Remedying Education with Personalized Learning: Evidence from Randomized Field Experiment in India

Anuj Kumar
Assistant Professor
Warrington College of Business Administration
University of Florida
akumar1@ufl.edu

Amit Mehra
Associate Professor
Naveen Jindal School of Management
University of Texas Dallas
Amit.Mehra@utdallas.edu

This version

(April 2016)

Remedying Education with Personalized Learning: Evidence from Randomized Field Experiment in India

Abstract

Can Information and Communication Technology (ICT) enabled personalization remedy the educational production in resource-strapped schooling systems? We conduct a randomized field experiment on a group of residential schools in Hyderabad India to examine this question. In a school setting, students first learn concepts through class room instructions and then reinforce their learning by doing homework. In our experiment, students were first taught different topics in mathematics through classroom instructions, and then a randomly selected one half of them were assigned computer-generated adaptive homework (CGAHW) and the other half were offered paper-based traditional homework (PBTHW). In a PBTHW, a pre-decided fixed number of easy and hard questions were offered from different topics. In a CGAHW, first half of the total questions were offered in the easy category, and based on a student's performance on these questions, later questions were adaptively generated such that: (1) more questions were offered on the topics in which student incorrectly answered questions and (2) hard questions on a topic were offered when the student correctly answered easy questions on that topic. Thus, while all PBTHW students received the same number of easy and hard questions on different topics, CGAHW students received different numbers and difficulty levels of questions on different topics based on their individual learning needs. A total of 50 homework in each category were offered to students between October 2014 and April 2015, and their learning was assessed in two standardized exams offered in this period.

We found that CGAHW students on average obtained lower homework scores than PBTHW students, but they obtained 4.28 percent higher scores in exams than PBTHW students. Lower homework scores could be attributed to students receiving more questions in their weak areas in CGAHW. However, by doing more questions in their weak areas and less in their strong areas, students achieved personalized learning in CGAHW, and hence obtained higher exam scores. To provide evidence that personalized learning in CGAHW resulted in improvement in their exam scores, we show that students that were offered higher levels of personalization in CGAHW, obtained higher exam scores.

To further understand the differential effect of CGAHW on students of different abilities, we categorized students in low, medium, and high categories of ability based on their mathematics scores in standardized exams at the beginning of experiment. We found that personalized learning through CGAHW helped the students in low and medium ability categories but not in high ability category. Overall, we developed and deployed an adaptive homework generation application in a field set up to show how ICT-enabled personalized learning could improve educational production with existing school resources.

Key Words *– Economics of education; ICT and education; personalized learning; adaptive learning; randomized field study.*

Introduction

Recent World Bank reports on education outlines the poor quality of education in developing countries (World Bank Report 2005). In particular, the quality of secondary education has significantly deteriorated in developing countries in the early twentieth century. The average cognitive performance of secondary-level students (measured by the standardized Program on International Student Assessment (PISA) scores) in developing countries was found to be significantly lower than that of their counterparts in developed countries. Furthermore, PISA scores of a large proportion of the student population in many developing countries was found below the average scores in developed countries. For instance, scores of students at 95th percentile in Peru and Indonesia were below the average scores for OECD countries. Another study places the average standardized mathematics scores (Trends in International Mathematics and Science Study (TIMSS) scores) obtained by students from two large states (Rajasthan and Orissa) in India at 44th place out of 51 countries tested (World Bank Report 2009). This poor quality of education in developing countries is attributed to the inability of their existing secondary school infrastructures to meet the increased demand for secondary education, which is caused by a large number of students graduating from elementary schools due to the adoption of universal and inclusive elementary school schemes by these countries in the early twentieth century (World Bank Report 2009, Verspoor 2008).[1]

Since secondary education is a bridge between elementary and higher education, the poor quality of secondary education could impede human capital growth and hence lower the productivity and economic growth in these countries. While we know that improvement in the education infrastructure is urgently required in developing countries, it is not clear how to achieve it in a quick and cost effective way. The World Bank studies call for comprehensive systemic reform of the education system in these countries, but such reforms may require substantial investments and may take a long time to implement. Recent advances in information and communication technologies (ICT), however, offer a hope of supplementing the existing educational infrastructure in developing countries in a shorter time span. For this reason, the school systems all over the world have introduced computer use, the Internet, and a variety of educational applications of ICT in the last decade. While introducing ICT in schools involves substantial investment, the available evidence of benefits of ICT on student learning, in academic and practitioners studies, are mixed and do not provide a definitive answer (OECD 2015, Belo et al. 2014).

