Fictional Money, Real Costs:
Impacts of Financial Salience on Disadvantaged Students

Claire Duquennois[1]

October 2019
Latest version available here.

(Job Market Paper)

Abstract

Performance on examinations is a significant determinant of educational and economic opportunities. Using three data sets, I find evidence that students with lower socio-economic status (SES) indicators perform worse when randomly given an exam or assignment that features a larger share of mathematics questions in which money is salient. This pattern begins as early as in the fourth grade, is largest for the most disadvantaged and is responsive to income shocks. For students with SES indicators below the national median, a 10 percentage point increase in the share of monetary themed questions depresses exam performance by 0.026 to 0.038 standard deviations. The magnitude of the effect represents about 6% of the overall performance gap for below median SES students. Evidence from a homework platform shows that acquiring a mathematical skill takes differentially more time and effort for low SES students when it is practiced using monetary prompts. Using question-level data, I confirm the role of financial salience by comparing performance on monetary and highly similar non-monetary questions. Furthermore, by leveraging the randomized ordering of questions, I identify an attention capture effect on directly subsequent questions, providing evidence that the attention capture effects of poverty affect policy relevant outcomes outside of experimental settings.

[1]Ph.D. Candidate at the Department of Agricultural and Resource Economics, UC Berkeley. Email: claire.duquennois@berkeley.edu. 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 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 may be 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. Students of the same ability, but from different backgrounds, are known to respond differently to questions,[2] though there is limited understanding as to why (Freedle (2010), Editors (2010)). 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.[3] 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 as well. I exploit the natural variation in the financial salience of mathematics exams that is generated from monetary questions. I begin by documenting that disadvantaged students differentially underperform on mathematics exams and assignments when they feature a larger share of monetary themed questions. I observe this result in three different datasets spanning three different contexts: a homework platform in the US, an international cross-country standardized exam and a national educational assessment exam in Mexico. Using data from the two examinations, I find that a 10 percentage point increase in the financial salience of the exam depresses the performance of students with socio-economic status (SES) indicators below the national median[4] by 0.026 and 0.038 standard deviations depending on the context. 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, is largest for the most disadvantaged and is responsive to

[2]Differential performance by different ethnic and socio-economic groups has been documented on the SAT for instance. A proposed alternative scoring mechanism could shrink the performance differential between white and African-American test takers by a third (Freedle (2003), Santelices and Wilson (2010)).

[3]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 6 and figures 12 and 13 in the appendix.

[4]The datasets I use feature different SES indicators: parental education levels, a school marginalization index and the share of students in a school receiving free or reduced price lunch.

income shocks caused by rainfall. Furthermore, evidence from the homework platform shows that acquiring a mathematical skill requires differentially more time and effort for disadvantaged students when it is practiced using monetary themes.

There are a number of reasons why we might expect lower income students to perform differentially on these types of exercises. The literature on poverty and 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. I investigate an attention capture mechanism that draws on recent experimental findings from this literature. I identify the effect of attention capture by matching monetary themed questions to similar non-monetary themed questions, and by exploiting the randomized ordering of questions on the homework platform. This analysis of the itemized question level response data shows evidence of an attention capture effect. Comparison of student responses on monetary themed questions to highly similar non-monetary themed questions provides evidence that disadvantaged students underperform on questions that feature a monetary theme. Furthermore, by leveraging the randomized ordering of questions in the homework data, I observe a pattern of underperformance on questions that are placed subsequent to a monetary themed question. This pattern is consistent with an attention capture effect on subsequent questions for the low SES students and manifests in the other itemized exam data as well. A relationship between poverty and cognition has been observed in experimental settings using psychological tests. These findings show policy relevant impacts on student performance using real homework and examination scores.

The proposed mechanism draws heavily from recent ideas in the psychology of poverty literature regarding the relationship between cognitive functioning and poverty. This literature has suggested that poverty captures attention, generates intrusive and distracting thoughts that reduce an individual's cognitive resources (Mani et al. (2013), Shah et al. (2012), Shah et al. (2018), Tomm and Zhao (2016)). Though hard to differentiate, several mechanisms have been investigated. 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. This mechanism has been tested in a number of lab and field experiments (Mani et al. (2013), Shah et al. (2012), Spears (2011), Kaur et al. (2019)). 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. There have been a number of works evaluating the relationship between poverty and stress.

Haushofer and Fehr (2014) provide an extensive review of this literature, concluding that the majority of findings support a causal link. The impacts on cognition, however are not as well established, with some contradictory results (Mani et al. (2013), Carvalho et al. (2016), Kaur et al. (2019)). A particular challenge to identifying this mechanism is the difficulty in using actual income variation, as it correlates with changes in nutrition which are known to generate cognitive effects (particularly for children) even in the short run (Anderson et al. (2018), Gassman-Pines and Bellows (2018)). The mechanism that 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, the fact that this distraction occurs precisely when a low income individual is required to make potentially cognitively demanding decisions about financial resources makes any such effect important to understand for scholars who study decision making in the context of poverty.

This mechanism is reminiscent of the stereotype threat effect first posited by Steele and Aronson (1995), who suggested that an individual's performance on an examination is sensitive to priming about a stereotype of their group. This hypothesis has generated a significant amount of research, primarily in lab and field-lab settings (Spencer et al. (2016), Fryer Jr et al. (2008)). Empirical challenges and research preferences within disciplines, has limited field research on stereotype threats[5] and on poverty's effects on cognitive functioning. By utilizing real examination data, I address this gap in the literature and alleviate concerns of experimenter demand effects. I also remove concerns about 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 am able to estimate the effects of the tested mechanism under normal exam and homework conditions and show that the experimental results on the cognitive effects of poverty have external validity beyond the experimental setting. Though the effects on attention may be temporary, the impacts are economically meaningful because exam scores are frequently used to determine important economic opportunities such as eligibility for further education, placement in schools or access to scholarships. Furthermore, the homework effects I find imply impacts on the entire learning process. In this regard, I am addressing a gap in the cognitive functioning literature by investigating real costs and showing that effects that have thus far been measured using psychological tests also impact exam scores, a policy relevant metric.

[5]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.

The rest of the paper is organized as follows: Section 2 describes the three primary datasets used in my analysis. Section 3 presents estimation methods and results on aggregate exam and assignment performance, using exam level variation in financial salience. Section 4 investigates potential mechanisms using itemized question level data, providing evidence of an attention capture effect. 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

This paper uses data from three different sources to provide evidence that the share of monetary questions featured on an exam or assignment differentially affects the performance of low socio-economic status (SES) students. In addition to confirming the replicability of this result across a variety of contexts, each of these three datasets has distinct attributes allowing for a more thorough understanding of the mechanisms behind the general result. The ASSISTments homework platform in the US allows me to show effects on learning and effort and to exploit the randomized ordering of questions to identify attention capture effects. The cross-country Trends in International Mathematics and Science Study (TIMSS) exam provides examination setting evidence and shows that these results on exam performance and attention capture are widely generalizable. The Mexican Evaluación Nacional de Logros Académicos en Centros Escolares (ENLACE) exam provides evidence from a more traditional examination setting and allows me to exploit the panel nature of the data to show that the effects on examinations respond to income shocks. Table 1 summarizes the key attributes of each of these datasets.

2.1 Homework Platform Micro Data: ASSISTments

ASSISTments is a free online homework platform in the US operated by the Worcester Polytechnic Institute's Computer Science Department. Teachers create accounts on ASSISTments and then use the platform to assign homework to their students. Teachers can generate their own problems sets or use existing material. The most widely used format on ASSISTments is called 'skill builders.' Skill builders consist of a large pool of questions meant to practice a specific skill. When assigned a skill builder, students are expected to 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 an-

swering 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 skill builder.

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. 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,[6] the number of hints requested, and whether they completed the skill builder by answering three questions in a row correctly. ASSISTments data does not include any socio-economic indicators, though ASSISTments made an exception and agreed to match the schools in their user pool to National Center for Educational Statistics (NCES) data in order to supply me with de-identified school level SES indicators. These include school enrollment and a count of students enrolled in free and reduced price lunch programs.

I use the data for the 13 skill builders that feature both monetary and non-monetary themed practice questions. The main sample consists of 23,208 different student assignments covering thirteen different skill builders, featuring 1690 questions, of which 519 are coded as monetary themed.[7] Figure 1 shows that there is significant variation in the proportion of monetary questions featured on student assignments 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.

The ASSISTments data has the distinct advantage of random question ordering. This is key to identification of attention capture effects on subsequent questions, because this alleviates the concern that systematic placement of questions may be impacting estimates. Furthermore, the ASSISTments data provides insight into the learning process which can shed some light on the performance gaps that become evident in the examination data.

2.2 Cross-Country Exam Micro Data: TIMSS

The Trends in International Mathematics and Science Study (TIMSS) is an international standardized test in math and science administered by the International Association for the Evaluation of Educational Achievement (IEA) to a random sample of 4th and 8th

[6]To focus on students who are actually engaged in completing the assignment, time spent on a question is coded as NA if the student spends more than 8.8 minutes (the 90th percentile) or less than 5 seconds on a question. Outlier attempt counts beyond 8 attempts (the 90th percentile) are also coded as NA.

[7]Cleaning primarily involved limiting the sample to student assignments for which the SES indicator is observed and the monetary indicator is defined for all questions. Furthermore, though rarely exercised, teachers have the option of fixing the ordering of questions. For each assignment, I test that monetary questions are not correlated with a particular sequential positioning and drop any assignments where this correlation is significant at the 10% level. I also drop any skill builder that features multiple part questions.

graders in participating countries.[8] [9] 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. 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.

The advantage of the TIMSS data is that it features question level responses so that I observe student answers to each question on their exam. In addition to student responses, student, teacher and school surveys are also administered. Importantly, since 2011, most countries also administered a parental questionnaire for the 4th grade exam which reports basic occupational and educational categories of the parents.[10] 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. Most of these countries are middle to high income (see figure 4).[11] For this dataset I opt to use the highest reported parental education category as my primary SES indicator.[12] Furthermore, since SES may be associated with different education levels in different national contexts, I also generate an indicator variable for whether the highest parental education reported for a student falls

[8]TIMSS exam design, sampling and implementation is executed in coordination with participating countries via country representatives and national statistical organizations. For instance, the NCES, part of the U.S. Department of Education, is responsible for the collaboration and implementation of TIMSS in the US. In collaboration with TIMSS sampling experts, participating countries define their national target population, apply the TIMSS requirements to construct the country's sampling frame, and select a nationally representative sample of schools and students (see LaRoche et al. (2016)).

[9]For brevity I use the term country, though the IEA also works with regional authorities that wish to benchmark their performance.

[10]The 8th grade TIMSS data also provides these indicators but they are elicited from the students. I opt to use the 4th grade data out of concern that 8th grade student misreporting could correlate with exam performance.

[11]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.

[12]Occupational categories are more difficult to compare and interpret given the cross-country nature of this data. Nonetheless, results using highest family occupational category are broadly similar and reported in table 19 of the appendix.

below the national median as observed in the TIMSS data.[13] 469,849 students have taken the examination over the two rounds of 2011 and 2015. My main sample will consist of the 379,468 students for whom parental education is available. There is selection into the main sample due to non-random parental non-response. Nevertheless, because treatment is random within the sample, estimates are internally valid.[14]

Each year, a student taking a TIMSS exam is assigned one of 14 possible booklets. Each booklet consists of three components: a mathematics and a sciences section, followed by the student survey, all of which are separated by short breaks. For my estimations I will focus exclusively on the mathematics section of the exams. 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. 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. Prompt order within a prompt block does not vary. Among other goals, the TIMSS exams are designed to measure time trends in learning outcomes; therefore, eight blocks of mathematics questions get re-administered 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. Figure 2 illustrates the structure of the 14 TIMSS booklets that could be handed to a student in a given year. For example, a student handed booklet 1 would first complete their math section, which would consist of prompt blocks M01 and M02, and then move on to their science section after a short break. A student handed booklet 2 would complete their science section first and after a break complete their math section consisting of prompt blocks M02 and M03.

I do not observe the exact text of most of the prompts.[15] Information is available on each prompt, including some prompt characteristics such as the answer type (completed response or multiple choice),[16] topic area and cognitive domain and a brief thematic de-

[13]Because 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.

[14]My estimation exploits the fact that the random assignment of booklets to students is orthogonal to parental non-response, as demonstrated by the first column in table 18. Columns two through four in table 18 show 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.

[15]TIMSS readministers prompts across examination waves and thus does not releases the full set of prompts that were used.

[16]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. For simplicity I do not count partially correct answers as correct.

scriptor. 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). Furthermore, I also flag the four directly subsequent non-monetary questions as 'post' questions to track persistence of effects.

