instance, algebraic formula problems. The large question pools used for ASSISTments problem sets allow me to address this concern. Within each problem set I match monetary themed questions to almost identical non-monetary questions. Questions are matched if they are formulated similarly, require similar reading and vocabulary skills, and involve the application of the same mathematical process.[33] Figure 3 shows two examples of matched monetary and non-monetary questions. Additional examples of matched questions are presented in appendix figures A7 and A8.

Column 2 of figure 2 shows how, for monetary and non-monetary questions, question-level performance metrics and effort metrics vary by the percent of students on free and reduced price lunch in a school for the matched sub-sample of questions.[34] Column 2 of figure 2 shows that performance on these matched questions is very similar for students in the most advantaged schools. Monetary and non-monetary themed questions are about equally likely to be answered correctly and require about the same number of hints and attempts, though the monetary questions do appear to take a little longer. For students in the most disadvantaged schools, the differences are much more substantial. They are much less likely to answer the monetary questions correctly and require more hints and attempts, and differentially more time.

To estimate the difference between matched monetary and non-monetary questions with additional controls, I estimate the following,

$$Y_{iq} = \gamma_1 + \gamma_2 PFR_s * M_q + \gamma_3 M_q + \gamma_4 Seq_{iq} + ms_{ms} + \epsilon_{iq}. \tag{7}$$

Each observation is a student (i)'s performance on question (q). I regress question-level performance metrics (Y_{iq}) on the interaction between the proportion of students in the school receiving free or reduced price lunch (PFR_s) and an indicator for monetary themed questions (M_q) controlling for the monetary indicator and sequence position (Seq_{iq}). I include a matched question group by school fixed effect (ms_{ms}). This fixed effect is important as it restricts my variation so that I am comparing student performance within a school on questions that are nearly identical except for their monetary themed

[33]Question matching is done manually. ASSISTments problem sets are composed of several hundred questions. Of these questions, many share a similar template using similar wording but different numerical values. Each question shown in figure 3 and in the appendices is representative of several other questions that share the same question template. Some of the templates are monetary themed and are very similar to non-monetary themed templates as illustrated in the examples. In such a case these templates are matched creating a group of matched questions. Eleven matched question groups are identified.

[34]Note that restricting the data to matched questions significantly reduces the sample size from 91,092 to 23,771 active question observations.

content.

Results are reported in table 5 and reflect the pattern observed in column 2 of figure 2. Estimates for γ_3 in the second row show only a difference of 15 additional seconds of time spent (+0.13 sd) on monetary questions in the most advantaged schools, consistent with column 2 of figure 2. Summing the γ_2 and γ_3 coefficients in rows 1 and 2 for the different effort metrics, shows underperformance and increased effort on monetary questions for lower SES students, as compared to the performance of their schoolmates answering highly similar non-monetary themed questions. In a school where all students received free or reduced lunch, students are 5 percentage points less likely to answer the question correctly (-0.13 sd), request 0.29 more hints (+0.28 sd) make 0.1 more attempts (+0.12 sd) and require 54 more seconds (+0.46 sd) to answer the question as compared to classmates responding to a similar non-monetary themed question. This is evidence that underperformance on monetary questions by low SES students cannot be fully explained by monetary questions requiring a different set of mathematical or question answering skills beyond those implied by their topical content. The increased effort exerted by the low SES students on these questions is consistent with these problems capturing student attention as discussed in the scarcity literature. Here this generates a distraction from the task at hand as response accuracy decreases.[35]

The evidence presented in table 5 shows that monetary themed questions present a greater challenge to students in disadvantaged schools, even when these questions are formulated almost identically to non-monetary questions. Differences in mathematical skill cannot explain these results. It is possible however, that differences in experience applying math to money could explain some of the estimated effect presented in table 5. As discussed in section 4.1, an attention capture mechanism cannot be disentangled from this explanation by looking only at performance on the monetized questions. An attention capture effect can be isolated by looking at performance on subsequent questions. Furthermore, because the ordering of questions on the ASSISTments platform is randomized, the concerns in the TIMSS data that question placement might correlate with question unobservables are alleviated.

I leverage this randomized ordering to find evidence of an attention capture effect by comparing the performance of students in the same school on a question when it is placed subsequent to a monetary question versus when it is placed after a matched non-

[35]This is in contrast to a 'focus dividend' (Mullainathan and Shafir (2013)) which leads to a more effective use of scarce resources. The focus dividend is a distinct mechanism, as it governs the use of scarce resources. In this context, this fictional money is not actually being used or allocated by the students. Rather, the monetary theme is acting to distract and capture cognitive bandwidth, a mechanism more closely related to those explored in Mani et al. (2013), Carvalho et al. (2016) or Bartoš et al. (2021).

monetary question. I estimate the following,

$$Y_{iq} = \beta_1 + \beta_2 PFR_s * Post_{iq} + \beta_3 Post_{iq} + \beta_4 Seq_{iq} + \beta_5 Seq_{iq} * PFR_s + m_{m_{pre}} + \nu_{qs} + \epsilon_{iq}. \tag{8}$$

Each observation is a student (i)'s performance on question (q). I regress question-level performance metrics (Y_{iq}) on the interaction between the proportion of students in the school receiving free or reduced price lunch (PFR_s) and an indicator for being placed subsequent to a monetary themed question $(Post_{iq})$. I include controls for the question sequence position (Seq_{iq}) and the differential effect of question sequence position ($Seq_{iq} * PFR_s$). Finally, a fixed effect for the leading matched group of questions $(m_{m_{pre}})$ and a question by school fixed effect (ν_{qs}) are also included. These fixed effects allow me to compare the performance of students in the same school on the same question when it is placed after a monetary themed question or a very similar non-monetary question. To avoid having to consider the effects of repeated exposures, I limit my sample to questions that are positioned between the first and second matched question a student encounters and no more than 4 questions after the first matched question.[36] Results are reported in the first 4 columns of table 6. To address any selection concerns, as better performing students respond to fewer questions, columns 5-8 estimate equation 8 using only the first three assignment questions.[37]

In more advantaged schools, students experience reduced effort and better performance on questions subsequent to monetary themed questions suggesting that these questions are particularly effective learning tools in the high SES schools. Estimates of β_3 in the second row of of columns 1-4 show that for students with $PFR_s = 0$, a question positioned after a similar monetary question is 7.7 percentage points more likely to be answered correctly (+0.16 sd), requires 0.22 fewer hints (-0.18 sd), 0.11 fewer attempts (-0.1 sd) and 11 seconds less (-0.09 sd). All of these coefficients are statistically significant. Estimates in columns 5-8 are similar and statistically significant with the exception of the coefficient on attempts which becomes smaller in magnitude and statistically insignificant.

This pattern stands in sharp contrast to the effect of monetary questions on students in the more disadvantaged schools, as reflected by the β_2 coefficients in row 1 which more than offset the effects described above. The β_2 estimates suggest a pattern of differen-

[36] More formally, let Seq_{qi} be the position of question q in student i's sequence of questions. Seq_{M1i} and Seq_{M2i} are the positions of the first and second matched questions faced by student i. I subset the data to observations where $Seq_{M1i} < Seq_{qi} < Seq_{M2i}$ and $Seq_{qi} \leq (Seq_{M1i} + 4)$. Four questions are considered as the estimates presented in figure 1 suggest this is where effects are concentrated.

[37] All students must answer a minimum of three questions, since three questions must be answered correctly consecutively in order to complete and assignment.

tial underperformance by low SES students on questions that follow monetary themed questions when compared to their high SES peers, though the results are underpowered as only the estimate for time is statistically significant in both samples. When compared to schoolmates answering the same question, there is suggestive evidence that students in lower SES schools are less likely to answer a question correctly, require more hints and attempts and spend more time on a question when it is randomly positioned after a monetary themed question versus after a similar non-monetary themed one. In a disadvantaged school where $PFR_s = 1$, the sum of β_2 and β_3 in rows 1 and 2 of columns 1-4 show that a question positioned after a similar monetary question is 6 percentage points less likely to be answered correctly (-0.12 sd), requires 0.236 more hints (+0.19 sd), 0.125 more attempts (+0.12 sd) and 48 seconds more (+0.37 sd), though only the difference in time is statistically significant. Estimates in columns 5-8 are similar with the exception of the β_2 and β_3 estimates for attempts which both shrink in magnitude. Though underpowered, these results are suggestive of an attention capture of poverty impacting the performance of low SES students on questions that they consider shortly after a monetary themed question.

Because of the randomized question order in the ASSISTments data, any lagged effect of a monetary question on subsequent questions must be due to their positioning relative to a monetary question. In addition to an attention capture explanation, a subtly different cognitive fatigue mechanism could explain this pattern. Since low SES students find monetary questions differentially difficult, possibly due to experience differences, this might affect their performance on subsequent questions as they are differentially fatigued when they face them. It is worth emphasizing that the distinction between these mechanisms is subtle as first, they both carry the same implications for exam performance and second, such a cognitive fatigue mechanism is conceptually close to some of the limited cognition mechanisms discussed in the psychology of poverty literature that posit that economic decisions may be more difficult for the poor and thereby deplete their cognitive resources.[38]

To asses the impact of cognitive fatigue effects, I create an SES group specific measure of lagged difficulty. I use the SES indicator to categorize students into quartiles and then calculate the mean time spent by students in each quartile on the preceding matched questions. This measure of differential difficulty of the preceding question is added as a

[38]Though similar, the reasoning behind the limited cognition mechanism hinges on the poor facing more difficult economic trade-offs. Given the fictional, and often third person nature of the math questions, students are not explicitly being asked to consider difficult economic trade-offs. In this context, differential cognitive fatigue generated by the monetary questions is more likely due to differences in experience and familiarity with monetary applications.

control to the estimation strategy used in equation 8.

Table 7 presents the results of this estimation. Controlling for the differential difficulty of the leading matched question does not meaningfully alter the coefficients or significance levels of the coefficients of interest, pointing to an attention capture mechanism as the driver behind these results. A similar approach is applied to the TIMSS data in appendix A3 and an approach using placebo estimates is also presented and applied. Using the placebo approach suggests that cognitive fatigue may explain up to 34% of the effect on subsequent questions in the TIMSS data. Overall, the sum of the evidence points towards attention capture though some impact due to cognitive fatigue cannot be ruled out completely.

5 Implications for High Stakes Exams

Performance differences on high stakes entrance exams can significantly affect access to secondary and higher education and thus to economic opportunities. If questions on high stakes examinations put vulnerable students at a disadvantage, these questions could aggravate socio-economic disparities in access to education and contribute to the intergenerational transmission of inequality. As monetary questions are regularly featured on high stakes exams, the effects identified in this paper have the potential to significantly impact the educational opportunities of low SES students.

Using available information about high stakes tests, I am able to project my estimates onto high stakes exam scores and simulate the potential impact on access to further education.[39] The scholastic assessment test (SAT) is an important component of student applications to universities in the United States. A survey of current official practice exams suggests that monetary questions are regularly featured on the exam and can account for up to 20% of the questions on the quantitative section of the SAT.[40] Using the estimates in table 2 and official assessment statistics, a 20 percentage point reduction in the share of monetary questions on the SAT could improve expected performance by students with below median parental education levels by 6 points (0.052 standard deviations).[41]

[39] I cannot identify effects using a high stakes exam for two reasons. First, most high stakes entrance exams use only one examination booklet per examination wave, making it difficult to control for contemporaneous shocks that might differentially affect different socio-economic groups. Secondly, administered booklets and itemized question data are not generally available.

[40] Ten official practice tests for the SAT were accessed on the college board website in September 2019. Monetary questions on these practice tests ranged from 8.6% to 20% of the questions on the quantitative portion of the exam with a median of 13.8%.

[41] The SAT's 2018 Annual Report shows a standard deviation of 114 points on the quantitative section. This report also shows performance by parental education categories that can be used to determine that the median level of parental education of test takers was a bachelor's degree. The mean quantitative score

This represents about 7.2% of the quantitative section's performance gap between these groups.

In the US, SAT scores are generally only one of many components in a complex admission process. It is thus difficult to anticipate exactly how a change in score would affect access to higher education beyond the prediction that it would make access more equitable. Globally, though, there are many high stakes exams where scores are the sole determinant of eligibility for further education. In the following section, I use the TIMSS estimates to generate counterfactual exam scores on Mexico City's high school entrance exam and use these counterfactual scores to simulate how the change in exam scores would affect student allocation.

5.1 Simulating Effects on High School Placement in Mexico City

A consortium of public schools in Mexico City known as the Comisión Metropolitana de Instituciones Públicas de Educación Media Superior (COMIPEMS) uses competitive centralized admissions. All ninth graders wishing to attend one of these schools submit a ranked list of up to 20 high school programs and then take a comprehensive standardized exam. After exams have been scored, students are ranked and assigned to schools according to a serial dictatorship mechanism (see Abdulkadiroglu and Sonmez (2003)).

I consider how a 10 percentage point decrease in the share of monetary questions would change the allocation of students to high schools in Mexico city.[42] I use data from the 2004 and 2005 COMIPEMS entrance exam in which I observe parental education, student rankings of preferred high schools, actual high school placement, and performance on the COMIPEMS exam in the different subjects. With this data I generate counterfactual COMIPEMS scores using the estimates from column 2 of table 2 and a counterfactual ranking of students. This is fed into the allocation algorithm to generate a counterfactual allocation of students to schools. Details on the calculation of these counterfactual scores and the replication of the allocation algorithm are presented in appendix A4.