[1] Many developing countries in Asia and Africa have implemented the universal elementary educations schemes in the early twentieth century. For example, the Government of India adopted a universal elementary education scheme called "Sarva Shiksha Abhyihan". As a result, the number of students completing elementary education in India was expected to grow at five percent per year between 2001 and 2014, which translates into a projected increase in demand for secondary education from 40 million in 2007/08 to 57 million in 2017/18, a whopping 43 % increase (World Bank report 2009).

Therefore, carefully designed micro studies that rigorously measure the effect of individual functionality of ICT on student learning are needed to clearly understand how ICT could affect educational production. To this end, the present study examines how ICT-based personalization in homework could affect student learning and measure its effect rigorously with a randomized field experiment.

ICT-based personalization technologies have been widely used in ecommerce to learn individual customer needs based on their past browsing/purchase behavior and accordingly offer them personalized product recommendations to maximize sales. These recommendation systems have been successfully used in a variety of products and services, such as, movies (Netflix.com), music (Pandora.com and Spotify.com), miscellaneous products (Amazon.com), home services (Angie's List, HomeAdvisors.com), and dating services (match.com, eHarmony). In a similar vein, personalization technologies are developed to personalize the educational content and interactions with children as per their learning needs to maximize learning. An emerging example of such personalization technology is adaptive learning. As per the paper commissioned by the Bill and Melinda Gates foundation and authored by Education Growth Advisors

"*adaptive learning is a sophisticated data driven, and in some cases, nonlinear approach to instructions and remediation, adjusting to a learner's interaction and demonstrated performance level, and subsequently anticipating what type of content and resources learners need at a specific point in time to make progress*" (NMC Horizon Report: 2015 Higher Education Edition).

The basic idea of adaptive learning is to personalize instructions, learning material, and homework in accordance with individual student needs and thus enhance the effectiveness of learning resources. Although a variety of adaptive learning solutions are available in the market that are costly to deploy and maintain, there is very little rigorous empirical evidence of their efficacy. In the present research, we developed an adaptive software to generate personalized mathematics homework and deployed it in a group of secondary schools in India to empirically measure its effectiveness in improving students' cognitive achievements.

In a school setting, students initially learn concepts through class room instructions and then reinforce their learning by solving homework problems on those concepts. This is referred to as learning by doing. In developing countries, teachers with inadequate training and resources teach big class sizes of students.[2] These teachers have little motivation to adapt their class room instructions as per the needs of individual students and, thus, their instructions may result in heterogeneous learning among students of

[2] The average class size in secondary schools in developing countries is more than double the allowable class sizes in developed countries (Verspoor 2008)

different abilities. If one-size-fits-all homework is offered to student populations with different learning needs, the learning by doing through homework could be severely limited. However, if ICT-based application can identify individual learning needs of students and accordingly offer questions in homework, the learning by doing from homework could be significantly improved. For example, if more questions are generated on the topics in which students incorrectly answered earlier questions and fewer questions on topics in which they correctly answered earlier questions. Then, while students learn by solving more questions on topics that they did not understand well in class room instructions, they do not unnecessarily solve questions on the topics that they understood well.

To test this intuition, we conducted a randomized field experiment on 7th and 8th grade students in a group of residential schools in Hyderabad, a large metropolitan area, in India from October 2014 to April 2015. In this experiment, we randomly assigned half of the students to computer-generate adaptive homework (CGAHW) in mathematics, and the remaining half to the paper-based traditional homework (PBTHW). Questions in both categories of homework were drawn from the same question bank and homework in both categories on same topics contained the same total number of questions. The topic-wise breakup of easy and hard questions in a PBTHW were fixed and hence all students receiving a PBTHW got the same number of easy and hard questions on different topics. In a CGAHW, half of the total questions in the easy category (called base questions) were first offered on all topics included in the homework. The topic-wise breakup of base questions in a CGAHW was kept the same as in the corresponding PBTHW. Thereafter, a greater number of easy questions were generated on the topics in which the student incorrectly answered base questions, and hard questions on a topic were generated only when the student correctly answered easy questions on that topic. As a result, the topic-wise number of easy and hard questions for students receiving CGAHW varied based on their performance. During the experiment period, teachers first offered class room instructions to students on different topics in mathematics and then assigned homework on those topics. Overall, 50 homework covering different topics in 7th and 8th grade mathematics syllabus were offered to the students. The students were assessed for their learning in two paper-based exams, a half-yearly exam in December 2014 and a yearly exam in April 2015.