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 question in their booklet.[17] Out of 708 questions, 44 are flagged as monetary questions and feature 14 unique prompts. Figure 2 shows that there is variation in the proportion of monetary questions featured in the booklets as well as variation in the reported parental education categories, the two key sources of variation I exploit in this dataset.

2.3 National School Panel: ENLACE

The Mexican Evaluación Nacional de Logros Académicos en Centros Escolares (ENLACE) exams were administered throughout the country each June from 2006 to 2013. While ENLACE started out as a low stakes test, ENLACE results were broadly diffused, becoming one of the main metrics for school performance and eventually being linked to teacher salary bonuses (Vivanco (2013), de Hoyos (2014)). ENLACE was eventually discontinued because the growing performance incentives, combined with lack of implementation oversight, led to concerns about cheating.

School level subject results for all tested grades in all schools in Mexico are publicly available. The data also includes the school's marginalization index (1 to 5)[18] as defined by Mexico's National Population Council.[19] 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. Figure 3 show that there is variation in the proportion of monetary questions featured on exams within each grade. Figure 3 also shows

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

[18] Although the marginalization index does not change over time for most schools, there is some year on year variation. I opt to treat this index as time invariant, calculating the average for each school and rounding to the closest index category.

[19] Mexico's National Population Council (CONAPO) calculates 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, and electricity, dirt floors, household overcrowding, geographic isolation, and low incomes in employment. Further details are available at http://www.conapo.gob.mx.

the variation in the marginalization indicator across schools, the other source of variation necessary for my estimation.

I use a panel of school performance for 135,307 different schools between the years 2009 and 2013.[20] While the ENLACE data has the disadvantage of not being at the question level, nor even at the individual level, the panel structure presents certain advantages. In particular, by incorporating additional data, it allows some insight into how estimated effects respond to income shocks.

Rainfall has been shown to generate income shocks in the Mexican context (Munshi (2003)). I obtain the coordinates for Mexican municipalities and match these to rainfall data.[21] I use data from the Tropical Rainfall Measuring Mission (TRMM) [22] to calculate a drought indicator for each examination year. The drought indicator is set to one if the cumulative rainfall in the previous agricultural season (July-February) falls in the lowest decile of a locality's rainfall realizations between 1998 and 2018.

3 Impacts of Financial Salience on Aggregate Performance

I begin by presenting student level estimations that look at how the variation in the proportion of monetary questions featured on an exam or assignment impacts effort and performance. I find that more financially salient exams differentially depress the exam scores of lower SES students and that this effect is responsive to income shocks. I also identify that lower SES students have to exert differentially greater learning effort when faced with more financially salient homework assignments.

3.1 Impacts in Examination Settings: TIMSS

For the estimation using the TIMSS data, I exploit the random assignment of test booklets to students and the variation in the number of monetary questions between booklets. Columns 1 through 4 of table 3 provide a randomization check, confirming that within a

[20] 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. Finally, 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 also dropped from the final dataset.

[21] Municipality coordinates are available from the Instituto Nacional de Estadística y Geografía (INEGI).

[22] Specifically, I use the TRMM Multi-Satellite Precipitation Analysis (TMPA) Rainfall Estimate Product 3B43 Version 7, which merges satellite and gauge data to generate a monthly estimate on a $0.25°$ by $0.25°$ spatial resolution.

year there is no correlation between a student's SES categories and the share of monetary questions in the booklet they receive, overall and within a classroom.

The effect on a low SES student of receiving a financially salient booklet is estimated as follows,

$$E_{ib} = \Theta_1 + \Theta_2 LowP_i + \Theta_3 LowP_i * PM_b + \kappa_b + c_i + \epsilon_{bi}, \quad (1)$$

$$E_{ib} = \theta_1 + \sum_{p=2}^{5} \theta_{2p} P_i + \sum_{p=2}^{5} \theta_{3p} P_i * PM_b + \kappa_b + c_i + \epsilon_{bi}. \quad (2)$$

I regress standardized exam scores (E_{ib})[23] 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 (b). For the TIMSS estimations, I use an indicator for whether the reported parental education category is lower than the national median $(LowP_i)$ as observed in the TIMSS data in equation 1 or parental education category dummies(P_i) as specified in equation 2. I also include booklet fixed effects (κ_b) and country or class fixed effects (c_i) as controls.

Results are reported in table 3. Estimates in columns 5 and 7 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 6 and 8 show that this effect is negatively related to parental education, with the largest effect for the most disadvantaged students. Columns 7 and 8 include classroom fixed effects. The addition of classroom fixed effects does not significantly change the magnitudes of the θ_3 coefficients of interest, although the overall variation in performance due to parental education levels as estimated by θ_2 is significantly smaller within a classroom than within a country. This is likely due to selection across schools.

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.

[23] I use crude exam scores calculated as the mean performance on the questions in the question level data.

3.2 Impacts in Examination Settings: ENLACE

I apply a similar estimation approach to the ENLACE exam data,

$$E_{sgy} = \Theta_1 + \Theta_2 LowZ_s * PM_{gy} + \kappa_{gy} + \tau_{sy} + \rho_{gs} + \epsilon_{sgy}, \quad (3)$$

$$E_{sgy} = \theta_1 + \sum_{z=2}^{5} \theta_{2z} Z_s * PM_{gy} + \kappa_{gy} + \tau_{sy} + \rho_{gs} + \epsilon_{sgy}. \quad (4)$$

I regress the standardized school average[24] 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}). Here, SES indicators include an indicator for whether a school's marginalization index falls below the national median $(LowZ_s)$ in equation 3 or the school marginalization index dummies (Z_s) as specified in equation 4. I include a grade by year fixed effect (κ_{gy}) to control for overall difficulty of each particular exam booklet, school by year fixed effects (τ_{sy}) to control for local shocks that might affect overall performance in a school, and grade by school fixed effects (ρ_{gs}) to control for time invariant performance of a grade in a school.

Results are reported in table 4 and are qualitatively 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 figure 3, 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 0.038 standard deviations and up to 0.126 standard deviations in very disadvantaged schools. The overall performance gap between above and below median schools is 0.265 standard deviations. Thus the effect of a 10 percentage point increase in monetary salience represent about 14% of the overall performance gap.

3.3 Impacts in Examination Settings: Response to Income Shocks

In addition to confirming the TIMSS results, the ENLACE data has the advantage of being a school level panel, allowing me to observe how the estimated effects respond to fluctuations in income. In the context of Mexico, drought conditions have been shown to generate economically significant income variation (Munshi (2003)). To consider whether annual income variation impacts this effect, I add the relevant interaction terms with the drought indicator (D_{sy}),

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

$$E_{sgy} = \delta_1 + \delta_2 LowZ_s * PM_{gy} + \delta_3 LowZ_s * PM_{gy} * D_{sy} + \delta_4 PM_{gy} * D_{sy} + \kappa_{gy} + \tau_{sy} + \rho_{gs} + \epsilon_{sgy}. \tag{5}$$

Results are reported in table 5. Student performance fluctuations in response to income shocks are consistent with the hypothesis that income scarcity amplifies the negative effect of monetary questions on exam performance. δ_2, reported in the first row, shows that students in below median schools perform worse on exams that feature a higher percent of monetary questions. δ_3, reported in the second row, shows that this negative effect on exam scores is amplified, such that it more than doubles in magnitude in below median schools during drought years. Thus, in a drought year, facing an exam with 10 percentage points more monetary questions differentially depresses the performance of students in below median schools by 0.077 standard deviations. As one might expect, δ_4 is small and insignificant, as droughts do not affect performance on monetary questions in above median schools.

3.4 Impacts on Learning: ASSISTments

Analysis of the TIMSS and ENLACE data presents evidence that monetary questions differentially depress the exam performance of lower SES students. These impacts are of concern because examination performance often determines educational and economic opportunities. Yet, while the estimated effects are non-negligible, the underlying performance gap that exists between high and low SES students is substantially larger.

In this section, I present evidence that monetary questions may also contribute to this underlying performance gap through their impact on learning. Open any elementary school level math textbook and you will invariably find monetary themed examples being used to teach mathematical concepts. Thus, the same mechanism that depresses exam performance may also affects learning and skill acquisition. Using the user data from the ASSISTments homework platform, I find that lower SES students must exert differentially more effort to complete an assignment when their assignment features a greater proportion of monetary themed questions.

When assigned a skill builder by their teachers, students must log in and answer randomly selected questions from the skill builder's question pool until they answer three correctly in a row, at which point the system registers that they have mastered the assignment. For each student, I calculate the proportion of monetary questions they faced on their assignment as well as the mean number of attempts and hints they requested per question and the total time spent on the assignment.

In the following estimations, I use the assignment (a) data for all students (i) with complete question level data to estimate the effect of the proportion of monetary questions on several different dependent variables (Y_{ia}),

$$Y_{ia} = \alpha_1 + \alpha_2 PFR_s * PM_{ia} + \alpha_3 PM_{ia} + c_{ac} + \epsilon_{ia}. \quad (6)$$

I am interested in α_2, the interaction between the proportion of students in the school receiving free or reduced price lunch (PFR_s) and the proportion of questions the student faced that features a monetary theme (PM_{ia}), controlling for the proportion of monetary themed questions, as these are on average easier, and an assignment by class fixed effect (c_{ac}), which captures the general performance of students in that class on the assignment.

I first run a conditional logit to estimate the likelihood of mastering the assignment. I also estimate the equation above as a linear probability model. Next, I restrict the data to students who master the assignment and are actively engaged throughout the assignment [25] and run the same estimation on other dependent variables that capture learning effort such as the total time spent on the assignment, the number of questions the student answered, the mean number of hints they requested per question and the mean number of attempts they made on each question.

Results are presented in table 6. The estimates of α_2 using the conditional logit and the linear probability model reported in columns 1 and 2 are not statistically significant. Though the coefficients move in the hypothesized direction, I cannot reject that the proportion of monetary questions a student faces on their assignment has the same effect on the likelihood of mastering the assignment for students in advantaged and disadvantaged schools. However, the remaining columns in table 6 do present evidence that in order to achieve mastery on the assignment, students in disadvantaged schools have to differentially exert more learning effort when faced with an assignment featuring a higher share of monetary questions.

The coefficients in the first and second row suggest that any learning benefit monetary questions have for students in more advantaged schools is smaller, and even a disadvantage, for students in lower income schools where more students receive free and reduced price lunches. The mean value of the percent of students in the school receiving free or reduced price lunch is 0.252 while the minimum value is 0.012 (close to $PFR_s = 0$) and the maximum is 0.946 (close to $PFR_s = 1$). When $PFR_s = 0$, a 10 percentage point increase in the proportion of monetary questions a student faces decreases the time spent on mastering the assignment by 7.05 seconds (-0.020 sd). For the mean school where 25% of

[25] I define active engagement as students whose time spent on each assigned problem falls between 5 seconds and 8.8 minutes (the 90th percentile).

students receive free or reduced price lunch, assignment time is decreased by 1.8 seconds. However, in a school where all students receive free or reduced price lunch, much like the most disadvantaged school in the data, $PFR_s = 1$ and assignment time would increase by 13.75 seconds (+0.04 sd) over a mean value of 330 seconds. A similar pattern holds for the other measures of learning effort. A 10 percentage point increase in the share of monetary questions faced by a student reduces the number of questions needed to complete the assignment by 0.063 (-0.023 sd) questions if $PFR_s = 0$, 0.032 at the mean and increases the number of questions to mastery by 0.063 (+0.023 sd) if $PFR_s = 1$. Similarly, the mean number of attempts made on a question is reduced by 0.01 (-0.02 sd) when $PFR_s = 0$, 0.005 at the mean and increases by 0.012 (+0.024 sd) when $PFR_s = 1$. Finally, a 10 percentage point increase in the share of monetary questions faced by a student reduces the mean number of hints requested by 0.016 (-0.021 sd) if $PFR_s = 0$, 0.006 at the mean and increases the mean number of requested hints by 0.024 (+0.0323 sd) if $PFR_s = 1$. Thus, when comparing two classmates completing the same assignment, the student who randomly faces a larger share of monetary questions does not need to exert as much effort to complete the assignment in the wealthier schools. By contrast, in the more disadvantaged schools, the advantage presented by a larger share of monetary questions is reduced and even a net disadvantage.