Table 8 presents estimated impacts for the two years of data. The mean difference between simulated and actual math scores as well as the mean change in rank for students in each SES category are reported in panel a of table 8.[43] Once ranked using the coun-

for students with below median levels of parental education was 495, while the mean score for those with parental education at or above the median level was 578.

[42]This would amount to changing the thematic content of about 3 of the 28 questions that comprise the mathematic components of the COMIPEMS tests.

[43]Because the mathematics section covers only 22% of the exam, the effect on aggregate exam scores and rank is relatively small. I also conservatively opt not to modify the scores of students missing their parental education indicator, though their performance is similar to that of the lower SES students.

terfactual scores, the ranking of students in disadvantaged groups improves, at a cost to those in the more advantaged groups.

Of primary interest is the change in the number of examined students who are ineligible to be assigned to a high school because they fail to meet the 31 point cutoff. For these students, the counterfactual of being eligible to go to high school has the potential to significantly alter the course of their lives. As illustrated in panel b of table 8, the simulation suggests that reducing the share of monetary themed questions on the exam by 10 percentage points would reduce the number of ineligible students by 2.7% for students whose parents are primary educated. Overall, an additional 128 students pass this cutoff and become eligible for high school placement.

The counterfactual rankings also change which students get assigned to a requested school as illustrated in panel c of table 8.[44] Because more students are able to meet the 31 point cutoff, the total number of assigned and unassigned students both increase by 0.07 and 0.06 percentage points respectively. Assignment to high demand high schools is a zero sum game where improved performance by lower SES students results in some displacement of higher SES students. In the simulated allocation, students from higher SES groups are more likely to remain unassigned by 0.05 percentage points. By contrast, in the more disadvantaged groups, students are now more likely to get assigned with a net improvement of 0.11 percentage points.

Finally, the counterfactual scores also change whether students get to attend a more highly preferred school as summarized in panel d of table 8.[45] On average students get assigned to slightly less preferred schools as more students are passing the 31 point threshold, generating more competition and displacement. Here again, heterogeneity is important. Students with highly educated parents experience the bulk of this negative effect, 0.18% being assigned to schools lower on their ranked list. In contrast, 0.14% of students with the least educated parents receive a preferred school assignment.

As the simulation shrinks the test score differential between higher and lower SES students, the allocation of educational opportunities becomes, predictably, more equitable. Note that the effects in this context are relatively small as I only generate counterfactual scores on the mathematics portion of the exam, which accounts for only 22% of a student's score. In contexts where mathematics is weighted more heavily, impacts would be

[44]Not receiving an assignment means that the student did not rank sufficiently high to be placed in any of the schools they listed on their application. In this event, students go through another secondary selection process that allocates unassigned students to the remaining open slots.

[45]Note that I do not use any outside metric of school quality and rely solely on the preference ranking elicited from the students. This listing of school preferences may be endogenous to student expectations about their performance, and these expectations would incorporate expectations regarding monetary questions. The effect of this endogeneity is not reflected in the simulation.

substantially larger.

6 Discussion and Conclusion

Every year, millions of people around the world take examinations that significantly impact their future economic outcomes. Performance on an exam may determine whether they receive a degree or get licensed, which school they can attend and even whether they are eligible to continue their schooling. Societies rely on examinations because they are an efficient way of assessing and ranking a population by ability. The legitimacy of this approach, however, relies heavily on the perception of examinations as fair and objective, and a belief that the skills tested are good proxies for the skills assessors are interested in.

In this paper, I show that lower SES students perform differentially worse on mathematics exams that feature higher shares of monetary themed questions. This performance differential increases with socio-economic disadvantage. Investigation of question-level response data shows evidence of depressed performance on monetary questions, even when compared to questions that are virtually identical. Furthermore, performance is depressed on subsequent questions as well, indicating an attention capture effect as posited in the psychology of poverty literature.

Should monetary themed questions be used in the teaching of mathematical concepts? The evidence evaluated here suggests that in many settings these questions are effective teaching tools. Students engage with these questions but lower SES students find these questions differentially difficult. Nevertheless, being able to apply mathematical concepts to monetary transactions is an important, even critical, skill. To the extent that equipping students with critical life skills is an important goal of early education, one might argue that lower SES students may benefit from more practice using monetized examples to help overcome this disadvantage.

Should monetary themed questions be featured on mathematics exams? It depends on what the examination is supposed to be assessing. If assessing the ability to engage in monetary transactions is a primary goal of the examination, then it would be appropriate. Most high stakes academic mathematics exams are designed to evaluate student preparation for more advanced mathematics studies. To the extent that more advanced mathematics studies do not necessarily center around monetary themes, opting for questions featuring non-monetary content would likely improve examination equity.

Beyond the implications for educational testing, I present non-experimental evidence of attention capture due to poverty, and show that even a temporary effect can have an important impact on a policy relevant outcome. This evidence that lower SES students

underperform and make errors when distracted by a monetary theme has implications beyond the educational setting. Despite being temporary, this effect would impact financial choices made under conditions of scarcity, as it would mechanically activate each time a disadvantaged individual must make a financial decision. These findings support the recommendations made by Mani et al. (2013) that policy makers be cautious of imposing cognitive taxes on the poor, with the additional caveat that this is particularly relevant for financially salient bureaucratic processes.

Policymakers may not be able to prevent this attention capture effect from creating a cognitive cost and inducing errors. However, minimizing the potential to make errors and the possible consequences of these errors is a conceivable avenue for policy intervention. Further research identifying cognitively demanding decisions and processes in which such errors are being committed is warranted. Similarly, educators cannot fully insulate low SES students from the disadvantage generated by the use of monetary examples without depriving them of an important life skill. Given this, it would be valuable to better understand how these effects might be shaping educational choices, aspirations and outcomes. Furthermore, adjusting assessment goals and strategies, by avoiding these monetary topics on high stakes exams where financial literacy is not explicitly being assessed, is a feasible and relatively simple policy. This could prevent these effects from limiting the long run educational opportunities of disadvantaged students and help reduce the intergenerational transmission of poverty.

References

Abdulkadiroglu, Atila and Tayfun Sonmez, "School choice: A mechanism design approach," *American Economic Review*, 2003, *93* (3), 729–747.

Akee, Randall KQ, William E Copeland, Gordon Keeler, Adrian Angold, and E Jane Costello, "Parents' incomes and children's outcomes: a quasi-experiment using transfer payments from casino profits," *American Economic Journal: Applied Economics*, 2010, *2* (1), 86–115.

Akyol, Ş Pelin, James Key, and Kala Krishna, "Hit or miss? Test taking behavior in multiple choice exams," Technical Report, National Bureau of Economic Research 2016.

Baldiga, Katherine, "Gender differences in willingness to guess," *Management Science*, 2014, *60* (2), 434–448.

Bandiera, Oriana, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, and Munshi Sulaiman, "Labor markets and poverty in village economies," *The Quarterly Journal of Economics*, 2017, *132* (2), 811–870.

Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry, "A multifaceted program causes lasting progress for the very poor: Evidence from six countries," *Science*, 2015, *348* (6236).

Banerjee, Abhijit V, Swati Bhattacharjee, Raghabendra Chattopadhyay, and Alejandro J Ganimian, "The Untapped math skills of working children in India: Evidence, possible explanations, and implications," *Unpublished manuscript*, 2017.

Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Ian Levely, "Psychological Effects of Poverty on Time Preferences," *The Economic Journal*, 2021.

Ben-Shakhar, Gershon and Yakov Sinai, "Gender differences in multiple-choice tests: the role of differential guessing tendencies," *Journal of Educational Measurement*, 1991, *28* (1), 23–35.

Bernheim, B Douglas, Debraj Ray, and Şevin Yeltekin, "Poverty and self-control," *Econometrica*, 2015, *83* (5), 1877–1911.

Bertrand, Marianne and Adair Morse, "Information disclosure, cognitive biases, and payday borrowing," *The Journal of Finance*, 2011, *66* (6), 1865–1893.

_ , Sendhil Mullainathan, and Eldar Shafir, "A behavioral-economics view of poverty," *American Economic Review*, 2004, *94* (2), 419–423.

Bowles, Samuel, Herbert Gintis, and Melissa Osborne Groves, *Unequal chances: Family background and economic success*, Princeton University Press, 2009.

Breza, Emily, Supreet Kaur, and Yogita Shamdasani, "The morale effects of pay inequality," *The Quarterly Journal of Economics*, 2018, *133* (2), 611–663.

Bursztyn, Leonardo and Robert Jensen, "How does peer pressure affect educational investments?," *The quarterly journal of economics*, 2015, *130* (3), 1329–1367.

Carvalho, Leandro S, Stephan Meier, and Stephanie W Wang, "Poverty and economic decision-making: Evidence from changes in financial resources at payday," *American Economic Review*, 2016, *106* (2), 260–84.

Danaher, Kelly and Christian S Crandall, "Stereotype threat in applied settings reexamined," *Journal of Applied Social Psychology*, 2008, *38* (6), 1639–1655.

Deckers, Thomas, Armin Falk, Fabian Kosse, and Hannah Schildberg-Hörisch, "How does socio-economic status shape a child's personality?," *Unpublished manuscript*, 2015.

DellaVigna, Stefano, "Psychology and economics: Evidence from the field," *Journal of Economic literature*, 2009, *47* (2), 315–72.

Dobbie, Will and Roland G Fryer Jr, "Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone," *American Economic Journal: Applied Economics*, 2011, *3* (3), 158–87.

Dustan, Andrew, Alain De Janvry, and Elisabeth Sadoulet, "Flourish or fail? The risky reward of elite high school admission in Mexico City," *Journal of Human Resources*, 2017, *52* (3), 756–799.

Editors, "SYMPOSIUM: Bias in the SAT? Continuing the Debate," *Harvard Educational Review*, 2010, *80* (3), 391–394.

Freedle, Roy, "Correcting the SAT's ethnic and social-class bias: A method for reestimating SAT scores," *Harvard Educational Review*, 2003, *73* (1), 1–43.

_ , "On replicating ethnic test bias effects: The Santelices and Wilson study," *Harvard Educational Review*, 2010, *80* (3), 394–404.

Gasiorowska, Agata, Lan Nguyen Chaplin, Tomasz Zaleskiewicz, Sandra Wygrab, and Kathleen D Vohs, "Money cues increase agency and decrease prosociality among children: Early signs of market-mode behaviors," *Psychological Science*, 2016, *27* (3), 331–344.

Gertler, Paul J, Sebastian W Martinez, and Marta Rubio-Codina, "Investing cash transfers to raise long-term living standards," *American Economic Journal: Applied Economics*, 2012, *4* (1), 164–92.

Haushofer, Johannes and Ernst Fehr, "On the psychology of poverty," *Science*, 2014, *344* (6186), 862–867.

_ **and Jeremy Shapiro**, "The short-term impact of unconditional cash transfers to the poor: experimental evidence from Kenya," *The Quarterly Journal of Economics*, 2016, *131* (4), 1973–2042.

_ **and** _ , "The long-term impact of unconditional cash transfers: experimental evidence from Kenya," *Busara Center for Behavioral Economics, Nairobi, Kenya*, 2018.

International Association for the Evaluation of Educational Achievement (IEA), "TIMSS Assesment 2011-4th Grade," *Publisher: TIMSS and PIRLS International Study Center, Lynch School of Education, Boston College*, (accessed June 2017).

_ , "TIMSS Assesment 2015-4th Grade," *Publisher: TIMSS and PIRLS International Study Center, Lynch School of Education, Boston College*, (accessed June 2017).

Jr, Roland G Fryer, Steven D Levitt, and John A List, "Exploring the impact of financial incentives on stereotype threat: Evidence from a pilot study," *American Economic Review*, 2008, *98* (2), 370–75.

Jurajda, Štěpán and Daniel Münich, "Gender gap in performance under competitive pressure: Admissions to Czech universities," *American Economic Review*, 2011, *101* (3), 514–18.

Karlan, Dean, Aishwarya Lakshmi Ratan, and Jonathan Zinman, "Savings by and for the Poor: A Research Review and Agenda," *Review of Income and Wealth*, 2014, *60* (1), 36–78.

Kaur, Supreet, Sendhil Mullainathan, Suanna Oh, and Frank Schilbach, "Does Financial Strain Lower Productivity?," Technical Report, Working Paper 2019.

Kim, Jinhee, Benoit Sorhaindo, and E Thomas Garman, "Relationship between financial stress and workplace absenteeism of credit counseling clients," *Journal of Family and Economic Issues*, 2006, *27* (3), 458–478.

Koch, Alexander, Julia Nafziger, and Helena Skyt Nielsen, "Behavioral economics of education," *Journal of Economic Behavior & Organization*, 2015, *115*, 3–17.

Kremer, Michael, Gautam Rao, and Frank Schilbach, "Behavioral development economics," in "Handbook of Behavioral Economics: Applications and Foundations 1," Vol. 2, Elsevier, 2019, pp. 345–458.

Lavecchia, Adam M, Heidi Liu, and Philip Oreopoulos, "Behavioral economics of education: Progress and possibilities," in "Handbook of the Economics of Education," Vol. 5, Elsevier, 2016, pp. 1–74.

Levitt, Steven D and John A List, "What do laboratory experiments measuring social preferences reveal about the real world?," *Journal of Economic perspectives*, 2007, *21* (2), 153–174.