The schools in the experiment were required by the state law to admit students by random lottery from the pool of applicants belonging to economically poor households. Thus, our sample of students, even though from a few schools, was representative of children from economically poor households in a large metropolitan area in India. The students in our experimental setup stayed in the school dormitories and were allocated a fixed amount of time for different activities in a day, such as meals, class room instructions, self-study, recreation, and sleeping. As a result, both categories of students in our experiment had similar class room instructions, learning resources, available time for study, and peer groups. Moreover, the confounding effect of family environment on student learning was absent. The only difference in available learning resources between the two categories of students was the nature of their homework. Moreover, the two categories of homework were similar in all respects except that CGAHW offered personalized questions as per the student's needs but PBTHW offered fixed questions. Therefore, the difference in exam scores for the two categories of students in our experiment measured the causal effect of offering personalized questions to students on their cognitive achievements.

We found that while CGAHW students obtained lower homework scores than PBTHW students, they obtained 4.28 percent higher scores in the exams than PBTHW students. Lower CGAHW scores could be due to students receiving more questions in their weak areas in such homework. By solving more questions in their weak areas and less in their strong areas, CGAHW students achieved higher learning by doing through their homework, and hence obtained higher exam scores. The 4.28 percent increase in exam scores for CGAHW students translates into an eight percent increase over the average score of 51.8 percent for PBTHW students. To provide empirical evidence that higher personalization in homework resulted in higher learning by doing, we show that students who were offered higher personalization in homework obtained higher exam scores. We further examined the effect of adaptive homework on cognitive achievements of students of different baseline mathematical abilities (as measured by their scores in mathematics exams in the previous year). We found that adaptive homework benefited students in the bottom and middle terciles but not in the top tercile of ability. Thus, offering personalized homework could reduce the achievement gap between high and low ability students.

Overall, we developed a simple and cost effective adaptive homework generation software and deployed it in a school settings with limited resources, and show that offering personalized homework significantly improves students' cognitive achievements. Our results have implications for remedying education production in situations of limited educational resources and are, therefore, particularly relevant for remedying the educational production in developing countries.

We organize the remainder of this paper as follows. In §2, we provide a review of relevant papers in this domain. We introduce our field study setting in §3, describe our data in §4, and explain our econometric specifications and results in §5. Finally, in §6, we conclude and outline future research possibilities and limitations.

2.0 Past Literature

The role of schooling in human capital development has been of great interest to policy makers. Policy makers in the US have instituted the “Equality of Educational Opportunity study” (known as the Coleman Report) that examined the relationship between school inputs and student scholastic achievements (Coleman 1966). Economists formalized this input-output relationship of Coleman Report into formal “educational production functions (EPF)”, which states that educational inputs, such as school resources, teacher quality, family effects, and peer effects result in educational production that is measured by student cognitive (scholastic) achievements. Numerous studies have shown that quality and quantity of schooling inputs significantly affect student cognitive achievement and thus affect the productivity, national growth rates (Hanushek and Kimko 2000), and labor market outcomes (Mincer 1970, Psacharopoulous and Patrinos 2004). Similarly, other studies found that labor market returns were significantly affected by individual differences in cognitive achievements (Mulligan 1999, Murane et al. 2000, and Lazear 2003).