These estimates are consistent with the findings using the TIMSS and ENLACE data. The results on time suggest that monetary questions would impact lower income students' performance when placed under a time constraint, as is common in examination settings. These effects on learning would impact much of the learning process in early mathematics education. This suggests that, while these monetary themed questions may present certain pedagogical advantages, these advantages are not evenly distributed and are even a disadvantage for the most vulnerable students, creating an unrelenting drag on their learning process. Over the course of an education, the few seconds of extra effort on each 6 minute assignment would certainly add up into a non-negligible effort cost that would contribute to explaining the underlying performance gap in educational outcomes between low and high SES students.

4 Identifying Attention Capture

While the evidence presented so far does suggest a differential impact of monetary questions on exam performance and learning effort, it does not clearly identify the mechanisms or show evidence of an attention capture effect. Monetary questions may differ from other mathematics questions. They may be used to 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. In the following sections, I exploit the itemized question level responses of the ASSISTments and TIMSS datasets to present evidence of an attention capture mechanism. I begin with an analysis of the itemized ASSISTments data where I exploit the precise thematic content of skill builders and the large number of questions to conduct a matching exercise to investigate the possibility that effects are driven solely by monetary questions being used to test different skills. Next, I exploit the random ordering of questions to clearly identify attention capture effects by looking at the lagged effects of monetary questions on subsequent questions. Finally, I show evidence that these effects are generalizable, as similar results are observable in the itemized TIMSS data.

4.1 Controlling for Question Characteristics: Matching

Figure 5 shows how question level performance and effort metrics vary by the percent of students on free and reduced price lunch in a school based on whether a question is monetary themed or not. These plots suggest that at all levels of free and reduced price lunch shares, the monetary questions in the ASSISTments skill builders are easier for students. However as the estimates in section 3.4 indicate, this advantage 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.

Figure 5 is constructed using all of the questions in the ASSISTments data. An important concern may be that monetary questions are used to test a very different set of mathematical skills in which low SES students are disadvantaged. For instance, these questions may be more likely to test numerical operations rather than geometric reasoning. While this is undoubtedly the case in most settings, including in the ENLACE and TIMSS exams, it is worth noting that the ASSISTments skill builders are very narrow in thematic content, as teacher use them to practice very specific mathematical skills such as 'Writing a Linear Equation from a Situation', 'Finding the Whole from the Percent and Part in a Word Problem' or 'Percent Increases and Decreases'. Nevertheless, 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. To address this

concern, within each skill builder I match monetary themed questions to almost identical non-monetary questions. Questions are matched if they are formulated similarly and involve the application of the same mathematical process. Figure 6 shows two examples of matched monetary and non-monetary questions.

Figure 7 plots the performance metrics by the share of students receiving free or reduced price lunch and monetary theme for the matched sub-sample. Note that restricting the data to matched questions significantly reduces sample size from 133,997 to 33,295 question observations. Nonetheless, figure 7 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 more formally estimate the difference between matched monetary and non-monetary questions, I estimate the following,

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

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. I include a school by matched question group fixed effect ms_{qi}. 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 thematic content. Question level performance metrics include whether the student answered the questions correctly,[26] how many hints were requested, how many attempts were made, and the time spent on the question.

Results are reported in table 7 and reflect the pattern observed in figure 7. Although the results are somewhat under-powered in this small sample, estimates of the differential, γ_2, are significant at the 5% level for hint requests and at the 10% level for completion time. The coefficients on answering correctly and attempts are consistent with underperformance and increased effort on monetary questions, although they are not statistically significant at conventional levels. Estimates for γ_3 are insignificant except for completion time. Thus, when a school has a low free and reduced price lunch rate, students within the

[26] ASSISTments does allow for partial credit; however, most of the data is either a 0 or 1.

same school perform similarly on monetary questions (though they require 11.42 seconds more time) as compared to their performance on almost identical non-monetary themed questions. However in more disadvantaged schools, compared to their performance on almost identical non-monetary themed questions, students within the same school request more hints and require even more time. In a school where all students received free or reduced lunch, students experience an additional disadvantage as compared to students in wealthier schools, and request an additional 0.518 (+0.24 sd)hints, and spend an additional 23.16 seconds (0.26 sd) (in addition to the additional 11.42 seconds experienced in all schools) on a monetary question as compared to their schoolmates answering a matched non-monetary question. This is evidence that underperformance on monetary questions by low SES students cannot be fully explained by the possibility that monetary questions require a different set of mathematical or question answering skills beyond those implied by their topical content.

4.2 Evidence of Attention Capture

The evidence presented in table 7 shows that monetary themed questions present a greater challenge to students in disadvantaged schools. This evidence, however is insufficient to clearly identify attention capture. Students in lower income households may not have as many opportunities to apply mathematical skills to monetized situations. For instance, they may be less likely to receive an allowance with which they can make purchases or they may be less likely to be put in charge of making small purchases in a shop or market where they must collect change. This explanation could lead to underperformance on monetary questions and generate the pattern of results in table 7 and in the aggregate effects estimated in section 3. Attention capture cannot be disentangled from this possible explanation by looking only at performance on the monetized questions.

If non-negligible, the attention capture effect can be identified by looking at performance on subsequent questions. These questions are not monetary themed but are potentially affected by the attention capture effect generated by the preceding monetary question. The randomized ordering of questions in the ASISSTments data can be exploited to identify whether there is such a lagged performance effect on subsequent questions for low SES students, as random ordering alleviates any concerns that question placement might be based on question unobservables.

I leverage this randomized ordering to identify the 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-monetary

question. To avoid having to consider the effects of repeated exposures and selection as students complete their assignments, 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.[27] I estimate the following,

$$Y_{iq} = \beta_1 + \beta_2 PFR_s * Post_{iq} + \beta_3 Post_{iq} + m_{q_{pre}} + \nu_{qs} + \epsilon_{iq}. \quad (8)$$

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 a fixed effect for the leading matched group of questions $(m_{q_{pre}})$ and a question by school fixed effect (ν_{qs}). 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.

Results are reported in table 8 and are consistent with an attention capture effect. Compared to their peers answering the same question, students in a school where everyone receives free or reduced price lunch request 0.389 (+0.15 sd) more hints and spend an additional 27 seconds (+0.28 sd) on the question if it follows a monetary themed question rather than a similar non-monetary themed question. The β_2 coefficients on answering correctly and attempts are also consistent with an attention capture effect, though not statistically significant. This stands in sharp contrast to the effect of monetary questions on students in more advantaged schools as estimated with the β_3 coefficients. These 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 more advantaged schools.

4.2.1 Attention Capture versus Cognitive Fatigue

Because of the randomized question order in the ASSISTments data, the lagged effect of a monetary question on subsequent questions must be due to their positioning relative to a monetary question. There is a possible alternative mechanism to the attention capture explanation. If low SES students find monetary questions differentially difficult, this might affect their performance on subsequent questions if they are differentially fatigued when they face them. This explanation could have the same implications for exam and assignment performance as estimated above but the underlying explanatory mechanism

[27] More formally, let S_{qi} be the position of question q in student i's sequence of questions. S_{M1i} and S_{M2i} are the positions of the first and second matched questions faced by student i. I subset the data to observations where $S_{M1i} < S_{qi} < S_{M2i}$ and $S_{qi} \leq (S_{M1i} + 4)$

would be subtly different.

To distinguish attention capture from fatigue effects, I create a measure of lagged difficulty that is adjusted for different SES groups. I divide the students by quartile based on the share of students in their school receiving free or reduced price lunch. I 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 control to the estimation strategy used in equation 8.

Table 9 presents the results of this estimation. The positive and significant coefficient in the third row suggests that students do spend a little bit more time on questions that follow questions that were differentially more difficult for them. This effect is small, however, and controlling for it does not does not meaningfully alter the coefficients or significance levels of the β_2 coefficients of interest, supporting the hypothesis that an attention capture mechanism is driving these results.

4.2.2 Attention Capture versus Stalled Learning

If low SES students struggle with monetary themed questions, they may not benefit as much from practicing using these questions, compared to similar non-monetary questions. Thus low SES students answering questions that follow a monetary themed question have not received as much effective practice as their peers who answered a similar non-monetary question. This could plausibly generate results similar to those in table 8, where the depressed performance of low SES students after monetary questions is simply due to having had one less practice question than their schoolmates.

To consider this possibility, I add a sequence control to the estimation strategy used in equation 8. Results are presented in table 10. The coefficients in the third row show that, as a student proceeds through a homework assignment, the likelihood that they answer a question correctly increases with each problem and the effort they must expend on each problem is reduced, in that they require fewer hints, make fewer attempts and spend less time on each subsequent question. If low SES students receive no learning gains from the monetary questions, the sequence position of the subsequent questions is effectively reduced by one. This could result in β_2 coefficients in the first row that offset the sequence coefficients in the third row.

Inspection of the relative magnitudes of the estimates in the first row as compared to the estimates in the third row shows that the depressed performance of low SES students is substantially larger than the effect of having answered one less previous practice question. Thus, underperformance following monetary themed questions truly is underperformance and not simply lack of improvement.

This evidence supports the attention capture hypothesis rather than a stalled learning effect. This is further supported by the evidence presented in the next section. Indeed, the TIMSS examinations are not designed as learning tools and question topics and themes can drastically change from one question to the next. In this setting, we would not expect improvement from learning as students proceed through the examination, ruling out the stalled learning explanation for the evidence we will consider next.

4.3 Confirming Attention Capture in Cross-Country Examination Data

The randomized ordering used on ASSISTments skill builders had the distinct advantage of allowing the estimation of the effect on subsequent questions without the concern that placement after a monetary question may correlate with some question unobservable that differentially depresses low SES performance. Nevertheless, ASSISTments is a small platform used primarily in the United States. In the following sections, I examine the TIMSS exam data to consider whether this effect is generalizable to an examination setting and a cross-county dataset. I find results consistent with an attention capture effect in the TIMSS data as well.

4.3.1 Suggestive Evidence Using Question Aggregates

Before imposing structure on my estimation methods, a simple approach suggests that there is indeed a performance gap on monetary and subsequent questions in the TIMSS data as well. A simple regression on aggregate TIMSS question statistics suggests a pattern of underperformance on monetary and subsequent questions for lower SES students. For different groups of students, I estimate the following,

$$\bar{C}_{q,p<nm} = \Phi_1 + \Phi_2 M_q + \Phi_3 Post_q + \Phi_4 \bar{C}_{q,p>=nm} + \epsilon_{q,p<nm}, \tag{9}$$

$$\bar{C}_{q,p} = \phi_{1p} + \phi_{2p} M_q + \phi_{3p} Post_q + \phi_{4p} \bar{C}_{q,uni} + \epsilon_{q,p}. \tag{10}$$

For students whose parental education falls below the national median, I regress the mean performance on each question, as measured by correct answers, $(\bar{C}_{q,p<nm})$ on the monetary indicator (M_q), an indicator for non-monetary questions placed within four questions after a monetary question $(Post_q)$, and the mean performance of students with parental education above the national median $(\bar{C}_{q,p>=nm})$ to control for question difficulty. I repeat the same procedure for each of the parental education categories, p, other than university graduates, using the mean performance of students with university educated parents $(\bar{C}_{q,uni})$ to control for question difficulty.

Results are reported in table 11. Φ_2 estimates are negative and the magnitude of the penalty increases for lower parental education. Estimates for Φ_3 follow the same pattern, consistent with the hypothesized attention capture effect.

4.3.2 Confirming Attention Capture in Cross-Country Exams

While the above results are indeed suggestive, I can exploit the itemized TIMSS student level micro data to compare an individual student's performance on monetary and subsequent questions to their performance on other mathematics questions to see whether a pattern consistent with an attention capture effect is also present in the TIMSS data.

To do so, I estimate the following,

$$C_{iq} = \Lambda_1 + \Lambda_2 LowP_i * M_q + \Lambda_3 LowP_i * Post_q + \mu_q + \eta_i + \epsilon_{qi} \tag{11}$$

$$C_{iq} = \lambda_1 + \sum_{p=2}^{5} \lambda_{2p} P_i * M_q + \sum_{p=2}^{5} \lambda_{3p} P_i * Post_q + \mu_q + \eta_i + \epsilon_{qi}. \tag{12}$$

I regress an indicator for a correct response (C_{qi}) on the interaction between an 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. All specifications use student (η_i) and question (μ_q) fixed effects to control for student and question unobservables. 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.

Results are reported in table 12, columns 1 and 2. Students with above median educated parents (column 1) or with university educated parents (column 2) are the omitted categories. Both sets of Λ_2 and Λ_3 coefficients are of interest. As on the ASSISTments platform, lower SES students' performance is differentially depressed on monetary questions and the questions that follow them. Furthermore, the sets of λ_2 and λ_3 coefficients are negative and inversely related to parental education, indicating a larger effect for the most disadvantaged students.