_ , _ , Susanne Neckermann, and Sally Sadoff, "The behavioralist goes to school: Leveraging behavioral economics to improve educational performance," *American Economic Journal: Economic Policy*, 2016, *8* (4), 183–219.

Lodder, Paul, How Hwee Ong, Raoul PPP Grasman, and Jelte M Wicherts, "A comprehensive meta-analysis of money priming," *Journal of Experimental Psychology: General*, 2019, *148* (4), 688.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao, "Poverty impedes cognitive function," *Science*, 2013, *341* (6149), 976–980.

McLoyd, Vonnie C, "Socioeconomic disadvantage and child development.," *American psychologist*, 1998, *53* (2), 185.

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

Ors, Evren, Frédéric Palomino, and Eloic Peyrache, "Performance gender gap: does competition matter?," *Journal of Labor Economics*, 2013, *31* (3), 443–499.

Ostrow, Korinn and Neil Heffernan, "Advancing the State of Online Learning: Stay Integrated, Stay Accessible, Stay Curious," in Robert Feldman, ed., *Learning Science: Theory, Research, and Practice*, McGraw Hill Education, 2019.

_ , D. Selent, Y. Wang, E. Van Inwegen, N. Heffernan, and J. J. Williams, "Assessment of Learning Infrastructure (ALI): The Theory, Practice, and Scalability of Automated Assessment," *Proceedings of 6th International Conference on Learning Analytics and Knowledge, 2016*, (data accessed May 2018).

Pekkarinen, Tuomas, "Gender differences in behaviour under competitive pressure: Evidence on omission patterns in university entrance examinations," *Journal of Economic Behavior & Organization*, 2015, *115*, 94–110.

Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel, "Poverty, depression, and anxiety: Causal evidence and mechanisms," *Science*, 2020, *370* (6522).

Riener, Gerhard and Valentin Wagner, "Shying away from demanding tasks? Experimental evidence on gender differences in answering multiple-choice questions," *Economics of Education Review*, 2017, *59*, 43–62.

Rohrer, Doug, Harold Pashler, and Christine R Harris, "Do subtle reminders of money change people's political views?," *Journal of experimental psychology: General*, 2015, *144* (4).

Santelices, Maria Veronica and Mark Wilson, "Unfair treatment? The case of Freedle, the SAT, and the standardization approach to differential item functioning," *Harvard Educational Review*, 2010, *80* (1), 106–134.

Schilbach, Frank, Heather Schofield, and Sendhil Mullainathan, "The psychological lives of the poor," *American Economic Review*, 2016, *106* (5), 435–40.

Schofield, Heather, "The economic costs of low caloric intake: Evidence from India," *Unpublished Manuscript*, 2014.

Secretaría de Educación Pública, "Evaluacion Nacional del Logro Academico en Centros Escolares," *enlace.sep.gob.mx*, (data accessed June 2018).

Shah, Anuj K, Jiaying Zhao, Sendhil Mullainathan, and Eldar Shafir, "Money in the mental lives of the poor," *Social Cognition*, 2018, *36* (1), 4–19.

_ , Sendhil Mullainathan, and Eldar Shafir, "Some consequences of having too little," *Science*, 2012, *338* (6107), 682–685.

_ , _ , and _ , "An exercise in self-replication: Replicating Shah, Mullainathan, and Shafir (2012)," *Journal of Economic Psychology*, 2018.

Spears, Dean, "Economic decision-making in poverty depletes behavioral control," *The BE Journal of Economic Analysis & Policy*, 2011, *11* (1).

Spencer, Steven J, Christine Logel, and Paul G Davies, "Stereotype threat," *Annual Review of Psychology*, 2016, *67*, 415–437.

Steele, Claude M and Joshua Aronson, "Stereotype threat and the intellectual test performance of African Americans," *Journal of Personality and Social Psychology*, 1995, *69* (5), 797.

Stricker, Lawrence J and William C Ward, "Stereotype Threat, Inquiring About Test Takers' Ethnicity and Gender, and Standardized Test Performance 1," *Journal of Applied Social Psychology*, 2004, *34* (4), 665–693.

Vohs, Kathleen D, "Money priming can change people's thoughts, feelings, motivations, and behaviors: An update on 10 years of experiments.," *Journal of Experimental Psychology: General*, 2015, *144* (4), e86.

_ **, Nicole L Mead, and Miranda R Goode**, "The psychological consequences of money," *Science*, 2006, *314* (5802), 1154–1156.

Wei, Thomas E, "Sticks, stones, words, and broken bones: New field and lab evidence on stereotype threat," *Educational Evaluation and Policy Analysis*, 2012, *34* (4), 465–488.

Zhao, Jiaying and Brandon M Tomm, "Attentional trade-offs under resource scarcity," in "International Conference on Augmented Cognition" Springer 2017, pp. 78–97.

Zhou, Xinyue, Kathleen D Vohs, and Roy F Baumeister, "The symbolic power of money: Reminders of money alter social distress and physical pain," *Psychological Science*, 2009, *20* (6), 700–706.

Tables

Table 1: Dataset Features

	TIMSS	ASSISTments	ENLACE*
Setting	International Exam	US Homework	Mexico Exam
Stakes	Low	Completing the assignment	For teachers and schools
Variation in Financial Salience	Across randomly assigned booklets	From random question selection	Across grades and years
SES Indicator	Student: Parental education	School: Share of schoolmates in free or reduced lunch programs	School: Local marginalization index
Panel			Yes
Itemized Question Data	Available	Available	
Question Matching		Possible	
Randomized Question Ordering		Yes	

Note: The ENLACE data and results are presented in appendix A2.

Table 2: Financial Salience and Aggregate Performance in TIMSS

	Standardized Score			
	(1)	(2)	(3)	(4)
Below Nat. Median	-0.438		-0.268	
	(0.00403)		(0.00388)	
Post Secondary		-0.315		-0.209
		(0.00527)		(0.00496)
Upper Secondary		-0.493		-0.334
		(0.00492)		(0.00478)
Lower Secondary		-0.712		-0.488
		(0.00700)		(0.00676)
Primary or None		-0.779		-0.523
		(0.00762)		(0.00735)
Below Nat. Median x Prop Mon Q.	-0.260		-0.262	
	(0.0472)		(0.0428)	
Post Sec x Prop Mon Q.		-0.0812		-0.0515
		(0.0635)		(0.0582)
Upper Sec x Prop Mon Q.		-0.0865		-0.0844
		(0.0584)		(0.0539)
Lower Sec x Prop Mon Q.		-0.160		-0.147
		(0.0817)		(0.0744)
Prim/No x Prop Mon Q.		-0.219		-0.242
		(0.0845)		(0.0748)
Constant	0.160	0.314	0.100	0.213
	(0.00167)	(0.00218)	(0.00155)	(0.00215)
FE: Booklet x Year	Yes	Yes	Yes	Yes
FE: Country	Yes	Yes	.	.
FE: Class	No	No	Yes	Yes
N	379468	379468	379160	379160

Note: Robust standard errors in parentheses. Observations are at the student by examination level with a student-level SES indicator: parental education. Omitted categories are students with parental education at or above the national median for columns 1 and 3 and university educated parents for columns 2 and 4. The proportion of monetary questions in a booklet is a value from 0 to 1.

Table 3: Monetary and Subsequent Questions in TIMSS

	Question Answered Correctly (=100)					
	(1)	(2)	(3)	(4)	(5)	(6)
Below Nat. Median x Mon Q.	-0.885		-1.207		-0.753	
	(0.123)		(0.131)		(0.151)	
Post Sec. x Mon Q.		-0.800		-0.0641		-0.459
		(0.165)		(0.177)		(0.188)
Upper Sec. x Mon Q.		-1.351		-0.502		-0.338
		(0.152)		(0.163)		(0.185)
Lower Sec. x Mon Q.		-1.948		-1.682		-0.919
		(0.213)		(0.229)		(0.262)
Primary/No x Mon Q.		-1.786		-2.548		-1.514
		(0.227)		(0.241)		(0.306)
Below Nat. Median x 4 Post	-0.680		-0.891		-0.395	
	(0.0880)		(0.0969)		(0.109)	
Post Sec. x 4 Post		-0.614		-0.397		-0.482
		(0.119)		(0.129)		(0.136)
Upper Sec. x 4 Post		-1.012		-0.717		-0.677
		(0.109)		(0.120)		(0.134)
Lower Sec. x 4 Post		-0.926		-0.998		-0.653
		(0.153)		(0.170)		(0.191)
Primary/No x 4 Post		-0.925		-1.213		-0.534
		(0.161)		(0.180)		(0.222)
FE: Student	Yes	Yes	Yes	Yes	Yes	Yes
FE: Question	Yes	Yes	Yes	Yes	Yes	Yes
FE: Below Med. x Diff.	No	.	Yes	.	Yes	.
FE: Below Med. x Seq.	No	.	Yes	.	Yes	.
FE: Below Med. x QType x Country	No	.	Yes	.	Yes	.
FE: Below Med. x QTopic x Country	No	.	Yes	.	Yes	.
FE: Par. Edu. x Diff.	.	No	.	Yes	.	Yes
FE: Par. Edu. x Seq.	.	No	.	Yes	.	Yes
FE: Par. Edu. x QType x Country	.	No	.	Yes	.	Yes
FE: Par. Edu. x QTopic x Country	.	No	.	Yes	.	Yes
FE: Class x Mon Q.	No	No	No	No	Yes	Yes
FE: Class x 4 Post	No	No	No	No	Yes	Yes
Dependent Variable Mean	49.56	49.56	49.56	49.56	49.56	49.56
Dependent Variable SD	23.56	23.56	23.56	23.56	23.56	23.56
N	9564201	9564201	9564201	9564201	9563918	9563918

Note: Standard errors in parentheses clustered at the student-level. Observations are at the question by student-level with a student-level SES indicator: parental education. When a question is answered correctly the indicator is set to 100, 0 otherwise. Omitted categories are students with parental education at or above the national median for columns 1, 3 and 5 and university educated parents for columns 2, 4 and 6. Difficulty is a 20 bin binned indicator based on the performance on a question by students with university educated parents. Sequence is a 5 bin binned indicator based on the the position of a question within the exam booklet. Question type indicates whether a question is multiple choice or completed response. Question topic indicates categorized questions based on the topics listed in panel b of figure A6.

Table 4: Monetary Questions in ASSISTments

	Correct (=100)	Hints	Attempts	Time (Sec)
	(1)	(2)	(3)	(4)
Pct. Free/Red. Lunch x Mon Q.	-10.39	0.241	0.145	25.40
	(1.676)	(0.0430)	(0.0380)	(4.585)
Monetary Question	6.318	-0.220	-0.0950	-2.506
	(0.569)	(0.0153)	(0.0122)	(1.469)
Sequence Position	0.162	-0.00281	-0.00368	-1.624
	(0.0618)	(0.00237)	(0.00109)	(0.139)
P-value of F-test: $\delta_2 + \delta_3 = 0$	0.002	0.532	0.089	0.000
FE: Problem Set x School	Yes	Yes	Yes	Yes
Dependent Mean	66.63	0.675	1.505	102.9
Dependent SD	46.23	1.290	0.938	125.0
N	91092	91092	88198	91092

Note: Standard errors in parentheses clustered at the student assignment level. Observations are at the student by question-level with a school level SES indicator: the share of students receiving free or reduced lunch. Inactive observations are dropped (5 sec $<$ Time $<$ 15 min). Outlier attempt counts beyond 5 attempts are dropped in column 3. Monetary question is a dummy variable set to 1 if the question features a monetary theme. The omitted category is non-monetary questions.

Table 5: Matched Monetary Questions in ASSISTments

	Correct (=100)	Hints	Attempts	Time (Sec)
	(1)	(2)	(3)	(4)
Pct. Free/Red. Lunch x Mon Q.	-6.636	0.283	0.111	38.71
	(3.016)	(0.0771)	(0.0695)	(8.480)
Monetary Question	1.228	0.00932	-0.00948	15.27
	(0.889)	(0.0218)	(0.0191)	(2.431)
Sequence Position	0.149	0.00184	-0.00485	-2.049
	(0.109)	(0.00289)	(0.00162)	(0.210)
P-value of F-test: $\gamma_2 + \gamma_3 = 0$	0.025	0.000	0.069	0.000
FE: Matched Group x School	Yes	Yes	Yes	Yes
Dependent Mean	73.31	0.492	1.414	100.4
Dependent SD	42.93	1.058	0.859	116.4
N	23771	23771	23245	23771

Note: Standard errors in parentheses clustered at the student assignment level. Observations are at the student by question-level with a school level SES indicator: the share of students receiving free or reduced lunch. Observations are limited to monetary questions and questions that have been matched to a monetary question. Inactive observations are dropped (5 sec $<$ Time $<$ 15 min). Outlier attempt counts beyond 5 attempts are dropped in column 3. Monetary question is a dummy variable set to 1 if the question features a monetary theme. The omitted category is non-monetary questions.