A large number of studies have been conducted to understand the relationship between common inputs, such as school organization (class size, school facilities, administrative expenditure), teacher background (educational level, experience, sex, race), peer inputs (aggregate of student socio-demographic characteristics and achievement) and family background (parental education, income, family size) on students educational outcomes (see Hanushek 1979, 2003, 2007 for survey of these studies). While some studies show a positive relationship between the school inputs and student cognitive achievements (Krueger 1999, Angrist and Lavy 1999), others show negative or no correlations between the two (Hanushek 2003, Rivkin et al. 2005). The contradiction in the results of prior studies has been attributed to imprecise input/output measures and faulty econometric design (Hanushek 2007, Todd and Wolpin 2003). Most prior studies examined the relationship between students’ cognitive achievements and contemporaneous school inputs, but the students’ cognitive achievements at a time are determined by their genetic endowment of mental capacity and the history of their received inputs, such as family environment, peer effects, and school resources. Todd and Wolpin (2003) proposed a value-added form of EPF in a panel-data setting to account for the unobserved history of inputs. Accordingly, we use the value-added form of EPF specification in the present study to measure the effect of personalized homework (a school input) on students’ cognitive achievements.

ICT has emerged as an important school input to promote student learning. Sosin et al. 2005 compare the efficacy and efficiency of the various types of available ICT on the performance of students in economics classes across the US universities. The visualization tools and multimedia in ICT can help better illustrate the educational content and create better student engagement with the content. Moreover, the interactive features of ICT can help students interactively engage with the learning content as per their needs. While ICT has the potential to improve student learning, unfortunately the empirical evidence available from studies in developed countries is not encouraging (Leuven et al. 2004, Krueger and Rouse 2004, Machin et al. 2006, OECD 2015). From the developing countries perspective, Banerjee et al. (2007) show evidence of improvement in students' cognitive achievements with computer-assisted learning in elementary schools in India. While most of these studies look at the effect of ICT-mediated contents on students' cognitive achievements, we examine the effect of ICT-enabled adaptive homework on students' cognitive achievements.

Both, economists and psychologists, agree that learning is the product of experience, i.e., learning takes place through the attempt to solve a problem and therefore takes place during activity. Psychologists referred to it as "Learning by Doing". For example, Anzai and Simon (1979), proposed a theory of the processes that enable a student to learn while engaged in solving a problem. In a similar vein, economists found that economic production increases with time, with the same capital, labor, and technology. They attribute this increase in production to technology knowledge or learning to use the technology more efficiently (Arrow 1962). Applying this in the context of educational production, students should learn more by solving more homework questions, which in turn, should result in their higher cognitive (scholastic) achievements. Below we illustrate how learning by doing can be maximized with computer-generated adaptive homework.

In a school setting, students are first exposed to concepts through class room instructions. Class room instructions, tailored for an average student, could lead to heterogeneous learning across students of different abilities. A well-trained and motivated teacher, improvises and personalizes class room instructions to cater to the needs of different students and thus, as far as possible, imparts uniform learning across students of different abilities. The basic purpose of homework is to reinforce initial learning through class room instructions with learning by doing, i.e., allow the students to learn concepts by doing homework questions. While solving homework questions, students can refer to their books and class notes to clarify their doubts on the concepts required to solve questions and thus learn them. Moreover, students learn concepts incrementally by first solving easy questions involving a few simple steps, followed by solving hard questions involving multiple steps. For this reason, a well-designed homework first offers easy questions on a concept followed by hard questions.

One-size-fits-all homework could be highly inefficient in delivering "Learning by Doing" in cases when significant heterogeneity exists across students in their levels of understanding of different concepts covered in the homework. We illustrate this with a simple example. Consider a homework that offers 10 questions from two topics, T1 and T2. A traditional homework would equally distribute 10 questions between two topics, and offer three easy and two hard questions from each topic. Now consider two students, S1 and S2, who acquired different levels of understanding on these topics through the class room instructions. While, S1 understands T1 very well but T2 not so well, S2 understands T2 very well but T1 not so well. In a traditional homework, S1 and S2, respectively, learn topics T2 and T1 by solving three easy and two hard questions on each of these topics. An adaptive homework first offers two easy questions on each topic. As S1 understands T1 well, she is likely to correctly answer both easy questions on T1 and incorrectly answer both easy questions on T2. In the next round, S1 will be offered one hard question on T1 and two easy questions on T2. S1 is likely to answer the hard question on T1 correctly and one easy questions on T2 correctly (presuming that S1 learns concepts in T2 by doing questions). Thereafter, S1 will be offered the remaining three questions in the homework from T2 only. In all, S1 receives seven questions from the topic (T2) that she did not understand well, and only three questions from the topic (T1) that she understood well. Similarly, S2 would receive seven questions from T1 and three questions from T2. Thus, the adaptive homework personalizes the quantity and difficulty levels of questions to cater to the learning needs of individual students. Through such homework, students are likely to achieve higher "Learning by Doing" by solving more questions on the topics that they did not understand through class room instructions.