Though the results in columns 1 and 2 are consistent with attention capture, it is important to clarify that, because question ordering on the TIMSS exam is not randomized across students, it is not possible to cleanly identify the attention capture effect. Indeed, many of the concerns that were mentioned in the previous sections also apply to the TIMSS data. Monetary questions may be used to test mathematical topics that are differentially difficult for low SES students. Disadvantaged students may have fewer

opportunities to apply their math skills to monetized situations. Importantly, unlike in the ASSISTments data, because question order is not random, we should further be concerned that subsequent questions may also systematically cover different mathematical skills or topics in a way that is unobservable. Though we cannot use the TIMSS data as conclusively as the ASSISTments data to identify attention capture effects – because of the non-random ordering of questions – there are a number of approaches that can be applied to improve upon the estimation strategy outlined in equations 11 and 12 and to appease some of the concerns above.

I begin by comparing question observables by question type. Because many TIMSS questions are not released, I do not observe all of the actual questions. TIMSS does, however, disseminate some information about each prompt on the exam, including a topical descriptor,[28] the question type of the prompt, the topical area the prompt is designed to test and the cognitive domain exercised by the prompt. Comparisons of observable prompt characteristics across the question indicators of interest are illustrated in figure 8. Unsurprisingly, it is clear that monetary questions 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, however, do not share the unifying monetary theme and have a distribution of question characteristics that is broadly similar to the other questions in the TIMSS booklets.

I use these question observables to augment the controls used in equation 11. I construct additional fixed effects designed to capture differential performance due to placement,[29] difficulty,[30] question type and topic.[31] Columns 3 and 4 of table 12 give estimates of equations 11 and 12 using these additional fixed effects. Adding the additional controls does not change the qualitative features of the estimates. There is a small impact on the magnitude of the estimated effect, which actually becomes more pronounced.

One may be concerned that the effect stems from differences in teaching patterns between more advantaged and disadvantaged classrooms. In columns 5 and 6, I add ad-

[28] Each prompt is labeled with a name that broadly describes the topic and theme of the prompt. Examples include 'Total number of people on a ship', 'Multiply 23 and 19', 'Cost of ice cream', or 'Stickers bought by Mr. Brown'.

[29] I include parental education by sequence fixed effects, where sequence is a constructed categorical variable indicating whether a question is featured in the first five questions of the exam, second five and so forth.

[30] I include 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.

[31] 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.

ditional fixed effects controlling for classroom performance on the questions of interest. These additional controls do not change the qualitative features of the estimates, though the magnitude of the estimated effect is slightly reduced. Thus the individual SES indicator of parental education still has explanatory power for performance on these questions, even controlling for class performance.

My preferred specifications are those in columns 3 and 4. Aggregated as in table 3, these estimates suggest that each additional monetary question featured on an 25 question exam would depress the score of a student whose parents have an education level below the national median by 0.193 percentage points or about 0.008 standard deviations[32] compared to students whose parents have an education level above the national median. Since each question represents about 4% of the mathematics section, this is equivalent to a 0.02 standard deviation decrease for a 10 percentage point increase in the share of monetary questions, which is consistent with the estimates in column 5 of table 3. As these exams can feature up to five monetary questions in a single booklet, this could amount to an exam score impact of between 0.04 and 0.05 standard deviations for students given a monetary intensive booklet who have below median parental education, and up to a 0.063 standard deviation decrease if their parents have a primary education or less.

I explore different combinations of the above fixed effects in table 20 of the appendix. Though 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. Table 21 in the appendix 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.

4.3.3 Monetary Questions as Events

It is possible to think of students proceeding though exams and encountering 'events' in the form of monetary themed questions and to graphically visualize these effects. As students in these different datasets often encounter multiple monetary questions, setting up the estimation as an event study is not entirely straightforward. By limiting the TIMSS data to the subset of booklets that only feature one monetary question or two consecutive monetary questions, so as to have clearly defined pre and post periods, an event study approach is possible. This subset of the data covers 19 of the 24 booklets that include mone-

[32]Since each question is worth approximately 4% of the exam score, using the estimates from column 5, I calculate the direct effect as $-1.206 * 0.04$ with an additional effect on four subsequent questions of $-0.906 * 0.04 * 4$ for a total of 0.193 percentage points or about 0.008 standard deviations as the standard deviation of the exam scores is 23.56.

tary questions. Figure 9 plots the differential performance on questions based on question placement relative to the monetary question in the booklet.[33] The figure clearly illustrates the sharp drop in performance on monetary question for students whose parental education is below the national median, and the continued effect on subsequent questions as well.

4.3.4 Attention Capture versus Cognitive Fatigue

The estimates in table 12 and figure 9 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 previous section. 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 13. 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.[34] I then estimate my preferred specification of equation 11 as estimated in column 3 of table 12.

The resulting pairs of $\hat{\Lambda}_3^{placebo}$ and $\hat{\Lambda}_2^{placebo}$ coefficients are plotted in figure 10. 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 results

[33]Figure 9 plots estimates for coefficients π_{2t} from the following estimating equation:

$$C_{iq} = \pi_1 + \sum_{t=-6, t\neq -1}^{10} \pi_{2t}(T_q = t) * LowP_i + \mu_q + \eta_i + ... + \epsilon_{iq} \tag{13}$$

Where T_q is a question's position relative to the monetary question. Additional fixed effects include the same fixed effects used in my preferred specification in column 3 of table 12: Below median by difficulty, below median by sequence, below median by country by question type and below median by country by question topic. Standard errors are clustered at the student level. Questions more than 5 questions prior to or 9 questions subsequent to the monetary event are binned together and coded as -6 and 10.

[34]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.

above. In addition to verifying the above results, looking at the correlation between the coefficient pairs can also help decompose the role of cognitive fatigue due to the differential difficulty of preceding questions 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.

$$\hat{\Lambda}_{3p}^{placebo} = \psi_1 + \psi_2 \hat{\Lambda}_{2p}^{placebo} + \epsilon_p \tag{14}$$

Results are reported in table 14 and plotted in figure 10. ψ_2 is indeed positive and 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 10, 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 29% of the estimated effect on subsequent questions.[35] I interpret the remainder as evidence of attention capture.

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 exam design on high stakes examinations puts vulnerable students at a disadvantage, these tests could aggravate socio-economic disparities in access to education. 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.

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

[35] Estimates in table 14 imply that $\mathbb{E}(\hat{\Lambda}_3^{placebo} | \hat{\Lambda}_2^{placebo} = -1.207) = -0.254$ or 29% of $\hat{\Lambda}_3 = -0.891$.

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 publicly available. Nonetheless, 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.

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.[36] Using the estimates in table 3 and official assessment statistics,[37] 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). 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 sections, I use my estimates from the TIMSS and ENLACE exams to generate counterfactual exam scores on Mexico City's high school entrance exam. I then perfectly replicate the placement algorithm used to allocate students to high schools across Mexico City. Finally, I use the counterfactual scores to simulate how the change in exam scores would affect student allocation.

5.1 Simulating Effects on Mexico City's High School Entrance Exam

In response to an inefficient high school enrollment process, a consortium of public schools in Mexico City known as the Comisión Metropolitana de Instituciones Públicas de Educación Media Superior (COMIPEMS) adopted a competitive centralized admissions process. All ninth graders wishing to attend one of these schools submit a ranked list of up to 20 high school programs and subsequently take a comprehensive standardized exam. After exams have been scored, students are ranked and assigned to schools according to

[36] 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%.

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

a serial dictatorship mechanism (see Abdulkadiroglu and Sonmez (2003)).

The COMIPEMS exams consists of 128 multiple choice questions covering multiple subjects including 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 student rankings of preferred high schools and performance on the COMIPEMS exam in the different subjects. I also observe parental education levels and current junior high school, which I match to the school marginalization levels reported in the ENLACE data.

I use two approaches to 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. The first uses the estimates derived from the ENLACE data in column 4 of table 4. These estimates suggest that a 10 percentage point decrease in the share of monetary questions should increase the mean score of students in very disadvantaged schools by 0.126 standard deviations. 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 binomial and add 1 point to the math COMIPEMS score of randomly selected students in very disadvantaged schools such that their aggregate performance on the mathematics section is improved in a manner consistent with the ENLACE estimate. I repeat the same procedure for students at each marginalization level using the relevant estimates.[38] Having simulated a new counterfactual mathematics score, I calculate their new counterfactual COMIPEMS score which I then use to generate a new counterfactual ranking of the students. The second approach is similar but uses parental education and employs the estimates from column 6 of table 3. The difference between actual and simulated counterfactual math scores and the mean difference between the actual and counterfactual rank for students in each SES category using these two different SES indicators is reported in table 15.

The effect of a 10 percentage point decrease in the share of monetary questions improves the performance of disadvantaged students on the mathematics portion of the exam. The gap in mean performance between the highest and lowest SES group is reduced by 3 to 20% on the mathematics section, depending on the SES indicator used.[39] Because the mathematics section covers only 20% of the exam, the effect on aggregate

[38] Exam scores are not adjusted for students with a missing SES indicator.

[39] Mexico City is relatively wealthy compared to much of Mexico and the vast majority of test takers are attending junior high schools that are considered very advantaged on the national scale.

exam scores is proportionally smaller.[40] Once ranked using the counterfactual scores,[41] the ranking of students in disadvantaged groups improves, at a cost to those in the more advantaged groups.

5.2 Mexico City's High School Placement Algorithm

Students in Mexico City are assigned to high schools according to a serial dictatorship mechanism based on their ranked exam performance and the list of preferred schools each student submits prior to taking the exam.

High schools first set the maximum number of students they will accept.[42] Students who fail to score above 30 or who fail to complete middle school are disqualified from attending high school. A computer program then proceeds through the ranked list of students, starting with the highest scoring student, and allocates each student to their top-ranked school with open seats remaining. If no seats remain at any of the schools listed by the student, the student is unassigned. After the first assignment process is complete, these students undergo a secondary selection process over several days that allocates unassigned students to the remaining open slots (Dustan et al. (2017)).

Following these rules, I replicate the placement algorithm used by the COMIPEMS's centralized admission system. Because I observe the school to which each student was actually assigned, I can verify that high school placement in Mexico City actually follows the rules described above.[43] My replication of the placement algorithm perfectly replicates actual student assignment when I use the students' true exam scores. I can thus accurately simulate student placements using my replicated algorithm and the counterfactual ranking based on the counterfactual scores in table 15.

[40] 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.

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

[42] 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 admissions for non-competitive schools.

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

5.3 Simulated High School Placements in Mexico City

Since so few schools are considered disadvantaged in Mexico City, I focus here on the results using parental education categories. Tables 16 and 17 present estimated impacts for the two years of data. Equivalent results using school marginalization levels are presented in tables 22 and 23 of the appendix. 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, as opposed to being ineligible, has the potential to significantly alter the course of their lives. As illustrated in table 16, 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 2.2% for households with below median parental education levels.[44] The most impacted group comprises students whose parents are primary educated; they experience a 2.9% reduction in ineligible students. Overall, an additional 128 students pass this cutoff using the counterfactual exam scores. Having additional low SES students pass this threshold cutoff score is the most straightforward effect of the policy simulation and likely the most meaningful effect in terms of improving educational opportunities for low SES students.

In addition to the impact on high school eligibility, the counterfactual scores also change which students get assigned to a high school that they requested.[45] Because more students are able to meet the 31 point cutoff, the total number of assigned and unassigned students both increase, though there is significant heterogeneity across parental education groups. Unlike the requirement of meeting the 31 point cutoff, 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. Overall, students from higher SES groups are more likely to remain unassigned and less likely to receive an assignment in the simulation. By contrast, in the more disadvantaged groups, students are more likely to get assigned and less likely to remain unassigned.

Finally, the counterfactual scores also change whether students get to attend a more highly preferred school as summarized in table 17. Overall, students who received an assignment using both the real and counterfactual scores, on average get assigned to slightly less preferred schools. This is not surprising, as more students are passing the 31 point threshold, generating more competition and some displacement. Here again,

[44] The median parental education level in this data is Upper Secondary.

[45] Recall that not receiving an assignment means that the student did not score 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.

heterogeneity is important. Students with highly educated parents experience the bulk of this negative effect. For the children of university educated parents, since their exam scores are unaffected, it is rare that they get assigned to a more preferred school in the simulated data while 111 are displaced into less preferred schools and 30 become unassigned. Conversely, students with the least educated parents are on net more likely to receive a preferred school assignment: 308 receive a preferred assignment while 180 receive a less preferred assignment. Furthermore, in this group 143 of previously unassigned students receive an assignment, while only 48 become unassigned. 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.