Table 6: ASSISTments Questions After Matched Monetary Questions

	Following first matched question				Following first matched question (within first three)			
	(1) Correct (=100)	(2) Hints	(3) Attempts	(4) Time (Sec)	(5) Correct (=100)	(6) Hints	(7) Attempts	(8) Time (Sec)
PFR Lunch x 4 Post Matched Mon Q.	-13.47	0.460	0.235	59.12	-21.88	0.524	-0.0656	69.88
	(8.647)	(0.243)	(0.221)	(23.74)	(16.74)	(0.462)	(0.505)	(40.97)
4 Post Matched Mon Q.	7.771	-0.224	-0.110	-11.45	10.74	-0.239	-0.0842	-20.96
	(2.416)	(0.0642)	(0.0560)	(6.400)	(3.890)	(0.102)	(0.0983)	(9.823)
Sequence Positon	3.531	-0.0651	-0.0691	-9.448	14.32	-0.257	-0.197	-23.98
	(0.739)	(0.0184)	(0.0167)	(1.859)	(3.856)	(0.104)	(0.0871)	(10.17)
PFR Lunch x Seq. Pos.	-5.685	0.104	0.129	7.568	-27.91	0.560	0.185	46.83
	(1.905)	(0.0495)	(0.0457)	(5.870)	(18.58)	(0.554)	(0.453)	(49.60)
P-value of F-test: $\beta_2 + \beta_3 = 0$	0.419	0.240	0.493	0.0148	0.429	0.465	0.727	0.154
FE: Leading Matched Q. Group	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE: Question x School	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Mean	64.37	0.731	1.556	113.9	64.12	0.687	1.563	116.4
Dependent SD	46.82	1.272	0.991	128.4	46.67	1.208	0.980	123.7
N	4521	4521	4278	4521	1795	1795	1685	1795

Note: Standard errors in parentheses clustered at the student assignment level. Observations are at the student by question-level with a school level SES indicator: the share of students receiving free or reduced lunch. The full sample in columns 1-4 consists of questions positioned between the first and second matched question a student faces and no more than four questions following the first matched question. Columns 5-8 limit the sample to questions within the first three questions a student faces. Inactive observations are dropped (5 sec $<$ Time $<$ 15 min). Outlier attempt counts beyond 5 attempts are dropped in column 3 and 7. 4 Post matched monetary question is a dummy variable set to 1 if a question is in the four questions that follow a monetary themed question. The omitted category are questions that follow matched non-monetary themed questions.

Table 7: ASSISTments Questions After Matched Monetary Questions with Lead Difficulty Controls

	Following first matched question				Following first matched question (within first three)			
	(1) Correct (=100)	(2) Hints	(3) Attempts	(4) Time (Sec)	(5) Correct (=100)	(6) Hints	(7) Attempts	(8) Time (Sec)
PFR Lunch x 4 Post Matched Mon Q.	-13.72	0.460	0.244	56.19	-24.51	0.457	-0.0620	61.92
	(8.664)	(0.243)	(0.223)	(23.87)	(17.28)	(0.466)	(0.520)	(42.90)
4 Post Matched Mon Q.	7.471	-0.223	-0.102	-15.09	10.48	-0.246	-0.0839	-21.77
	(2.600)	(0.0688)	(0.0576)	(6.791)	(3.932)	(0.103)	(0.0981)	(9.860)
Sequence Positon	3.523	-0.0651	-0.0689	-9.550	14.36	-0.256	-0.197	-23.85
	(0.738)	(0.0184)	(0.0167)	(1.861)	(3.857)	(0.104)	(0.0873)	(10.18)
PFR. Lunch x Seq. Pos.	-5.671	0.104	0.129	7.740	-28.33	0.549	0.185	45.56
	(1.903)	(0.0495)	(0.0458)	(5.872)	(18.55)	(0.553)	(0.453)	(49.60)
Quartile Mean Time on Lead Q.	0.00818	-0.0000142	-0.000246	0.0992	0.0232	0.000594	-0.0000303	0.0702
	(0.0234)	(0.000626)	(0.000521)	(0.0656)	(0.0399)	(0.00104)	(0.000843)	(0.107)
P-value of F-test: $\beta_2 + \beta_3 = 0$	0.387	0.245	0.450	0.0401	0.347	0.597	0.745	0.277
FE: Leading Matched Q. Group	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE: Question x School	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Mean	64.37	0.731	1.556	113.9	64.12	0.687	1.563	116.4
Dependent SD	46.82	1.272	0.991	128.4	46.67	1.208	0.980	123.7
N	4521	4521	4278	4521	1795	1795	1685	1795

Note: Standard errors in parentheses clustered at the student assignment level. Observations are at the student by question-level with a school level SES indicator: the share of students receiving free or reduced lunch. The full sample in columns 1-4 consists of questions positioned between the first and second matched question a student faces and no more than four questions following the first matched question. Columns 5-8 limit the sample to questions within the first three questions a student faces. Inactive observations are dropped (5 sec $<$ Time $<$ 15 min). Outlier attempt counts beyond 5 attempts are dropped in column 3 and 7. 4 Post matched monetary question is a dummy variable set to 1 if a question is in the four questions that follow a monetary themed question. The omitted category are questions that follow matched non-monetary themed questions. Quartile mean time on leading matched question is the mean time spent by students on the leading matched questions who share the same quartile of the SES indicator (share of students receiving free and reduced lunches in the school).

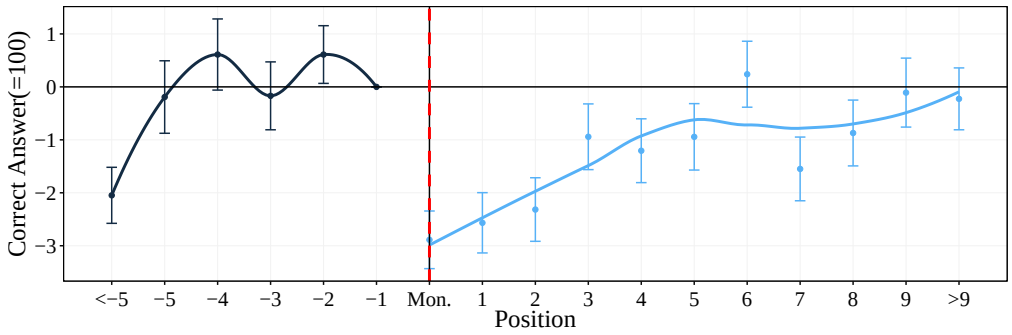
Table 8: COMIPEMS Simulation

Parental Education	All	Missing	University	Upper Secondary	Lower Secondary	Primary or Less
Panel a: Simulated change in performance.						
Mean Change in Math Points (out of 13)	0.05	0.00	0.00	0.05	0.07	0.10
Mean Change in Rank	0	-197	-167	-13	88	175
Panel b: High school eligibility.						
Remain Ineligible	1.64%	2.16%	0.56%	1.01%	1.89%	2.44%
Become Eligible	0.03%	0.00%	0.00%	0.01%	0.03%	0.07%
Panel c: Receiving a high school assignment.						
Remain Unassigned	18.39%	20.48%	17.43%	18.25%	18.38%	17.42%
Become Assigned	0.07%	0.01%	0.00%	0.07%	0.10%	0.16%
Become Unassigned	0.06%	0.08%	0.05%	0.06%	0.07%	0.05%
Remain Assigned	81.47%	79.42%	82.52%	81.63%	81.45%	82.37%
Panel d: Allocation to High Schools.						
Preferred High School Assignment	0.20%	0.02%	0.00%	0.19%	0.31%	0.34%
Less Preferred High School Assignment	0.20%	0.20%	0.18%	0.21%	0.21%	0.20%
Mean Change in Preference Rank	-.00028	-.00453	-.00445	-.00082	.00234	.00308
Observations	502943	81164	62592	130618	137216	91353

Note: The mean change in math points reports the difference in the mean performance of students' actual and simulated scores on the mathematics portion of the COMIPEMS exam. The mathematics and quantitative ability portions accounts for 28 points out of the 128 point exam. Scores on the COMIPEMS use round numbers only so simulated scores are generated by creating a random binomial and adding 1 point to the math COMIPEMS score of randomly selected students in disadvantaged SES groups such that their aggregate performance on the mathematics section is improved in a manner consistent with the TIMSS estimates. The scores of students with missing parental education indicators are not modified. The mean change in rank reports how the simulated scores change the rank position of the students in that category on average. Students completing the COMIPEMS exam receive a score that makes them either eligible for high school (if they score 31 points or more) or ineligible. If eligible, students receive a high school assignment if there are openings in their listed schools or remain unassigned in which case they undergo a secondary assignment process in the remaining openings. Panels b and c show the share of students whose eligibility and assignment statuses change when ranked according to the simulated scores. Panel d shows the share of students who are allocated to a more or less preferred high school when using the simulated scores. The mean change in preference rank is calculated using only students who receive a high school assignment in both the actual and simulated data.

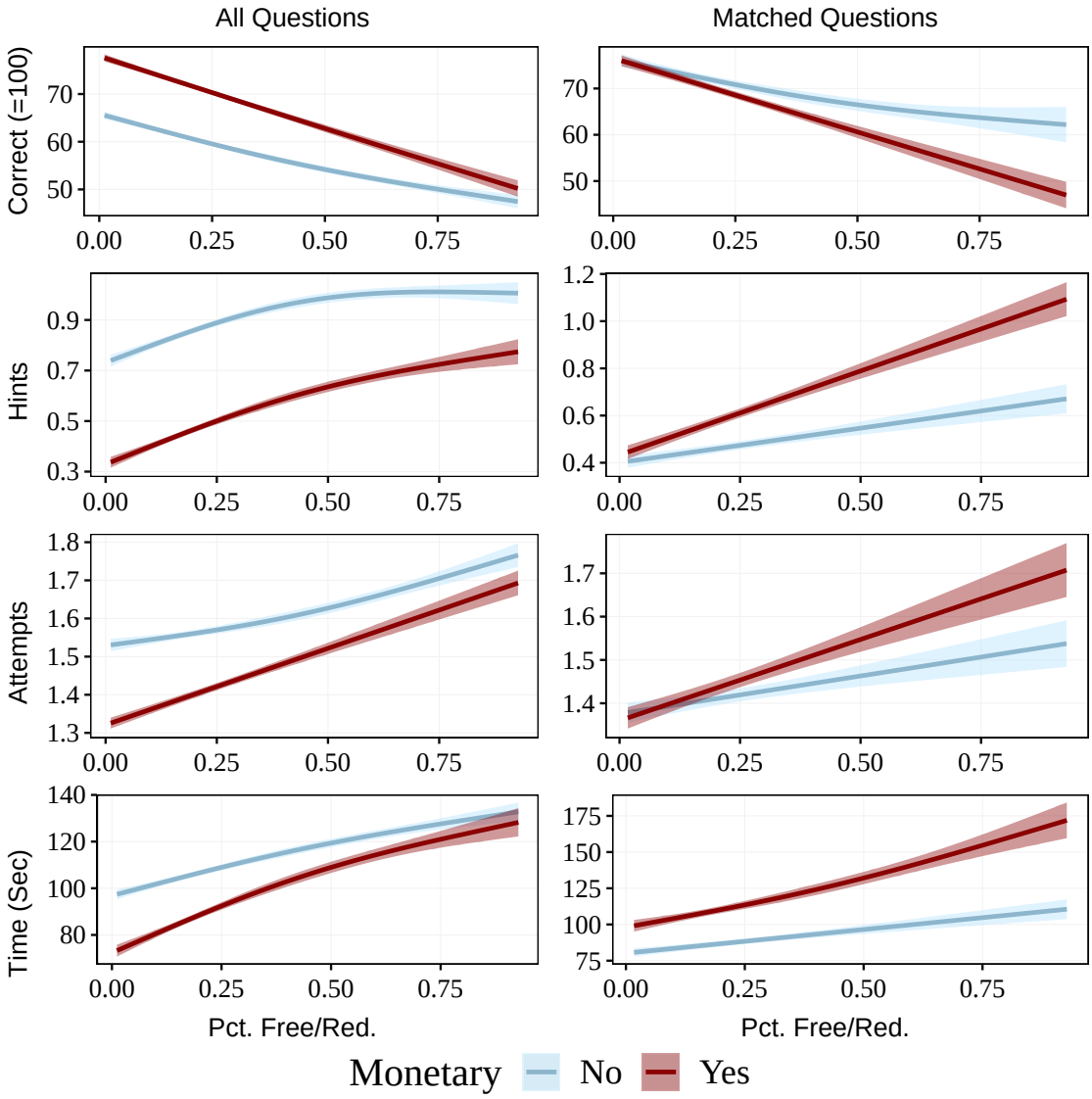
Figures

Figure 1: Differential Performance by Question Position Relative to Monetary Event for Below National Median Students in TIMSS



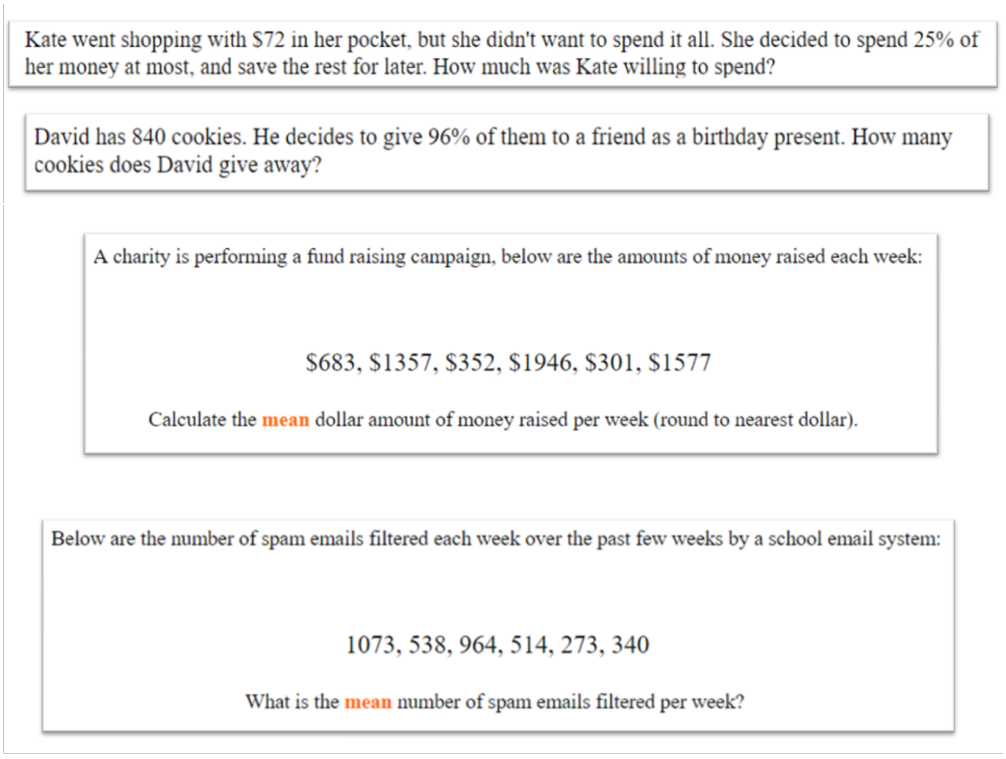
Note: This figure displays the event study estimates based on the positioning of a question relative to a monetary question. Plotted is the differential for students with below median parental education indicators in answering the question correctly (Correct=100) controlling for student and question fixed effects as well as the X_{pq} vector of fixed effects that controls for differential performance based on question observables. Data is limited to booklets that feature a single monetary question or booklets that feature only two sequential monetary questions. The question right before the monetary question is the omitted category. 95pct confidence intervals are displayed using standard errors clustered at the student-level.