The concept of adaptive learning and its enabling technologies are widely discussed in the context of higher education in the US. The basic idea of adaptive learning is to adjust the learning resources to a learner's demonstrated performance level and thus maximize one's learning. A variety of adaptive learning tools are available that learn the way students learn, and thus adjust content in real time, or provide customized exercises when they need it. Some notable examples are: MOOCulus – a Calculus course on MOOC offered at Ohio State University that feeds progressively harder questions to students based on their previous answers (go.nmc.org/ulus); Flat World Education (go.nmc.org/flatm); and INTUITEL – that employs adaptive learning to provide optimal guidance and feedback to students. While there are a number of adaptive learning tools available in the market, there is very little rigorous evidence of their efficacy. The reported positive effects of adaptive learning are mostly anecdotal (Zimmer 2014, Waters 2014, Fain 2014).

We conduct a randomized field study in a group of secondary schools in India, a developing country with limited resources to educate the second-largest human population in the world. Teachers at the secondary schools in India have neither the training nor the motivation to improvise or personalize their class room instructions as per student needs, and hence their instructions are likely to deliver highly heterogeneous learning across students. In such settings, the role of adaptive homework in efficiently imparting "Learning by Doing" is extremely crucial. We add to the academic literature on ICT and education by measuring the causal effect of personalized learning through adaptively-generated homework on student cognitive achievements.

3.0 Field Setup Description

We conducted a field study in two (one girls and one boys) residential schools in Hyderabad India. These schools are part of a group of 97 schools in different districts of the state of Andhra Pradesh governed by Andhra Pradesh Residential Educational Institutions (APREI) Society. These schools offer admissions to children from economically backward families (annual household income less than Indian Rupees 60,000.00 or US $ 1,000.00) in 5th grade. The admission process of these schools is advertised in the local newspapers in the Hyderabad metropolitan area, a large metropolitan area in India with population of over 11.5 million.[3] The schools receive over 25,000 valid applications for admissions in 5th grade in a year and select 240 students (80 for each school) by randomly drawing names from the applicant pool as per the Right to Education Act and APREI Society guidelines.[4] Once selected, the students continue in these residential schools till 12th grade. The parents of past admitted students are less educated (95% of the parents have completed less than 12th grade education) and are involved in marginal occupations, such as, farmer, carpenter, mason, laborer, fisherman, vegetable vendor, cab/lorry drivers, security guard, and tailor. Therefore, students selected from random lottery in our setting provide us the representative student population of economically and educationally backward households in a large metropolitan area in a developing country. The findings of our study would be especially applicable to such less-advantageous student populations in developing countries – the main focus of our study.

Students in these residential schools have fixed time allocations for different daily activities, such as meals, class room instructions, homework, self-study, physical activities, recreation, sleeping. Moreover, as students do not have daily interactions with their family and siblings, the unobserved influence of family on their cognitive achievements was absent in the study. This is an important point, because most of the prior EPF studies found the absence of data on family influence to be a significant confounding factor in evaluating the effect of school inputs on students' educational achievements. Therefore, the institutional arrangements in the field setup helps us cleanly identify the effect of CGAHW on students' cognitive achievements.