Overall, 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 20% of a student's score. In contexts where mathematics is weighted more heavily, impacts could be substantially larger.

6 Discussion and Conclusion

Every year, millions of people around the world take examinations that have the potential to 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 a relatively 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 actually 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 and responds to negative income shocks. Furthermore, the evidence suggests that a similar pattern holds for homework assignments. Lower SES students must exert differentially greater effort and spend differentially more time completing homework assignments when they feature a larger share of monetary questions. As monetary questions are a common tool in the instruc-

tion of mathematics, this potentially affects much of the learning process in mathematics education. 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? It depends on what the ultimate teaching goals are. It is clear that lower SES students face a disadvantage when confronted with these topics; however, 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, then 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.[46] 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.[47]

Beyond the implications for educational testing, I present non-experimental evidence of attention capture due to poverty, and show that it affects 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 be activated 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

[46] In its statement on testing fairness, the Educational Testing Service in the US frequently discusses the idea of 'construct-irrelevant variance', differences between test takers' scores that are caused by factors other than differences in the knowledge, skills, abilities, or traits the test is intended to measure (ETS (2014)). Thus, if a question or test is intentionally designed to test applications to monetary themes, differential performance would not be considered unfair.

[47] In fact, figure 3 shows that monetary themed questions become less common in higher grade examinations, likely because the subject becomes more abstract and conceptual.

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.

References

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

Anderson, Michael L, Justin Gallagher, and Elizabeth Ramirez Ritchie, "How the Quality of School Lunch Affects Students' Academic Performance," *The Education Digest*, 2018, *83* (6), 61–64.

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.

de Hoyos, Rafael E., "La evaluación que no evalúa," *Nexos*, 2014.

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.

ETS, "ETS Standards for Quality and Fairness," Technical Report, Educational Testing Service. 2014.

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.

Gassman-Pines, Anna and Laura Bellows, "Food instability and academic achievement: a quasi-experiment using SNAP benefit timing," *American Educational Research Journal*, 2018, *55* (5), 897–927.

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

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.

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

LaRoche, Sylvie, Marc Joncas, and Pierre Foy, "Sample Design in TIMSS 2015," *Methods and Procedures in TIMSS 2015*, 2016, pp. 3.1–3.37.

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

Munshi, Kaivan, "Networks in the modern economy: Mexican migrants in the US labor market," *The Quarterly Journal of Economics*, 2003, *118* (2), 549–599.

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.

Shah, Anuj K, 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.

Tomm, Brandon M and Jiaying Zhao, "Scarcity captures attention and induces neglect: Eyetracking and behavioral evidence," in "Proceedings of the 38th Annual Conference of the Cognitive Science Society" 2016, pp. 1199–1204.

Vivanco, E, "Evaluating the effects of school and teacher incentives: Quasi-Experimental Evidence from Mexico," *Manuscrito inédito*, 2013.

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.

Tables

Table 1: Dataset Features

	ASSISTments	TIMSS	ENLACE
Setting	Homework	Exam	Exam
Variation in Financial Salience	Yes	Yes	Yes
SES Indicator	School	Student	School
Panel			Yes
Itemized Question Data	Yes	Yes	
Question Matching	Yes		
Randomized Question Ordering	Yes		

Table 2: 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.

Table 3: Financial Salience and Aggregate Performance in TIMSS

	Proportion Mon Q. in Booklet				Standardized Score			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Below Nat. Median	-0.000141		-0.000234		-0.438***		-0.268***	
	(0.000198)		(0.000240)		(0.00407)		(0.00390)	
Post Secondary		-0.000266		-0.000270		-0.315***		-0.209***
		(0.000262)		(0.000293)		(0.00523)		(0.00489)
Upper Secondary		-0.000315		-0.000367		-0.493***		-0.334***
		(0.000243)		(0.000291)		(0.00493)		(0.00471)
Lower Secondary		0.000185		0.0000258		-0.712***		-0.488***
		(0.000349)		(0.000421)		(0.00715)		(0.00686)
Primary or None		0.000178		0.0000918		-0.779***		-0.523***
		(0.000380)		(0.000504)		(0.00802)		(0.00784)
Below Nat. Median x Prop Mon Q.					-0.260***		-0.262***	
					(0.0471)		(0.0427)	
Post Sec x Prop Mon Q.						-0.0812		-0.0515
						(0.0609)		(0.0558)
Upper Sec x Prop Mon Q.						-0.0865		-0.0844
						(0.0569)		(0.0519)
Lower Sec x Prop Mon Q.						-0.160*		-0.147*
						(0.0825)		(0.0755)
Prim/No x Prop Mon Q.						-0.219**		-0.242***
						(0.0895)		(0.0817)
Constant	0.0617***	0.0618***	0.0618***	0.0618***	0.160***	0.314***	0.100***	0.213***
	(0.000117)	(0.000153)	(0.000128)	(0.000178)	(0.00165)	(0.00215)	(0.00153)	(0.00211)
FE: Year	Yes	Yes	Yes	Yes	.	.	.	.
FE: Booklet x Year	No	No	No	No	Yes	Yes	Yes	Yes
FE: Country	.	.	.	.	Yes	Yes	.	.
FE: Class	No	No	Yes	Yes	No	No	Yes	Yes
N	379468	379468	379160	379160	379468	379468	379160	379160

Note: Standard errors in parentheses. *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. 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, 3, 5 and 7 and university educated parents for columns 2, 4, 6 and 8. The proportion of monetary questions in a booklet is a value from 0 to 1.

Table 4: Financial Salience and Aggregate Performance in ENLACE

	Standardized Score			
	(1)	(2)	(3)	(4)
Below Median x Prop Mon Q.	-0.241***		-0.376***	
	(0.0333)		(0.0464)	
Advantaged x Prop Mon Q.		-0.160***		-0.0903**
		(0.0312)		(0.0408)
Middle x Prop Mon Q.		-0.296***		-0.220***
		(0.0448)		(0.0580)
Disadvantaged x Prop Mon Q.		-0.217***		-0.277***
		(0.0367)		(0.0507)
Very Disadvantaged x Prop Mon Q.		-0.837***		-1.256***
		(0.0823)		(0.118)
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 school level. Observations are weighted by the number of tested students. *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. 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 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.

Table 5: Financial Salience and Rainfall Effects in ENLACE

	Standardized Score	
	(1)	(2)
Below Median x Prop Mon Q.	-0.376***	-0.315***
	(0.0601)	(0.0626)
Below Median x Prop Mon Q. x Drought		-0.453**
		(0.181)
Prop Mon Q. x Drought		0.0881
		(0.0824)
FE: Grade x Year	Yes	Yes
FE: Year x School	Yes	Yes
FE: School x Grade	Yes	Yes
N	1870964	1870964

Note: Standard errors in parentheses are clustered at the municipality level. Observations are weighted by the number of tested students. *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. 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. Drought is an indicator variable set to 1 if rainfall during the prior agricultural season (Jul- Feb) falls in the lowest decile of a locality's rainfall realizations between 1998-2018. The proportion of monetary questions in a booklet is a value from 0 to 1.

Table 6: Financial Salience and Aggregate Performance in ASSISTments

	Mastery		Effort Conditional on Mastered			
	Cond. Logit	Linear	Questions	Time (Sec)	Mean Hints	Mean Attempts
Prop. Free/Red. Lunch x Prop. Mon Q.	-0.607	-0.0318	1.256**	207.9***	0.396**	0.216**
	(0.466)	(0.0683)	(0.543)	(74.21)	(0.166)	(0.107)
Prop. Mon Q.	0.707***	0.0744***	-0.627***	-70.45***	-0.161***	-0.0905***
	(0.163)	(0.0196)	(0.156)	(20.72)	(0.0564)	(0.0254)
FE: Classroom x Assignment	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Mean	0.757	0.787	4.669	329.5	0.386	2.339
Dependent SD	0.429	0.410	2.735	348.0	0.749	0.489
N	19043	22962	12125	12125	12125	10894

Note: Standard errors in parentheses clustered at the school level. *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Observations are at the student by assignment level with a school level SES indicator: the share of students receiving free or reduced lunch. Mastery is a dummy variable set to 1 if a student has mastered the assignment by sequentially answering three questions in a row correctly. Questions is the number of questions a student answered on their assignment. Time is the amount of time in seconds spent on the assignment. Mean hints is the mean number of hints requested by the student on questions in the assignment and mean attempts is the mean number of attempts made by the student on questions in the assignment. Students who exceed more than 8 attempts on any single question are not included in the attempts estimation. Effort measures are estimated on the sub-sample of student assignments where the student masters the assignment and is actively engaged throughout the assignment (defined as the time spent on any question is such that 5 sec $<$ Time $<$ 8.8 min).

Table 7: Matched Monetary Questions in ASSISTments

	Correct	Hints	Attempts	Time (Sec)
	(1)	(2)	(3)	(4)
Prop. Free/Red. Lunch x Mon Q.	-0.0609	0.518**	0.153	23.16*
	(0.0471)	(0.239)	(0.162)	(12.18)
Monetary Question	0.0272	-0.0690	-0.0890	11.42**
	(0.0191)	(0.0983)	(0.0628)	(5.071)
FE: Matched Group by School	Yes	Yes	Yes	Yes
Dependent Mean	0.736	0.986	2.695	90.58
Dependent SD	0.428	2.121	1.434	88.78
N	29277	29277	28207	29277

Note: Standard errors in parentheses clustered at the school level. *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. 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 $<$ 8.8 min). Monetary question is a dummy variable set to 1 if the question features a monetary theme. The omitted category is non-monetary questions.

Table 8: Questions After Matched Monetary Questions in ASSISTments

	Correct	Hints	Attempts	Time (Sec)
	(1)	(2)	(3)	(4)
Prop. Free/Red. Lunch x 4 Post Matched Mon Q.	-0.106	0.791**	0.402	35.14**
	(0.0753)	(0.361)	(0.291)	(15.22)
4 Post Matched Mon Q.	0.0695***	-0.402***	-0.188	-8.223**
	(0.0255)	(0.101)	(0.114)	(3.685)
FE: Leading Matched Q. Type	Yes	Yes	Yes	Yes
FE: Question x School	Yes	Yes	Yes	Yes
Dependent Mean	0.651	1.433	2.891	101.1
Dependent SD	0.466	2.530	1.600	96.40
N	5409	5409	4963	5409

Note: Standard errors in parentheses clustered at the school level. *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. 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 questions positioned between the first and second matched question a student faces and no more than 4 questions following the first matched question. Inactive observations are dropped (5 sec $<$ Time $<$ 8.8 min). 4 Post matched monetary question is a dummy variable set to 1 if a question follows a monetary themed question. The omitted category are questions that follow matched non-monetary themed questions.

Table 9: Questions After Matched Monetary Questions in ASSISTments with Controls for Preceding Differential Difficulty

	Correct	Hints	Attempts	Time (Sec)
	(1)	(2)	(3)	(4)
Prop. Free/Red. Lunch x 4 Post Matched Mon Q.	-0.110	0.817**	0.413	33.11**
	(0.0735)	(0.351)	(0.284)	(16.02)
4 Post Matched Mon Q.	0.0632**	-0.361***	-0.173	-11.34***
	(0.0269)	(0.0938)	(0.123)	(4.244)
Quartile Mean Time on Leading Matched Q.	0.000225	-0.00144	-0.000549	0.111***
	(0.000204)	(0.00165)	(0.000944)	(0.0396)
FE: Leading Matched Q. Group	Yes	Yes	Yes	Yes
FE: Question x School	Yes	Yes	Yes	Yes
Dependent Mean	0.651	1.433	2.891	101.1
Dependent SD	0.466	2.530	1.600	96.40
N	5409	5409	4963	5409

Note: Standard errors in parentheses clustered at the school level. *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. 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 questions positioned between the first and second matched question a student faces and no more than 4 questions following the first matched question. Inactive observations are dropped (5 sec $<$ Time $<$ 8.8 min). 4 Post matched monetary question is a dummy variable set to 1 if a question follows a monetary themed question. The omitted category are questions that follow matched non-monetary themed questions.