Figure 2: ASSISTments Question Statistics by Question Type



Note: This figure displays performance and effort metrics by the share of students receiving free or reduced lunch rates in the school based on whether the question is monetary themed or not. Metrics in the first column are calculated using all questions in the data. Metrics in the second column are calculated using only the matched subset of questions. The shaded area displays the 0.95 confidence interval of the estimate.

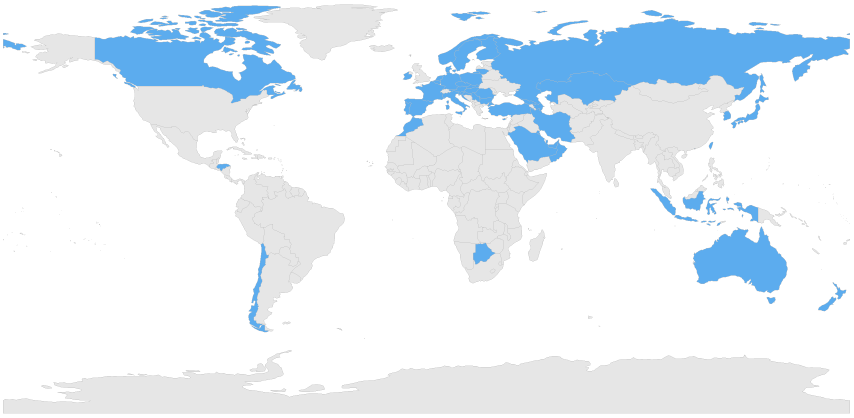
Figure 3: Two Examples of Matched ASSISTments Questions



Note: Additional examples are available in appendix figures A7 and A8.

For Online Publication: Appendix

Figure A1: Countries participating in 4th grade TIMSS



Note: Mapped countries only show countries participating in the 4th grade TIMSS in 2011 and/or 2015 in which parental questionnaires were administered. Countries participating in 2015 4th grade exams that administered parental questionnaires include: Abu Dhabi, Australia, Bahrain, Flemish Belgium, Buenos Aires, Bulgaria, Canada, Chile, Chinese Taipei, Croatia, Cyprus, Czech Republic, Denmark, Dubai, Finland, France, Georgia, Germany, Hong-Kong, Hungary, Indonesia, Ireland, Iran, Italy, Japan, Kazakhstan, Republic of Korea, Kuwait, Lithuania, Morocco, Northern Ireland, Netherlands, Norway, New Zealand, Oman, Ontario, Poland, Portugal, Qatar, Quebec, Russian Federation, Saudi Arabia, Singapore, Serbia, Slovak Republic, Slovenia, Spain, Sweden, Turkey, United Arab Emirates. Countries participating in 2011 4th grade exams that administered parental questionnaires include: Abu Dhabi, Australia, Austria, Azerbaijan, Botswana, Chinese Taipei, Croatia, Czech Republic, Dubai, Finland, Germany, Georgia, Honduras, Hong-Kong, Hungary, Ireland, Iran, Italy, Lithuania, Morocco, Malta, Northern Ireland, Norway, Oman, Poland, Portugal, Qatar, Romania, Russian Federation, Saudi Arabia, Singapore, Slovak Republic, Slovenia, Spain, Sweden, Quebec, United Arab Emirates.

A1 Parental survey non-response in TIMSS

Table A1 shows that students whose parents do not complete the parental questionnaire perform worse than their peers, even when controlling for classroom fixed effects, and these students also do worse on their exam if it features a higher share of monetary questions. These estimates are comparable to those for students with parental SES indicators below the national median. Since my estimation exploits the random assignment of booklets to students, which is orthogonal to parental non-response, as demonstrated by table A3, estimates are internally valid.

Table A1: Missing Parental Education

	Standardized Score			
	(1)	(2)	(3)	(4)
Missing Parental Education	-0.297	-0.233	-0.283	-0.221
	(0.00368)	(0.00356)	(0.00509)	(0.00474)
Missing Par Edu. x Prop Mon Q.			-0.227	-0.200
			(0.0578)	(0.0519)
FE: Booklet x Year	Yes	Yes	Yes	Yes
FE: Country	Yes	.	Yes	.
FE: Class	No	Yes	No	Yes
N	445342	445190	445342	445190

Note: Robust standard errors in parentheses. Observations are at the student by examination level with a student-level SES indicator: parental education. The omitted category is students with reported parental education levels. The proportion of monetary questions in a booklet is a value from 0 to 1.

Table A2: TIMSS Booklet Structure

	Part 1		**Part 2**	
	First Block	Second Block	First Block	Second Block
Booklet 1	M01	M02	S01	S02
Booklet 2	S02	S03	M02	M03
Booklet 3	M03	M04	S03	S04
Booklet 4	S04	S05	M04	M05
Booklet 5	M05	M06	S05	S06
Booklet 6	S06	S07	M06	M07
Booklet 7	M07	M08	S07	S08
Booklet 8	S08	S09	M08	M09
Booklet 9	M09	M10	S09	S10
Booklet 10	S10	S11	M10	M11
Booklet 11	M11	M12	S11	S12
Booklet 12	S12	S13	M12	M13
Booklet 13	M13	M14	S13	S14
Booklet 14	S14	S01	M14	M01

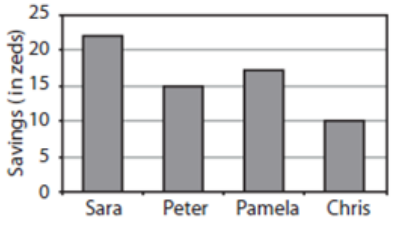
Note: A student handed booklet one would complete their math section first in part 1 and after a short break their science section in part 2. The math component of their exam would consist of prompt blocks M01 and M02. In contrast, a student handed booklet two would complete their science section first in part 1 followed by their math section in part 2. Their math section would consist of prompt blocks M02 and M03. Thus about half of the math prompts are identical between booklets 1 and 2. Additionally, 8 prompt blocks are re-administered between 2011 and 2015.

Figure A2: Example Monetary Questions from the 2011 TIMSS

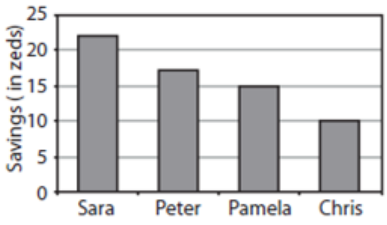
John was given the following table by his teacher and was asked to identify the graph that correctly displays the data. Which graph below should he choose?

Name	Savings
Sara	22 zeds
Peter	15 zeds
Pamela	17 zeds
Chris	10 zeds

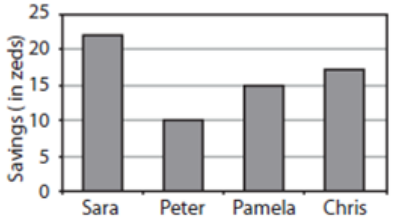
A.



B.



C.



D.

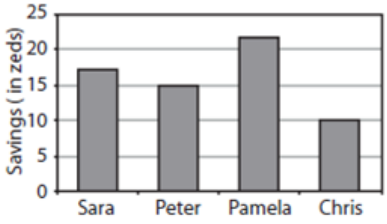
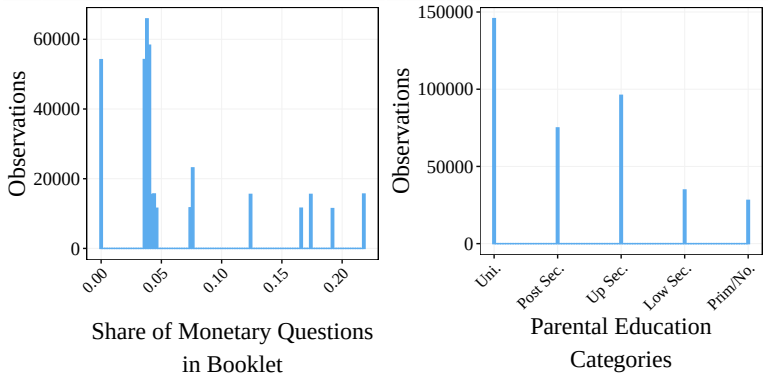
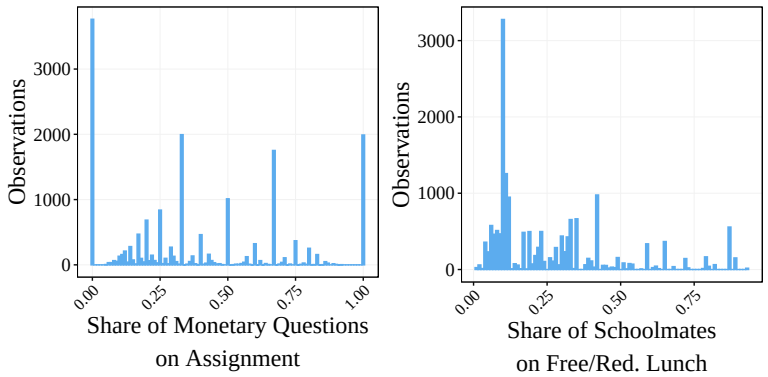


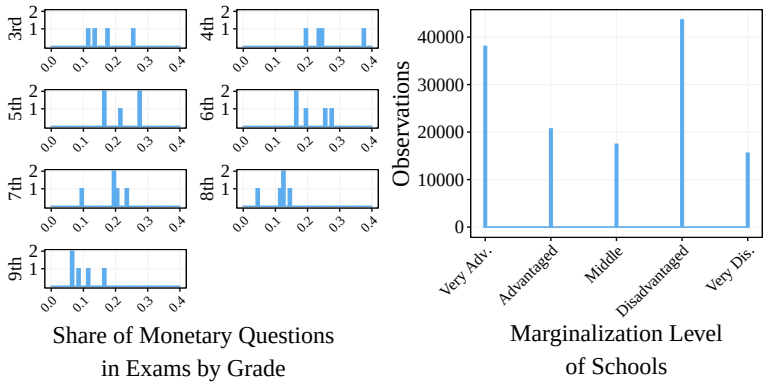
Figure A3: Variation in Financial Salience and SES Indicators



Panel a: TIMSS



Panel b: ASSISTments



Panel c: ENLACE*

Note: This figure shows the variation in the share of financially salient questions in student assignments and exams in the first column for all three datasets. The second column displays the variation in the SES indicators in all three datasets. The ENLACE data and results are presented in appendix A2.

Table A3: Exam booklet randomization check

	Proportion Mon Q. in Booklet			
	(1)	(2)	(3)	(4)
Below Nat. Median	-0.000141		-0.000234	
	(0.000198)		(0.000239)	
Post Secondary		-0.000266		-0.000270
		(0.000263)		(0.000294)
Upper Secondary		-0.000315		-0.000367
		(0.000242)		(0.000291)
Lower Secondary		0.000185		0.0000258
		(0.000346)		(0.000419)
Primary or None		0.000178		0.0000918
		(0.000378)		(0.000499)
FE: Year	Yes	Yes	Yes	Yes
FE: Class	No	No	Yes	Yes
N	379468	379468	379160	379160

Note: Robust standard errors in parentheses. Observations are at the student by examination level with a student-level SES indicator: parental education. Omitted categories are students with reported parental education levels above the national median in columns 1 and 3 and university educated parents in columns 2 and 4. The proportion of monetary questions in a booklet is a value from 0 to 1.

Table A4: Question Randomization Check

	Prop. MonQ.	MonQ. in 1st 3
Pct. Free/Red. Lunch	0.00313	0.00974
	(0.00826)	(0.0285)
FE: Problem set	Yes	Yes
Dependent Mean	0.391	1.195
Dependent SD	0.324	1.017
N	17428	16015

Note: Robust standard errors in parentheses. Observations are at the student by problem set level with a school level SES indicator: share of students in the school receiving free or reduced lunch. Observations in the second column only include student assignments in which at least three questions were assigned.

Table A5: Financial Salience and Exam-level Performance in TIMSS

	Standardized Score			
	(1)	(2)	(3)	(4)
Below Nat. Median	-0.438		-0.268	
	(0.00975)		(0.00728)	
Post Secondary		-0.315		-0.209
		(0.00906)		(0.00793)
Upper Secondary		-0.493		-0.334
		(0.0112)		(0.00855)
Lower Secondary		-0.712		-0.488
		(0.0131)		(0.00941)
Primary or None		-0.779		-0.523
		(0.0168)		(0.0123)
Below Nat. Median x Prop Mon Q.	-0.260		-0.262	
	(0.0885)		(0.0732)	
Post Sec x Prop Mon Q.		-0.0812		-0.0515
		(0.108)		(0.0629)
Upper Sec x Prop Mon Q.		-0.0865		-0.0844
		(0.101)		(0.0711)
Lower Sec x Prop Mon Q.		-0.160		-0.147
		(0.137)		(0.0821)
Prim/No x Prop Mon Q.		-0.219		-0.242
		(0.177)		(0.153)
Constant	0.160	0.314	0.100	0.213
	(0.00221)	(0.00426)	(0.00157)	(0.00323)
FE: Booklet x Year	Yes	Yes	Yes	Yes
FE: Country	Yes	Yes	.	.
FE: Class	No	No	Yes	Yes
N	379468	379468	379160	379160

Note: Standard errors in parentheses clustered at the booklet level. Observations are at the student by examination level with a student-level SES indicator: parental education. Omitted categories are students with parental education at or above the national median for columns 1 and 3 and university educated parents for columns 2 and 4. The proportion of monetary questions in a booklet is a value from 0 to 1.

Table A6: TIMSS Exam-level Results by Occupation

	Standardized Score	
	(1)	(2)
Small Business	-0.163	-0.114
	(0.00651)	(0.00596)
Clerical	-0.249	-0.131
	(0.00497)	(0.00467)
Skilled Labor	-0.373	-0.220
	(0.00659)	(0.00626)
General Labor	-0.518	-0.292
	(0.00989)	(0.00930)
Never Wk. for Pay	-0.438	-0.217
	(0.0105)	(0.00924)
Small Business x Prop Mon Q.	0.0142	-0.0646
	(0.0792)	(0.0715)
Clerical x Prop Mon Q.	-0.0179	-0.0374
	(0.0598)	(0.0550)
Skilled Labor x Prop Mon Q.	-0.0968	-0.0815
	(0.0776)	(0.0720)
General Labor x Prop Mon Q.	-0.182	-0.215
	(0.114)	(0.103)
Never Wk. for Pay x Prop Mon Q.	-0.161	-0.212
	(0.122)	(0.102)
Constant	0.148	0.0857
	(0.00195)	(0.00178)
FE: Booklet x Year	Yes	Yes
FE: Country	Yes	.
FE: Class	No	Yes
N	379468	379160

Note: Robust standard errors in parentheses. Observations are at the student by examination level with a student-level SES indicator: parental occupation. The omitted category is students with professional parental occupations. The proportion of monetary questions in a booklet is a value from 0 to 1.

A2 Estimation using ENLACE exams

The Mexican Evaluación Nacional de Logros Académicos en Centros Escolares (ENLACE) exams were administered to all public and private schools in the country each June from 2006 to 2013. ENLACE exams were low stakes exams for students but were used to assess school and teacher performance. The census of school level subject results for grades 3 through 9 in all schools in Mexico is publicly available. The data also includes the school's marginalization index (1 to 5) as defined by Mexico's National Population Council.[1] All students in Mexico in the same grade take the same exam with new exams being written for all the grades each year.[2] ENLACE examination booklets are also publicly available. Within each booklet, I tally the total number of mathematics prompts and the number featuring a monetary theme.[3] Panel c of figure A3 shows that there is variation in the proportion of monetary questions featured on exams within each grade. Panel c of figure A3 also shows the variation in the marginalization indicator across schools, the other source of variation necessary for my estimation.[4]

I use a panel of school performance for 135,307 different schools between the years 2009 and 2013.[5] Because multiple grades are tested each year, the ENLACE data allows me to look at heterogeneity of effect sizes by grade.

[1]Mexico's National Population Council (CONAPO) calculates local geographic marginalization indices using a principal components method based on percentage indicators of social exclusion collected in the census. Indicators include: illiteracy, incomplete primary education, lack of running water, sewage systems, electricity, dirt floors, household overcrowding, geographic isolation, and low incomes in employment. Further details are available at http://www.conapo.gob.mx. The index is then grouped into five levels of marginalization: very disadvantaged, disadvantaged, middle, advantaged and very advantaged which are provided for the schools in the data. Although the marginalization index does not change over time for most schools, the index changes for a small share of schools with CONAPO's scheduled recalculation of the indicator. To simplify interpretation and presentation of the results I opt to maintain the discreet nature of this index and treat this index as time invariant, calculating the average for each school and rounding to the closest index category.

[2]In 2011 two different test booklets were used for the 3rd and 4th grades in certain regions. As the data does not indicate which booklet was used, these observations are dropped from the final dataset.

[3]Figure A4 illustrates an example of a monetary themed question on a 4th grade ENLACE exam.

[4]The histogram of school marginalization in panel c of figure A3 does not weigh school observations by the number of students. When weighted, the median student attends a very advantaged school. For this analysis, schools that are not very advantaged are coded as below median.

[5]In many schools, examinations were administered in several sessions throughout the day. Performance data is reported for each session. I construct a single school level subject result for each grade by calculating a weighted average of the performance in the different sessions using the number of tested students as weights. Though some data is available for the earlier years, the number of examined students is not included in the 2006 and 2007 data. Furthermore, the data in 2008 does not disaggregate performance by subject. Analysis is thus focused on the years 2009-2013.

A2.1 Impacts in ENLACE Examinations

Using the ENLACE exam data, I estimate

$$E_{sgy} = \Theta_1 + \Theta_2 LowZ_s * PM_{gy} + \kappa_{gy} + \tau_{sy} + \epsilon_{sgy}, \quad (1)$$

$$E_{sgy} = \theta_1 + \sum_{z=2}^{5} \theta_{2z} Z_s * PM_{gy} + \kappa_{gy} + \tau_{sy} + \epsilon_{sgy}. \quad (2)$$

Each observation represents the mean performance on the mathematics portion of the exam of grade (g) in school (s) in year (y). I regress the standardized school average for each grade and year (E_{gys}) on SES indicators interacted with the proportion of questions on that grade's exam that featured a monetary theme that year (PM_{gy}).[6] Here, SES indicators include an indicator for whether a school's marginalization index falls below the national median $(LowZ_s)$ in equation 1 or the school marginalization index dummies (Z_s) as specified in equation 2. I include a grade by year fixed effect (κ_{gy}) to control for overall difficulty of each particular exam booklet and school by year fixed effects (τ_{sy}) to control for local shocks that might affect overall performance in a school. In some specifications, I also add a grade by school fixed effect (ρ_{gs}) to control for the time invariant performance of a grade in a school. As exam booklets are assigned at the cohort level, standard errors are clustered at the booklet (grade by year) level.

Results are reported in table A7 and are consistent with the results using the TIMSS data. Students in disadvantaged schools see their mathematics exam scores further depressed when more monetary questions are featured on the exam. As illustrated in panel b of figure A3, the percentage of monetary questions featured on an exam can vary by up to 18 percentage points within a grade level. These estimates suggest that a 10 percentage point increase in the share of monetary themed questions differentially reduces performance in below median schools by 0.023 standard deviations and up to 0.126 standard deviations in very disadvantaged schools. The overall performance gap between above and below median schools is 0.43 standard deviations. Thus the effect of a 10 percentage point increase in monetary salience represents about 5.3% of the overall performance gap.

A2.2 Effect Size by Grade

The ENLACE data reports exam performance in a school for each grade between the 3rd and 9th grades. I investigate the grade heterogeneity in effect sizes by adding the relevant interaction terms with the indicators for the k grades.

[6] The standardization of the school averages is weighted by the number of students who took the exam.

Results are displayed graphically in figure A5. Figure A5 suggests that the the negative impact of financially salient questions on the exam performance of students in disadvantaged schools generally increases as they progress through the grades within a school. The negative impact reaches its largest magnitudes in the 6th and 9th grades which are the terminal grades for elementary and junior secondary schools. The transition from elementary to junior secondary school, between the 6th and 7th grade, is associated with a break in the overall trend as there is substantial selection and sorting of the students who enter into junior secondary school. Note that selection of students and differences in the topics tested across grades complicates across grade comparisons. Nonetheless, this is suggestive evidence that the magnitude of the negative impact of financially salient questions on the exam performance of students in disadvantaged schools generally increases as they progress through the grades within a school.

Figure A4: Example Page from 4th Grade ENLACE Mathematics

17. Cuatro amigos leen un libro. La cantidad que cada uno ha leído se muestra en la tabla.

Amigo	Cantidad leída
Daniel	$\frac{2}{3}$
Fernando	$\frac{1}{5}$
Manuel	$\frac{3}{4}$
Guillermo	$\frac{5}{6}$

¿Qué amigo ha leído menos?

A) Fernando.
B) Daniel.
C) Manuel.
D) Guillermo.

18. María va a realizar el pago de los siguientes recibos: teléfono $209.40, luz $198.50 y agua $100.30. ¿Cuánto pagará en total?

A) $507.02
B) $507.12
C) $508.02
D) $508.20

19. ¿Cuál de las siguientes figuras tiene menos ejes de simetría?

A) 

B) 

C) 

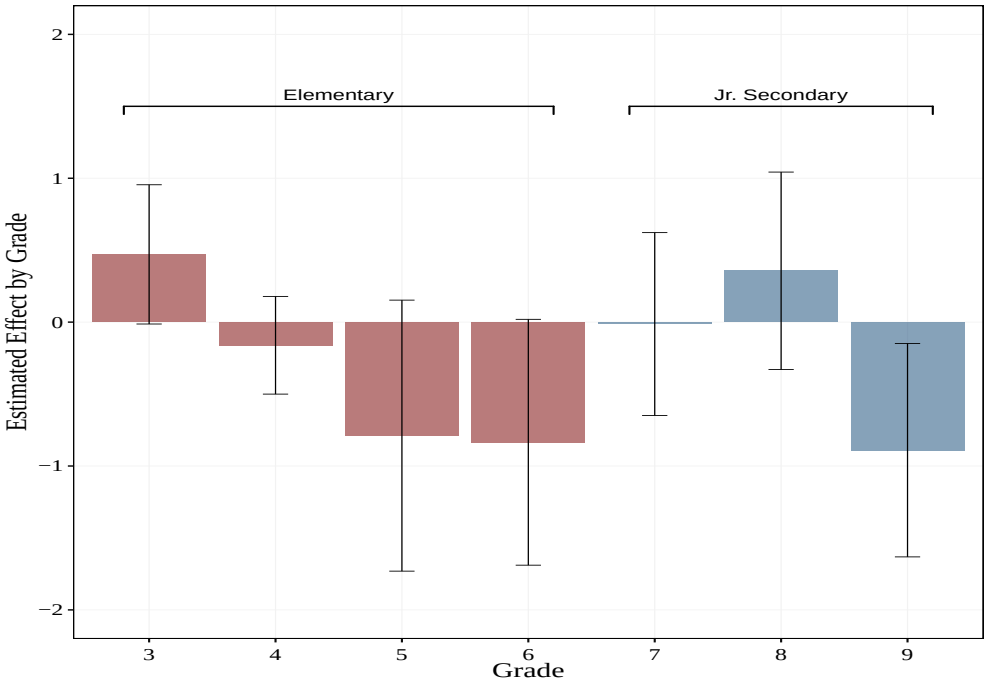
D) 

Table A7: Financial Salience and Aggregate Performance in ENLACE

	Standardized Score			
	(1)	(2)	(3)	(4)
Below Median x Prop Mon Q.	-0.240		-0.232	
	(0.123)		(0.158)	
Advantaged x Prop Mon Q.		-0.160		-0.0903
		(0.0585)		(0.0932)
Middle x Prop Mon Q.		-0.296		-0.220
		(0.128)		(0.151)
Disadvantaged x Prop Mon Q.		-0.217		-0.277
		(0.192)		(0.241)
Very Disadvantaged x Prop Mon Q.		-0.837		-1.256
		(0.413)		(0.473)
FE: Grade x Year	Yes	Yes	Yes	Yes
FE: Year x School	Yes	Yes	Yes	Yes
FE: School x Grade	No	No	Yes	Yes
N	1912259	1912259	1870964	1870964

Note: Standard errors in parentheses are clustered at the grade by year (booklet) level. Observations are weighted by the number of tested students. Observations are at the school by grade by year level with a school level SES indicator: the school's marginalization index. Omitted categories are schools at or above the median marginalization level of students for columns 1 and 3 and very advantaged schools for columns 2 and 4. The proportion of monetary questions in a booklet is a value from 0 to 1.

Figure A5: Impacts of Financial Salience in ENLACE by Grade



Note: This figure displays estimates of the performance differential by grade with 95% confidence intervals, clustered at the grade by year (booklet) level.