Table 10: Questions After Matched Monetary Questions in ASSISTments with Sequence Controls

	Correct	Hints	Attempts	Time (Sec)
	(1)	(2)	(3)	(4)
Prop. Free/Red. Lunch x 4 Post Matched Mon Q.	-0.107	0.792**	0.406	35.27**
	(0.0753)	(0.359)	(0.290)	(16.73)
4 Post Matched Mon Q.	0.0692***	-0.401***	-0.187	-8.092**
	(0.0251)	(0.101)	(0.113)	(3.713)
Sequence Positon	0.0152**	-0.0471*	-0.0427**	-5.432***
	(0.00589)	(0.0242)	(0.0192)	(0.625)
FE: Leading Matched Q. Type	Yes	Yes	Yes	Yes
FE: Question x School	Yes	Yes	Yes	Yes
Dependent Mean	0.651	1.433	2.891	101.1
Dependent SD	0.466	2.530	1.600	96.40
N	5409	5409	4963	5409

Note: Standard errors in parentheses clustered at the school level. *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. 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 questions positioned between the first and second matched question a student faces and no more than 4 questions following the first matched question. Inactive observations are dropped (5 sec $<$ Time $<$ 8.8 min). 4 Post matched monetary question is a dummy variable set to 1 if a question follows a monetary themed question. The omitted category are questions that follow matched non-monetary themed questions.

Table 11: Regressions on Question Means in TIMSS

	Mean Performance (Correct=100) by Students with Parental Education Level:				
	Below Nat. Median	Primary/No	Lower Sec	Upper Sec	Post Sec
Monetary Question	-1.529*	-3.042*	-2.265*	-1.057	-0.511
	(0.908)	(1.749)	(1.293)	(0.717)	(0.429)
4 Post Question	-1.001*	-1.594	-1.045	-0.844	-0.425
	(0.576)	(1.215)	(0.903)	(0.666)	(0.376)
Q. Mean for Par. Edu. Above Nat. Median	0.939***				
	(0.0118)				
Question Mean for Univ. Parental Edu.		0.784***	0.942***	1.026***	1.036***
		(0.0227)	(0.0166)	(0.0112)	(0.00684)
N	706	706	706	706	706

Note: Standard errors in parentheses clustered at the prompt level. *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Observations are at the question level with a student level SES indicator: parental education. Monetary Question is dummy variable set to 1 if a question is monetary themed. 4 Post is a dummy variable set to 1 if a question is non-monetary themed and positioned within 4 questions following a monetary question. Question means by parental education level are the mean of correct responses for each group where a correct response equals 100 and an incorrect response equals 0.

Table 12: 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)
Primairy/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)
Primairy/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. *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. 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 8.

Table 13: 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)
Primairy/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)
Primairy/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. *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. 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 8. Sample mechanically does not include the first for questions on an exam for which the differential difficulty controls are undefined.

Table 14: Regressions on Placebo Coefficients

	Placebo Post Estimates ($\hat{\Lambda}_3^{placebo}$)
Placebo Mon. Estimates ($\hat{\Lambda}_2^{placebo}$)	0.142***
	(0.0202)
N	1000

Note: *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. $\hat{\Lambda}_2^{placebo}$ are the estimates for Λ_2 from equation 11 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.

Table 15: COMIPEMS Simulation Summary Statistics

SES Indicator	Group	Observations	Math Mean		Mean Change
			Actual	Simulated	in Rank
School Indicator	Missing	16,437	13.10	13.10	−39
	Very Advantaged	391,249	14.15	14.15	−38
	Advantaged	90,112	13.56	13.61	144
	Middle	3,445	13.51	13.63	424
	Disdvantaged	1,666	12.61	12.75	537
	Very Disadvantaged	34	10.76	11.44	2,291
Parental Education	Missing	81,164	13.20	13.20	−197
	University	62,592	16.19	16.19	−167
	Upper Secondary	130,618	14.73	14.78	−13
	Lower Secondary	137,216	13.51	13.58	88
	Primary or Less	91,353	12.89	12.99	175

Table 16: COMIPEMS Simulation using Parental Education

		Ineligible		Eligible			
		Score under 31 points		Not Assigned		Assigned	
Total	Actual	8,373	1.67%	84,513	16.80%	410,057	81.53%
	Simulated	8,245	1.64%	84,584	16.82%	410,114	81.54%
Missing	Actual	1,756	2.16%	14,878	18.33%	64,530	79.51%
	Simulated	1,756	2.16%	14,937	18.40%	64,471	79.43%
University	Actual	349	0.56%	10,565	16.88%	51,678	82.56%
	Simulated	349	0.56%	10,592	16.92%	51,651	82.52%
Upper Secondary	Actual	1,334	1.02%	22,590	17.30%	106,694	81.68%
	Simulated	1,315	1.01%	22,594	17.30%	106,709	81.70%
Lower Secondary	Actual	2,637	1.92%	22,717	16.56%	111,862	81.52%
	Simulated	2,594	1.89%	22,727	16.56%	111,895	81.55%
Primary or Less	Actual	2,297	2.51%	13,763	15.07%	75,293	82.42%
	Simulated	2,231	2.44%	13,734	15.03%	75,388	82.52%

Table 17: COMIPEMS Simulation using Parental Education: Movement Detail

	Remain Ineligible or Unassigned	Become Assigned	More Prefered Assignment	Unchanged Assignment	Less Prefered Assignment	Become Unassigned	Change in Mean Preference Rank*
Total	92,510	376	987	407,732	1,019	319	-.00028
	18.39%	0.08%	0.20%	81.07%	0.20%	0.06%	
Missing	16,625	9	15	64,282	165	68	-.00453
	20.48%	0.01%	0.02%	79.20%	0.20%	0.08%	
University	10,911	3	3	51,534	111	30	-.00445
	17.43%	0.00%	0.00%	82.33%	0.18%	0.05%	
Upper Secondary	23,834	90	242	106,102	275	75	-.00082
	18.25%	0.07%	0.19%	81.23%	0.21%	0.06%	
Lower Secondary	25,223	131	419	111,057	288	98	.00234
	18.38%	0.10%	0.31%	80.94%	0.21%	0.07%	
Primary or Less	15,917	143	308	74,757	180	48	.00308
	17.42%	0.16%	0.34%	81.83%	0.20%	0.05%	

Note: For students who are assigned in both the actual and simulated data.

Figures

Figure 1: ASSISTments Variation

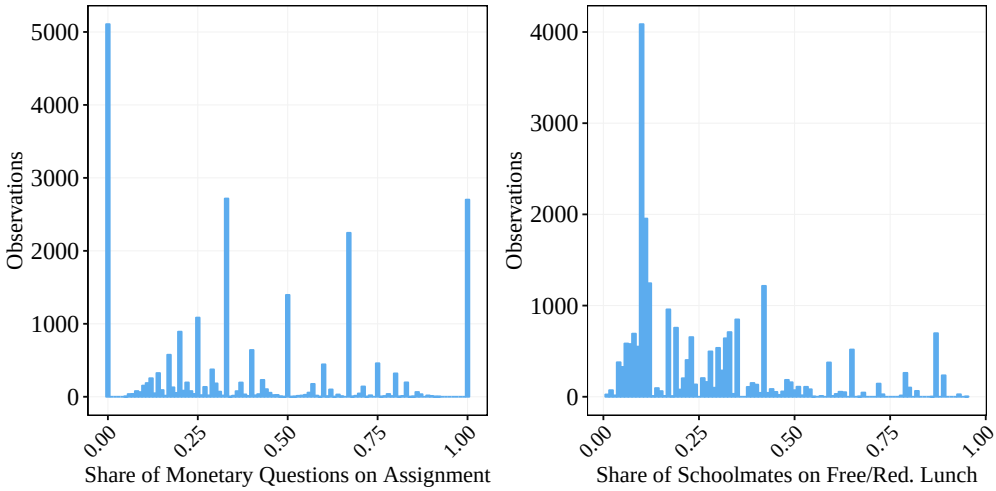


Figure 2: TIMSS Variation

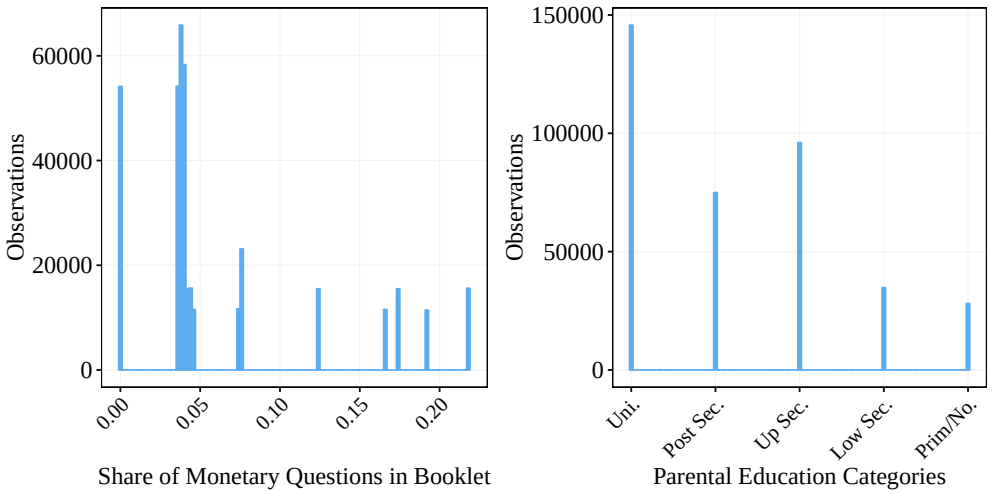


Figure 3: ENLACE Variation

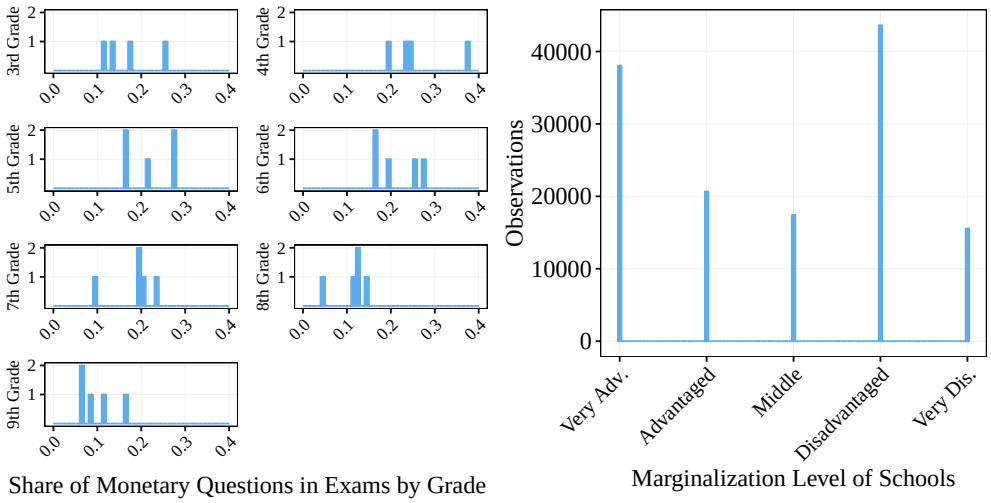
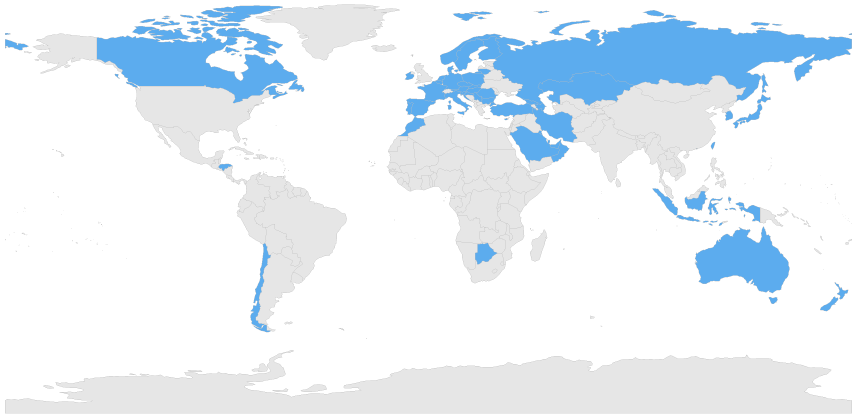


Figure 4: Countries participating in 4th grade TIMSS



Note: Mapped countries only show countries participating in the 4th grade TIMSS in 2011 and 2015 in which parental questionnaires were administerd.