Table A8: TIMSS Question Fixed Effects

	Question Answered Correctly (=100)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Below Nat. Median x Mon Q.	-0.885	-1.207	-0.694	-1.573	-1.399	-0.484	-0.478	-1.208	-1.436	-0.992
	(0.123)	(0.131)	(0.124)	(0.126)	(0.127)	(0.128)	(0.128)	(0.131)	(0.128)	(0.131)
Below Nat. Median x 4 Post	-0.680	-0.891	-0.716	-0.990	-1.024	-0.642	-0.674	-0.888	-0.858	-1.078
	(0.0880)	(0.0969)	(0.0893)	(0.0957)	(0.0969)	(0.0886)	(0.0903)	(0.0954)	(0.0969)	(0.0970)
FE: Student	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE: Question	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE: Below Med. x Diff.	No	Yes	No	Yes	Yes	No	No	Yes	Yes	Yes
FE: Below Med. x Seq.	No	Yes	Yes	No	Yes	No	Yes	No	Yes	Yes
FE: Below Med. x QType x Country	No	Yes	No	No	No	Yes	Yes	Yes	Yes	No
FE: Below Med. x QTopic x Country	No	Yes	No	No	No	Yes	Yes	Yes	No	Yes
Dep. Variable Mean	49.93	49.93	49.93	49.93	49.93	49.93	49.93	49.93	49.93	49.93
Dep. Variable SD	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00
N	9564201	9564201	9564201	9564201	9564201	9564201	9564201	9564201	9564201	9564201

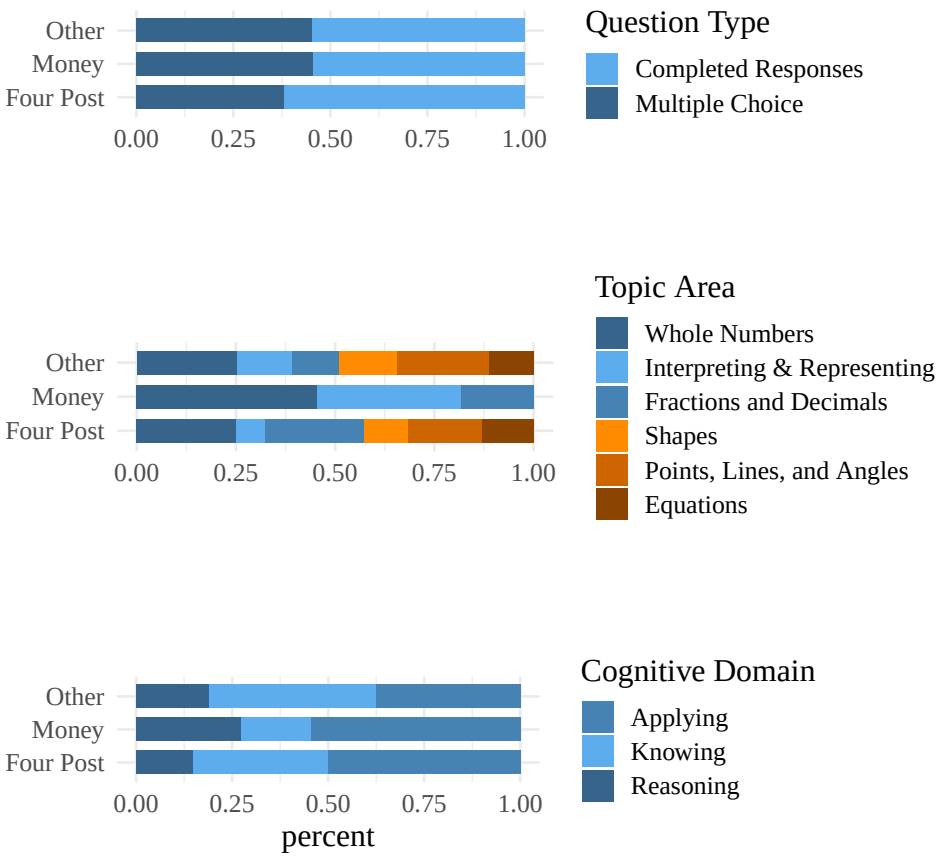
Note: Standard errors in parentheses clustered at the student-level. Observations are at the question by student-level with a student-level SES indicator: parental education relative to the national median. When a question is answered correctly the indicator is set to 100, 0 otherwise. Omitted categories are students with parental education at or above the national median. Difficulty is a 20 bin binned indicator based on the performance on a question by students with parental education above the national median. Sequence is a 5 bin binned indicator based on the the position of a question within the exam booklet. Question type indicates whether a question is multiple choice or completed response. Question topic indicates categorized questions based on the topics listed in panel b of figure A6.

Table A9: TIMSS Unanswered Questions

	Question Left Unanswered (=1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Below Nat. Median x Mon Q.	-0.00378		-0.00362		-0.00169	
	(0.000691)		(0.000723)		(0.000803)	
Post Sec. x Mon Q.		-0.000789		-0.00153		-0.000661
		(0.000848)		(0.000888)		(0.000927)
Upper Sec. x Mon Q.		0.00120		-0.00182		-0.00294
		(0.000813)		(0.000847)		(0.000942)
Lower Sec. x Mon Q.		-0.00120		-0.00264		-0.00266
		(0.00127)		(0.00133)		(0.00146)
Primary/No x Mon Q.		-0.00732		-0.00842		-0.00873
		(0.00144)		(0.00151)		(0.00182)
Below Nat. Median x 4 Post	-0.00147		-0.00296		-0.00232	
	(0.000545)		(0.000604)		(0.000666)	
Post Sec. x 4 Post		0.00157		-0.00148		-0.00166
		(0.000671)		(0.000742)		(0.000780)
Upper Sec. x 4 Post		0.00416		-0.000794		-0.00301
		(0.000641)		(0.000705)		(0.000784)
Lower Sec. x 4 Post		0.00351		-0.00151		-0.00324
		(0.000994)		(0.00110)		(0.00121)
Primary/No x 4 Post		0.00178		-0.00164		-0.00674
		(0.00115)		(0.00129)		(0.00152)
FE: Student	Yes	Yes	Yes	Yes	Yes	Yes
FE: Question	Yes	Yes	Yes	Yes	Yes	Yes
FE: Below Med. x Diff.	No	.	Yes	.	Yes	.
FE: Below Med. x Seq.	No	.	Yes	.	Yes	.
FE: Below Med. x QType x Country	No	.	Yes	.	Yes	.
FE: Below Med. x QTopic x Country	No	.	Yes	.	Yes	.
FE: Par. Edu. x Diff.	.	No	.	Yes	.	Yes
FE: Par. Edu. x Seq.	.	No	.	Yes	.	Yes
FE: Par. Edu. x QType x Country	.	No	.	Yes	.	Yes
FE: Par. Edu. x QTopic x Country	.	No	.	Yes	.	Yes
FE: Class x Mon Q.	No	No	No	No	Yes	Yes
FE: Class x 4 Post	No	No	No	No	Yes	Yes
Dep. Variable Mean	0.0598	0.0598	0.0598	0.0598	0.0598	0.0598
Dep. Variable SD	0.237	0.237	0.237	0.237	0.237	0.237
N	9564201	9564201	9564201	9564201	9563918	9563918

Note: Standard errors in parentheses clustered at the student-level. Observations are at the question by student-level with a student-level SES indicator: parental education relative to the national median. When a question is answered correctly the indicator is set to 100, 0 otherwise. Omitted categories are students with parental education at or above the national median for columns 1, 3 and 5 and university educated parents for columns 2, 4 and 6. Difficulty is a 20 bin binned indicator based on the performance on a question by students with university educated parents. Sequence is a 5 bin binned indicator based on the the position of a question within the exam booklet. Question type indicates whether a question is multiple choice or completed response. Question topic indicates categorized questions based on the topics listed in panel b of figure A6.

Figure A6: Question Characteristics by Category



Note: This figure displays the distribution of observable question characteristics in the TIMSS data based on whether a question is categorized as monetary or positioned within the four questions that follow a monetary question.

Figure A7: Additional Examples of Matched ASSISTments Questions

Bob works at a local convenience store and has been putting every penny earned into a savings account. Now, Bob wants to use a bank statement to graph the balance of the account.
How much money does Bob make each day given the bank statement below?
Bob's salary does not change during the time shown by the statement.

Bob's Account

Date	Balance (In $)
Day 7	898.00
Day 9	994.00
Day 15	1282.00
Day 16	1330.00

Joe is part of an environmental group that researches birds. Every so often, the group counts the number of birds in an area. Now, Joe wants to graph the number of birds to find the rate that the population grows each year.

Given the results shown below, how much does the population of birds grow each year? Assume the rate of growth is constant.

birds in area

Year of Result	Population Size
1964	1484
1971	2226
1973	2438
1979	3074

Liam is an artist. He paints portraits.

The table below shows the number of portraits painted in hours.

Number of Portraits	Time (in hours)
1	6
2	12
3	18

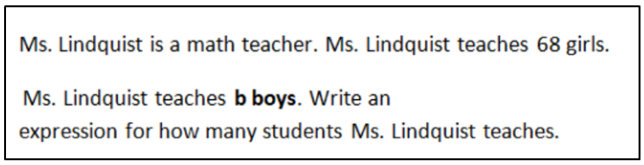
Do the numbers in the table represent a proportional relationship?

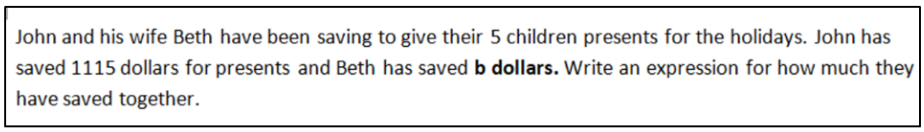
This table shows the amount earned by Isabella for selling ice cream.

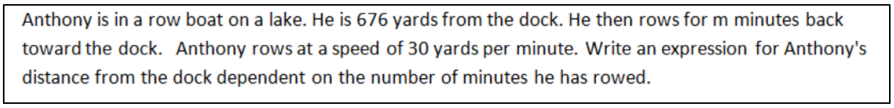
Cups Sold	Earnings ($)
3	15
6	30
9	45

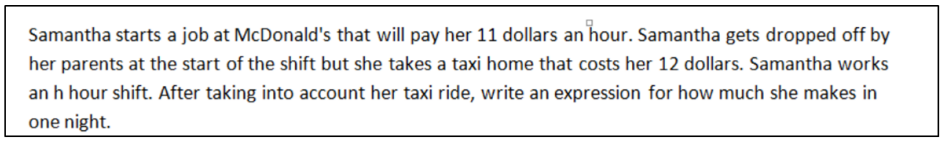
Do the numbers in the table represent a proportional relationship?

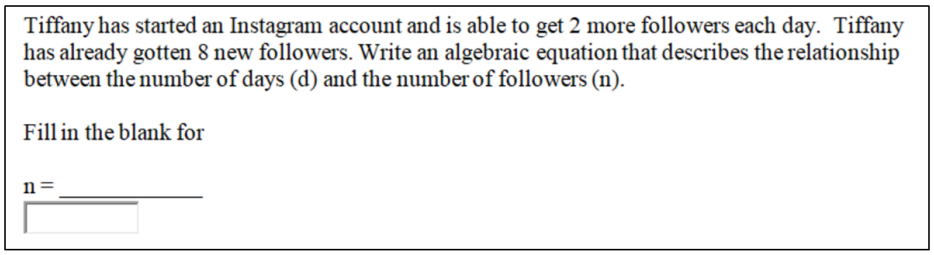
Figure A8: Additional Examples of Matched ASSISTments Questions

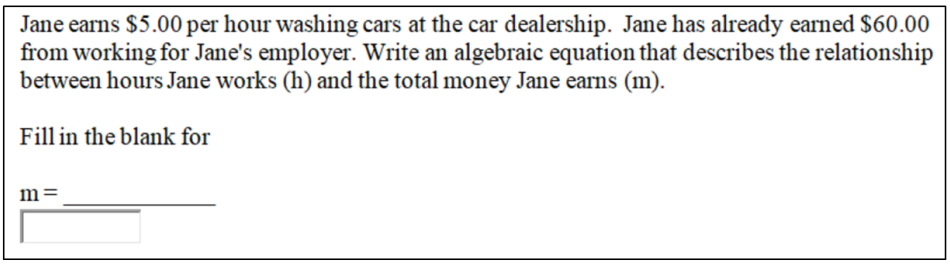












A3 Attention Capture versus Cognitive Fatigue in TIMSS

The estimates in table 3 and figure 1 are very much consistent with the attention capture hypothesis. Nevertheless, there remains the possibility that the estimates on subsequent questions reflect the effect of cognitive fatigue as discussed in the ASSISTments data. For each question, I calculate the share of students from each SES category that answered the question correctly and use this as an indicator for how difficult a question is for a student from a particular SES category. I generate four lags of this indicator to control for differential difficulty of the four questions leading up to a question. Results are reported in table A10. Controlling for the differential difficulty of leading questions in columns 3 and 4 seems to slightly reduce the magnitude of the estimates on subsequent questions by a small amount, though they remain negative and statistically significant, suggesting that this explanation cannot explain the entirety of the effect.

As an alternative method to estimate whether the effect on subsequent questions is due to cognitive fatigue from the differential difficulty of preceding questions, I generate 1000 placebo estimates from the data. Instead of flagging the true monetary questions, I flag a random set of questions as monetary and the 4 questions following this random set as post questions.[7] These placebos are then used to estimate equation 4.