Figure 5: ASSISTments Question Statistics by Question Type

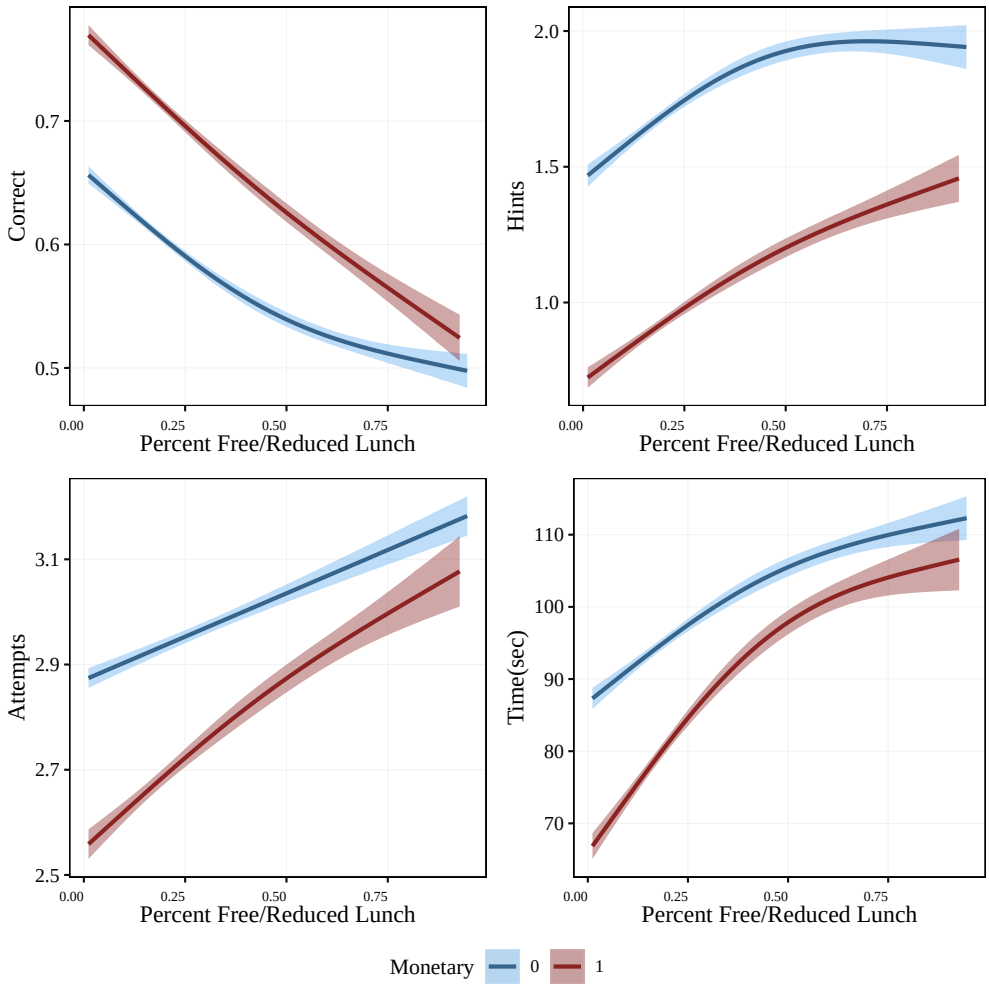


Figure 6: Examples of Matched ASSISTments Questions

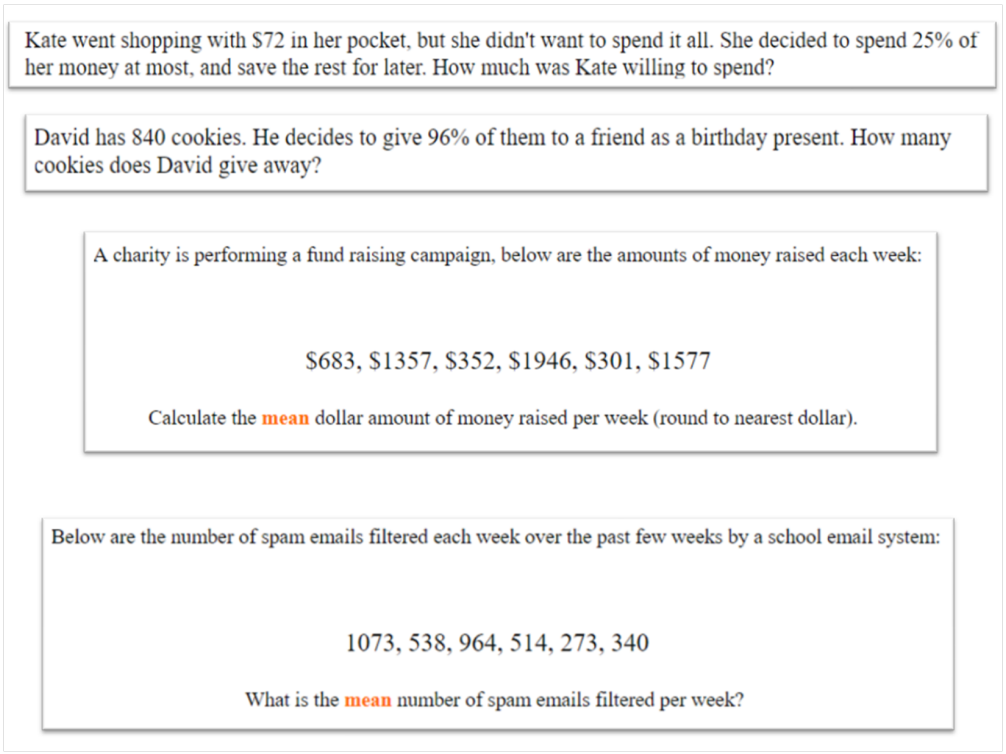
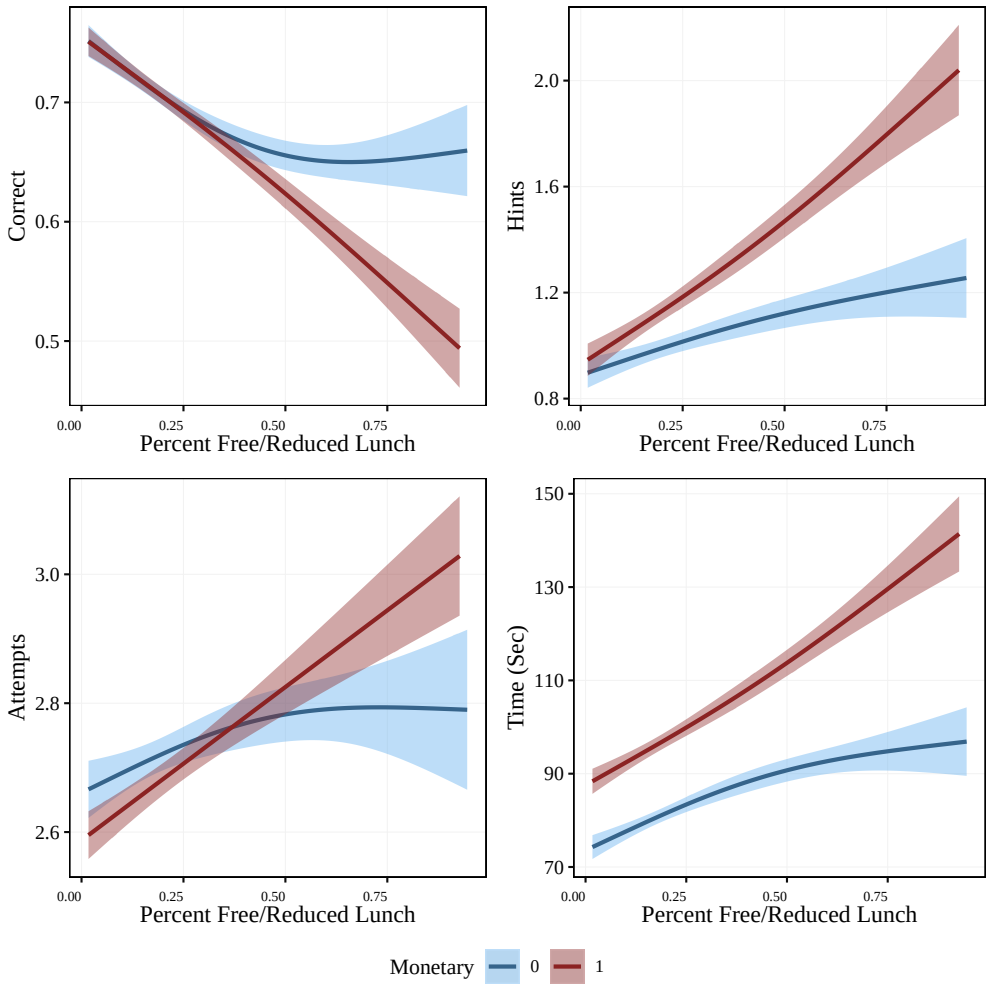


Figure 7: ASSISTments Matched Question Statistics by Question Type



Note: Data is limited to questions that are matched to a monetary themed question.

Figure 8: Question Characteristics by Category

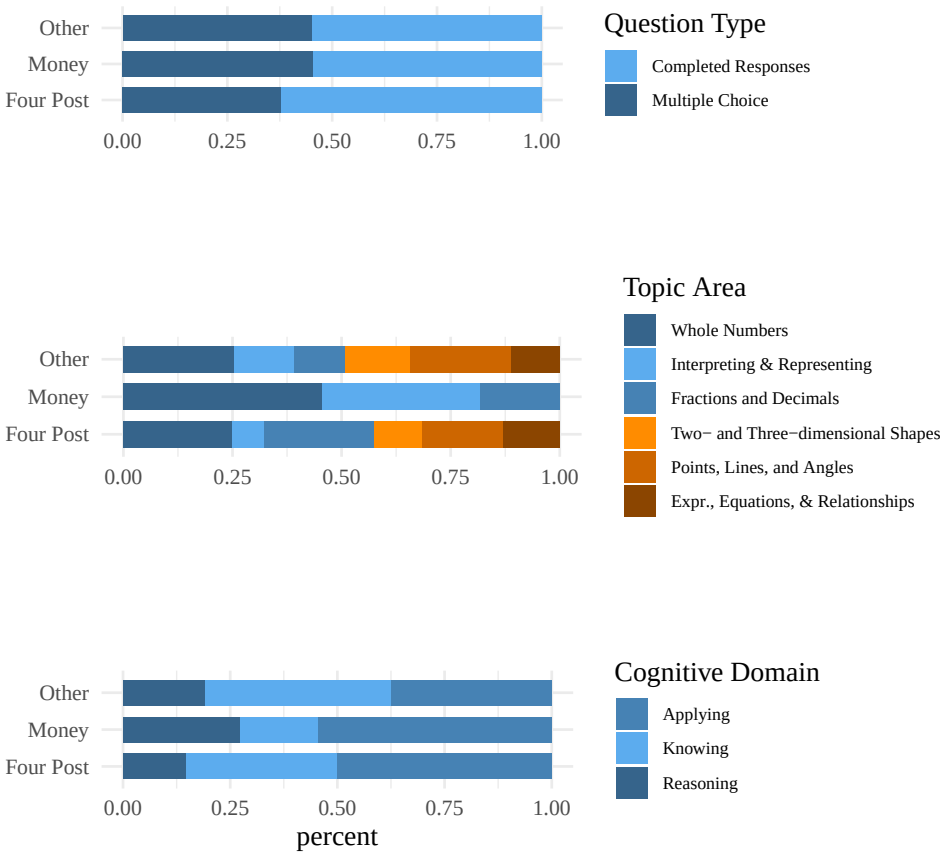
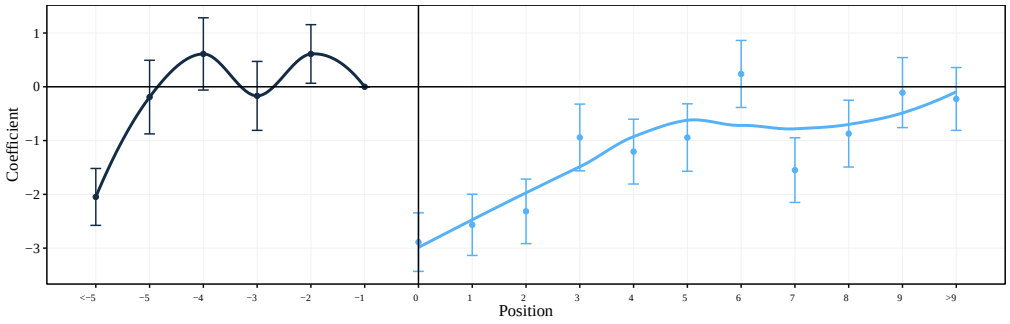
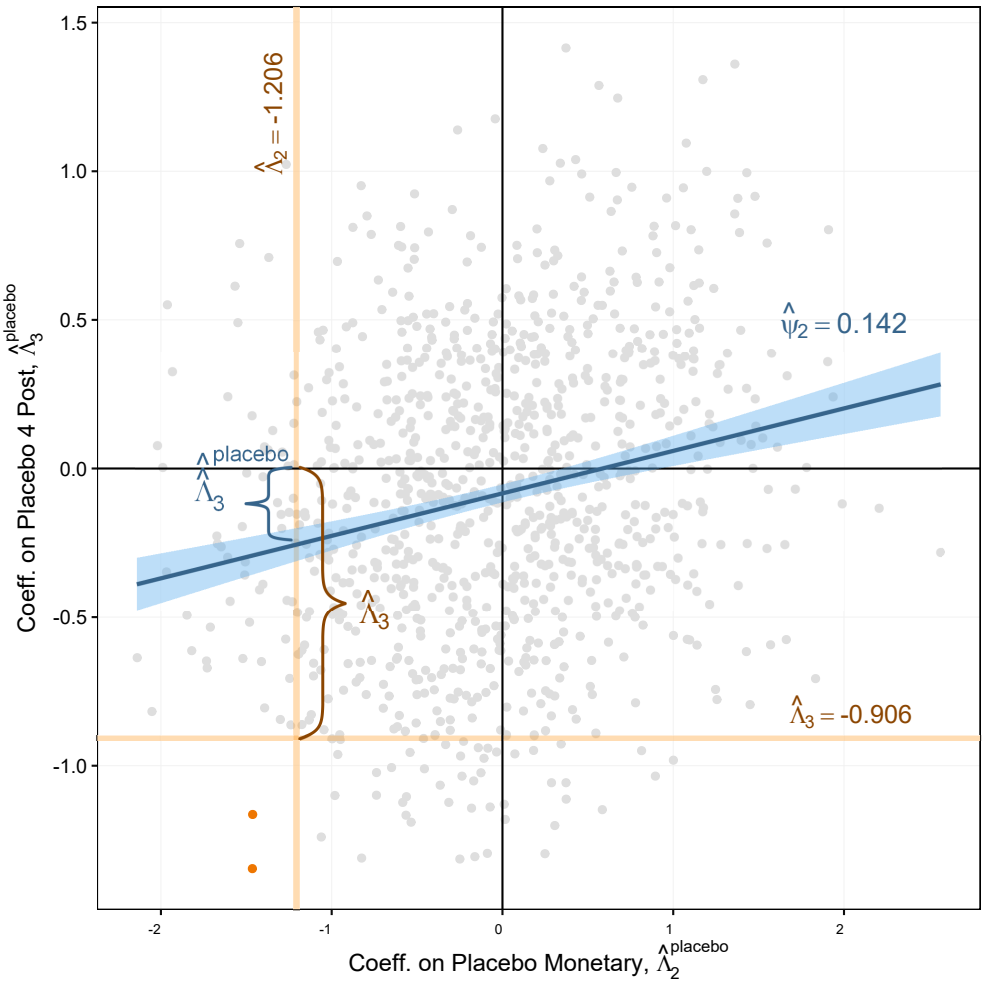


Figure 9: Differential Performance by Position Relative to Monetary Event for Below National Median Students in TIMSS



Note: Data is limited to booklets that feature a single monetary question or booklets that feature only two sequential monetary questions. Estimating equation includes Student, Question, Below Med. x Diff., Below Med. x Seq., Below Med. x QType x Country, Below Med. x QTopic x Country fixed effects.

Figure 10: Estimates from 1000 Placebo Estimations



Note:

Appendix

Table 18: Missing Parental Education

	Prop Mon Q. in Booklet	Standardized Score			
	(1)	(2)	(3)	(4)	(5)
Missing Parental Education	-0.000239	-0.297***	-0.232***	-0.283***	-0.219***
	(0.000292)	(0.00367)	(0.00354)	(0.00491)	(0.00457)
Missing Par Edu. x Prop Mon Q.				-0.233***	-0.211***
				(0.0530)	(0.0468)
FE: Year	Yes	.	.	.	.
FE: Country	.	Yes	.	Yes	.
FE: Class	Yes	No	Yes	No	Yes
FE: Booklet x Year	No	Yes	Yes	Yes	Yes
N	469697	469849	469697	469849	469697

Note: Standard errors in parentheses. *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Observations are at the student by examination level with a student level SES indicator: parental education. 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 19: TIMSS Main Results by Occupation

	Standardized Score	
	(1)	(2)
Small Business	-0.163***	-0.114***
	(0.00658)	(0.00597)
Clerical	-0.249***	-0.131***
	(0.00506)	(0.00465)
Skilled Labor	-0.373***	-0.220***
	(0.00672)	(0.00622)
General Labor	-0.518***	-0.292***
	(0.0100)	(0.00932)
Never Wk. for Pay	-0.438***	-0.217***
	(0.0106)	(0.00979)
Small Business x Prop Mon Q.	0.0142	-0.0646
	(0.0777)	(0.0700)
Clerical x Prop Mon Q.	-0.0179	-0.0374
	(0.0591)	(0.0532)
Skilled Labor x Prop Mon Q.	-0.0968	-0.0815
	(0.0784)	(0.0707)
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*
	(0.123)	(0.111)
Constant	0.148***	0.0857***
	(0.00192)	(0.00176)
FE: Booklet x Year	Yes	Yes
FE: Country	Yes	.
FE: Class	No	Yes
N	379468	379160

Note: Standard errors in parentheses. *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Observations are at the student by examination level with a student level SES indicator: parental occupation. Omitted categories are students with professional parental occupations. The proportion of monetary questions in a booklet is a value from 0 to 1.

Table 20: 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

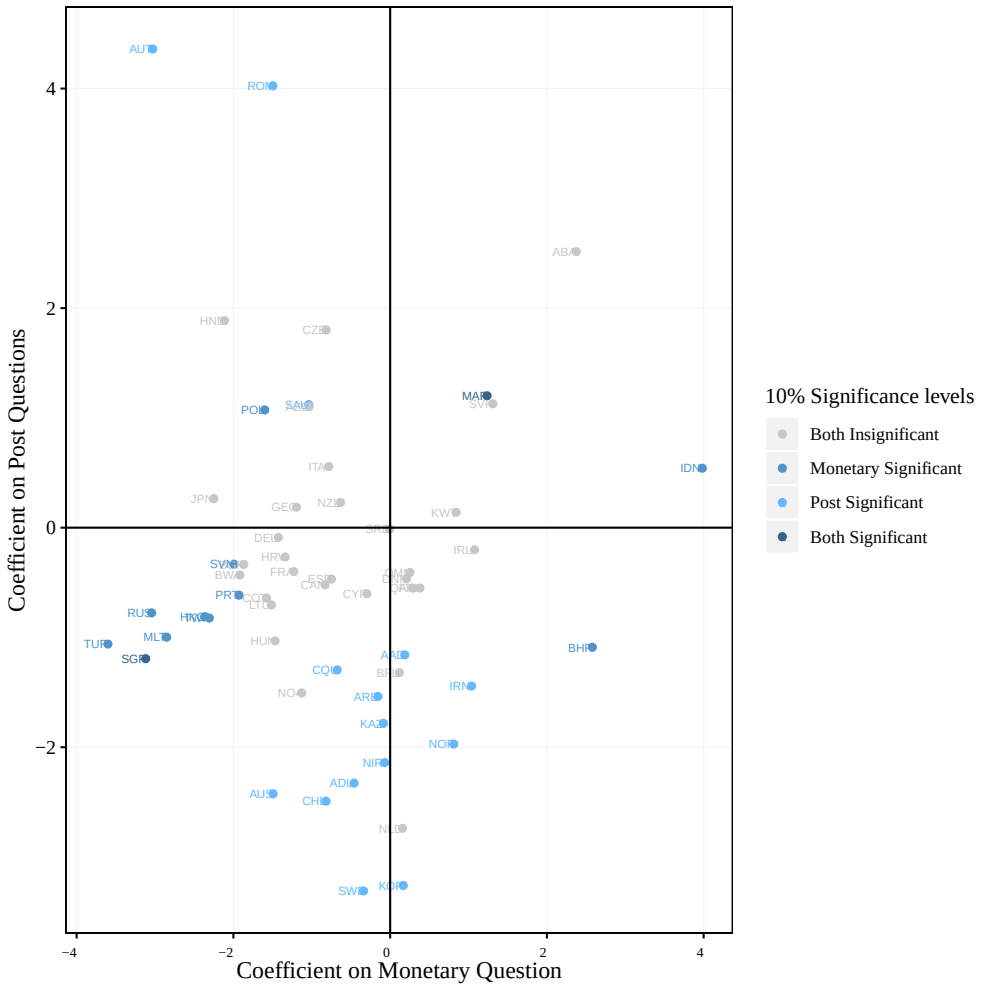
Note: Standard errors in parentheses clustered at the student level. *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. Observations are at the question by student level with a student level SES indicator: parental education relative to the national median. When a question is answered correctly the indicator 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 8.

Table 21: 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)
Primairy/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)
Primairy/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. *$p < 0.10$,** $p < 0.05$, *** $p < 0.01$. 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 8.

Figure 11: Country Estimates for Below National Median Students



Note: Estimating equation includes Student, Question, Below Med. x Diff., Below Med. x Seq., Below Med. x QType x Country, Below Med. x QTopic x Country fixed effects.

Table 22: COMIPEMS Simulation using School Indicators

		Ineligible		Eligible			
		Score under 31 points		Not Assigned		Assigned	
Total	Actual	8,373	1.67%	84,513	16.80%	410,057	81.53%
	Simulated	8,353	1.66%	84,518	16.81%	410,072	81.53%
Missing	Actual	392	2.39%	3,312	20.15%	12,733	77.47%
	Simulated	392	2.39%	3,314	20.16%	12,731	77.45%
Very Advantaged	Actual	6,245	1.60%	69,815	17.84%	315,189	80.56%
	Simulated	6,245	1.60%	69,844	17.85%	315,160	80.55%
Advantaged	Actual	1,615	1.79%	10,717	11.89%	77,780	86.32%
	Simulated	1,600	1.78%	10,695	11.87%	77,817	86.36%
Middle	Actual	74	2.15%	389	11.29%	2,982	86.56%
	Simulated	71	2.06%	386	11.21%	2,988	86.73%
Disadvantaged	Actual	45	2.70%	274	16.45%	1,347	80.85%
	Simulated	43	2.58%	273	16.39%	1,350	81.03%
Very Disadvantaged	Actual	2	5.88%	6	17.65%	26	76.47%
	Simulated	2	5.88%	6	17.65%	26	76.47%

Table 23: COMIPEMS Simulation using School Indicators: Movement Detail

	Remain Ineligible or Unassigned	Become Assigned	More Prefered Assignment	Unchanged Assignment	Less Prefered Assignment	Become Unassigned	Change in Mean Preference Rank*
Total	92,807	79	216	409,548	229	64	-.00007
	18.45%	0.02%	0.04%	81.43%	0.05%	0.01%	
Missing	3,703	1	1	12,725	4	3	-.00047
	22.53%	0.01%	0.01%	77.42%	0.02%	0.02%	
Very Advantaged	76,041	19	31	314,945	165	48	-.00082
	19.44%	0.00%	0.01%	80.50%	0.04%	0.01%	
Advantaged	12,282	50	157	77,550	60	13	.00256
	13.63%	0.06%	0.17%	86.06%	0.07%	0.01%	
Middle	457	6	21	2,961	0	0	.00905
	13.27%	0.17%	0.61%	85.95%	0%	0%	
Disdvantaged	316	3	6	1,341	0	0	.00594
	18.97%	0.18%	0.36%	80.49%	0%	0%	
Very Disadvantaged	8	0	0	26	0	0	0
	23.53%	0%	0%	76.47%	0%	0%	

Note: For students who are assigned in both the actual and simulated data.

Figure 12: 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) 

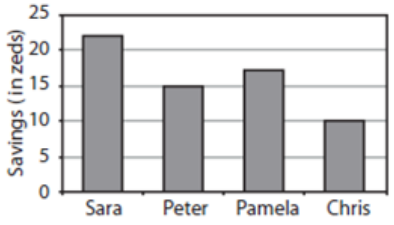
D) 

Figure 13: Example Monetary Questions from the 2011 TIMSS

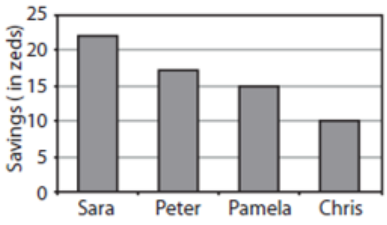
John was given the following table by his teacher and was asked to identify the graph that correctly displays the data. Which graph below should he choose?

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

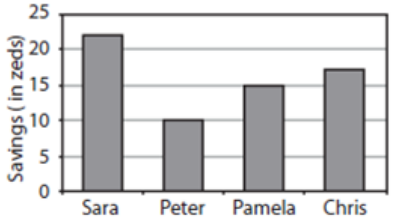
A.



B.



C.



D.