The resulting pairs of $\hat{\Lambda}_3^{placebo}$ and $\hat{\Lambda}_2^{placebo}$ coefficients are plotted in figure A9. The scatter plot suggests that it is highly unlikely that the two coefficients would both be jointly negative and of such a large magnitude by random chance, confirming the main results. In addition to verifying the main results, looking at the correlation between the coefficient pairs can also help decompose the role of cognitive fatigue in explaining the effect on subsequent questions. Suppose the differential difficulty of preceding questions generates differential cognitive fatigue and thus differential performance on subsequent questions. Under these conditions, if the randomly selected placebo monetary questions happen to be differentially difficult for the low SES students, then we would expect them to perform differentially worse on subsequent questions and vice versa. Thus we would expect the correlation between $\hat{\Lambda}_3^{placebo}$ and $\hat{\Lambda}_2^{placebo}$ to be positive. To investigate this, I estimate the following regression.

$$\hat{\Lambda}_{3p}^{placebo} = \psi_1 + \psi_2 \hat{\Lambda}_{2p}^{placebo} + \epsilon_p \tag{3}$$

Results are reported in table A11 and plotted in figure A9. ψ_2 is indeed positive and

[7] Some question blocks are repeated across the two years. To ensure that the distribution is representative of the actual distribution of monetary questions, I make sure to randomly select 6 questions from the non-repeated blocks and 8 questions from the repeated blocks.

statistically significant, suggesting that cognitive fatigue due to the differential difficulty of preceding questions does explain part of the magnitude of the estimated effect on subsequent questions. Nonetheless, as visible in figure A9, the predicted value of the coefficient on subsequent questions using the estimated placebos ($\hat{\hat{\Lambda}}_3^{placebo}$) is significantly smaller in magnitude than the estimate using the actual monetary questions, $\hat{\Lambda}_3$. I can reject that cognitive fatigue due to the differential difficulty of preceding questions explains the entirety of the effect on subsequent questions, supporting the attention capture hypothesis. When decomposed, I estimate that cognitive fatigue due to the differential difficulty of the previous questions explains approximately 34% of the estimated effect on subsequent questions.[8] I interpret the remainder as evidence of attention capture.

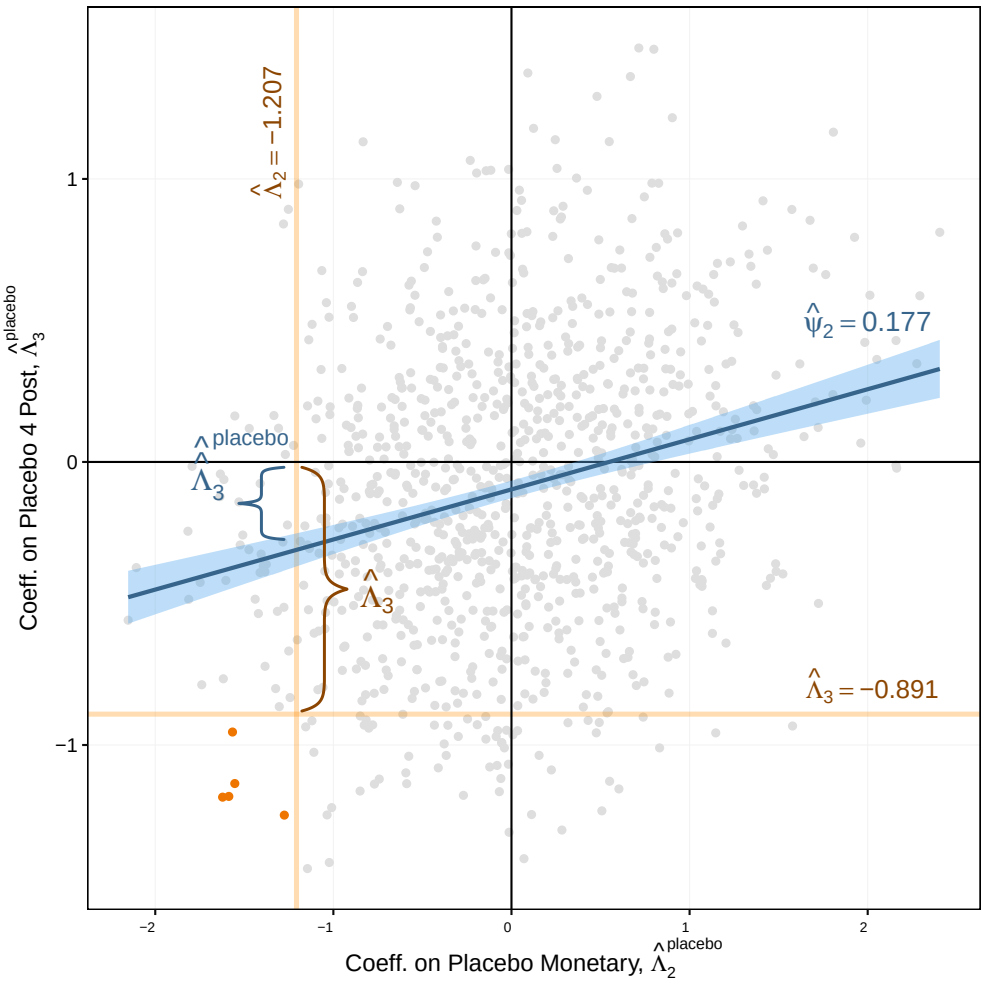
[8]Estimates in table A11 imply that $\mathbb{E}(\hat{\Lambda}_3^{placebo}|\hat{\Lambda}_2^{placebo} = -1.207) = -0.31$ or 34% of $\hat{\Lambda}_3 = -0.891$.

Table A10: Monetary and Subsequent Questions in TIMSS with Controls for Preceding Differential Difficulty

	Question Answered Correctly (=100)							
	(1)		(2)		(3)		(4)	
Below Nat. Median x Mon Q.	-1.108	(0.151)			-1.202	(0.153)		
Post Sec. x Mon Q.			0.190	(0.202)			0.199	(0.202)
Upper Sec. x Mon Q.			-0.450	(0.187)			-0.461	(0.187)
Lower Sec. x Mon Q.			-1.599	(0.265)			-1.640	(0.265)
Primary/No x Mon Q.			-2.249	(0.280)			-2.352	(0.282)
Below Nat. Median x 4 Post	-1.005	(0.102)			-0.878	(0.104)		
Post Sec. x 4 Post			-0.393	(0.136)			-0.370	(0.136)
Upper Sec. x 4 Post			-0.793	(0.126)			-0.754	(0.127)
Lower Sec. x 4 Post			-1.002	(0.179)			-0.904	(0.180)
Primary/No x 4 Post			-1.306	(0.191)			-1.156	(0.192)
Below Med. Performance on q-1					0.0612	(0.0102)		
Below Med. Performance on q-2					0.0312	(0.0107)		
Below Med. Performance on q-3					-0.0130	(0.00993)		
Below Med. Performance on q-4					-0.0498	(0.0103)		
Par. Edu. Group Performance on q-1							0.0327	(0.00644)
Par. Edu. Group Performance on q-2							-0.0160	(0.00660)
Par. Edu. Group Performance on q-3							0.0359	(0.00638)
Par. Edu. Group Performance on q-4							-0.0576	(0.00662)
FE: Student	Yes		Yes		Yes		Yes	
FE: Question	Yes		Yes		Yes		Yes	
FE: Below Med. x Diff.	Yes		.		Yes		.	
FE: Below Med. x Seq.	Yes		.		Yes		.	
FE: Below Med. x QType x Country	Yes		.		Yes		.	
FE: Below Med. x QTopic x Country	Yes		.		Yes		.	
FE: Par. Edu. x Diff.	.		Yes		.		Yes	
FE: Par. Edu. x Seq.	.		Yes		.		Yes	
FE: Par. Edu. x QType x Country	.		Yes		.		Yes	
FE: Par. Edu. x QTopic x Country	.		Yes		.		Yes	
Exam Mean	49.56		49.56		49.56		49.56	
Exam SD	23.56		23.56		23.56		23.56	
N	8046329		8046329		8046329		8046329	

Note: Standard errors in parentheses clustered at the student-level. Observations are at the question by student-level with a student-level SES indicator: parental education. When a question is answered correctly the indicator is set to 100, 0 otherwise. Omitted categories are students with parental education at or above the national median for columns 1, 3 and 5 and university educated parents for columns 2, 4 and 6. Difficulty is a 20 bin binned indicator based on the performance on a question by students with university educated parents. Sequence is a 5 bin binned indicator based on the the position of a question within the exam booklet. Question type indicates whether a question is multiple choice or completed response. Question topic indicates categorized questions based on the topics listed in panel b of figure A6. The sample mechanically does not include the first four questions on an exam for which the differential difficulty controls are undefined.

Figure A9: Estimates from 1000 Placebo Estimations



Note: This figure displays the coefficient estimates obtained for the monetary and four subsequent questions from 1000 placebo regressions where monetary questions were randomly assigned. Only five iterations yield estimated coefficients that were both jointly negative and of equal or larger magnitude than the estimates reported in column 3 of table 3. The regression line shows that there is a positive correlation between the differential difficulty of the leading question and differential performance on subsequent questions but that this effect can only explain about a third of the overall performance gap on subsequent questions.

Table A11: Regressions on Placebo Coefficients

	Placebo Post Estimates ($\hat{\Lambda}_3^{placebo}$)
Placebo Mon. Estimates ($\hat{\Lambda}_2^{placebo}$)	0.177
	(0.0208)
Constant	-0.0970
	(0.0157)
N	1000

Note: $\hat{\Lambda}_2^{placebo}$ are the estimates for Λ_2 from equation 4 when randomly selected questions are flagged as placebo monetary questions. $\hat{\Lambda}_3^{placebo}$ are the estimates for Λ_3 on the corresponding placebo subsequent questions.

A4 Simulating Effects and Replicating the COMIPEMS Placement Algorithm

The COMIPEMS exams consists of 128 multiple choice questions covering mathematics, Spanish, history and the natural sciences and is administered to about 250,000 students each year. Though I do not observe the exam booklets, practice COMIPEMS mathematics questions do feature monetary themed questions. I use data from the 2004 and 2005 COMIPEMS entrance exam in which I observe parental education, student rankings of preferred high schools, actual high school placement, and performance on the COMIPEMS exam in the different subjects.

I consider how a 10 percentage point decrease in the share of monetary questions would change the scores of students on the mathematics portion of the COMIPEMS test. I use the parental education indicator and the estimates from column 2 of table 2 to calculate a student's new counterfactual math COMIPEMS score and then generate a new counterfactual ranking of students.[9] These estimates suggest that a 10 percentage point decrease in the share of monetary questions should increase the mean score of students who have parents with a primary or less education by 0.0219 standard deviations.[10] Exam

[9]I only simulate the effect of monetary questions on the mathematics portion of the exam as this paper has focused on mathematics, and all of the estimates are derived using mathematics questions. I elected to focus on mathematics questions because monetary questions are a common feature in mathematics instruction and the structure of many mathematics exams and assignments (multiple short, distinct questions) helps with identification. Nevertheless, though not identified in this paper, it is possible these effects may apply to other subjects.

[10]The math portion on the COMIPEMS had a standard deviation of 5.12 points in 2004 and 5.26 points in 2005. Because scores on the COMIPEMS use round numbers only, I use these values to generate a random

scores are not adjusted for the substantial number of students with a missing SES indicator making the outcome of this simulation conservative. If a substantial share of these are low SES students, as their average performance suggests, there would be more movement in the allocation of students. Once ranked using the counterfactual scores, the ranking of students in disadvantaged groups improves, at a cost to those in the more advantaged groups.[11]

Students in Mexico City are assigned to high schools according to a serial dictatorship mechanism. High schools first set the maximum number of students they will accept.[12] Students who fail to score above 30 or who fail to complete middle school are disqualified from attending high school. The remaining students are placed based on their ranked exam performance and the list of preferred schools students submit prior to taking the exam. A computer program proceeds through the ranked list of students, starting with the highest scoring, and allocates students to their top-ranked school with open seats remaining. If no seats remain at any of the schools a student listed, the student is unassigned. After the first assignment process, unassigned students undergo a secondary selection process that allocates them to remaining openings (Dustan et al. (2017)). Following these rules, I perfectly replicate the placement algorithm used by the COMIPEMS's centralized admission system and verify that high school placement in Mexico City actually follows the rules described above.[13]

binomial and add 1 point to the math COMIPEMS score of randomly selected students with primary or less parental education indicators such that their aggregate performance on the mathematics section is improved in a manner consistent with the TIMSS estimate. I repeat the same procedure for students at each marginalization level using the relevant estimates.

[11]Ranking among students with identical exam scores is generated randomly.

[12]Many students receive the exact same COMIPEMS score. In the actual assignment process, once a school's available slots are filled, the school must elect to admit all or none of the students who receive the marginal score and would otherwise be assigned to that school based on the student's stated school preferences. Since I do not observe this rounding process, I cannot replicate it in the simulation. For competitive schools, where the lowest exam score of an admitted student was above 31, I use the number of students who were admitted into the school in each year as the maximum number admissible. I do not constrain the number of students admitted for non-competitive schools.

[13]In addition to the matching conditions above, UNAM and IPN affiliated schools have an additional minimum GPA requirement.