[3] http://www.indiaonlinepages.com/population/hyderabad-population.html

[4] The guidelines for a random lottery and the steps required to ensure transparency in this process are laid out in the APREI Society circular available at http://www.apresidential.gov.in/Circular%20Docs/V%20Admissions%20-%20Guidenlines-final%20for%20website.pdf

In our field study, four sections of 7th grade and two sections of 8th grade, were included in the two schools. Therefore, we had a total of 240 students (approx. 40 students per section for six sections) in our experiment. Out of these students, half of the students (20 students) in each of the six sections were randomly selected to get CGAHW and the remaining half in each section got PBTHW in mathematics. Therefore, 120 students were assigned to CGAHW and the remaining 120 students to PBTHW. A total of eight students, two assigned to PBTHW and six assigned to CGAHW, left school in the middle of the experiment period, and thus we have a sample of 118 students receiving PBTHW and 114 students receiving CGAHW in our analysis.[5]

3.1 Homework Generation

A comprehensive topic-wise question bank was developed from the mathematics syllabus of Andhra Pradesh Education Board and Central Board of Secondary Education (CBSE) at the 7th and 8th grade levels. The list of topics covered in the mathematics syllabus are provided in Appendix A. All questions were in a multiple-choice format, where correct answer was to be chosen from four (or sometimes five) alternative answers. Each question in the question bank was tagged to a: (1) specific topic/subtopic; and (2) hard or easy category. The questions were tagged to a topic/subtopic based on the chapter of the book from which it was taken. We used the following objective criteria to tag a question to easy or hard categories – if a single rule/formula is required to solve a question then it is an easy question, but if multiple rules/formula in multiple steps are required to solve a question then it is a hard question.[6] After initial tagging of questions, a group of experienced secondary school mathematics teachers were employed to verify the tagging of questions in the question bank.

The total number of questions offered and their topic-wise breakup in a PBTHW were decided by the mathematics class teacher in a section. Thereafter, these pre-decided number of questions on different topics were randomly drawn from the question bank on those topics and offered in paper-based format to students in PBTHW. Therefore, all students receiving a PBTHW received the same questions and the proportions of easy to hard questions on different topics in a PBTHW were the same as that in the question bank.[7] The students solved PBTHW in the allocated class in the afternoons, and they could

[5] With such marginal attrition in our student sample, our study does not suffer from a sample selection problem due to systematic attrition like many prior EPF studies (Becker and Walstad 1990).

[6] Example - hard question: For a given circumference, find the area of a circle. Easy question: For a given radius, find the area of a circle. To solve the hard question, radius of the circle is to be computed from the circumference formula, and then area of the circle can be computed from the radius. Only the second step is required to solve the easy question.

[7] An alternative PBTHW would be to randomly draw questions for each student separately, so that each student received different questions in PBTHW but all students received the same total number and topic-wise breakup of questions in their PBTHW. However, it was not feasible in our field setup because it would have required class

consult their mathematics books and class notes and use scrap paper to do rough work. The students were required to provide their choice of correct answers in the answer sheet.

In contrast, CGAHWs were generated on students' laptop computers. These laptop computers were provided to the schools by INTEL in collaboration with the education department of the state of Andhra Pradesh. Each CGAHW had the same total number of questions as the corresponding PBTHW on the same topics. In CGAHW, half of the total questions – called base questions – were first randomly drawn in the easy category from the question bank on those topics. The topic-wise breakup of base questions in a CGAHW were kept the same as in the corresponding PBTHW. Thereafter, the remaining 50 percent questions in the CGAHW were drawn from the question bank, based on the performance of the students in base questions, such that more questions were offered on the topics in which they missed base questions, and offered hard questions only when they answered easy questions on that topic correctly. The detailed algorithm of generating adaptive questions is provided in Appendix B. Thus, students could answer the base questions in CGAHWs in any sequence, but they had to answer adaptively-generated questions in the sequence in which they were generated. As a result, different students receiving CGAHW, received the same number of total questions, but different numbers of easy and hard questions, based on their performance.

The questions in CGAHW appeared in exactly the same format (letter size and fonts) as PBTHW and CGAHW students did not receive computer-generated hints on how to solve a question.[8] Similar to PBTHW students, CGAHW students solved their homework in allocated afternoon classes, and they could consult their mathematics book and class notes and use scrap paper to solve homework questions. The CGAHW students submit their answers by ticking the correct alternative answers on their laptop computers. Therefore, available time, resources, appearance of questions, and the process of solving questions were similar in the two categories of homework. The only differences between the two categories were: (1) personalized question offerings as per the student needs in CGAHW versus fixed questions in PBTHW and (2) while PBTHW students could answer questions in any sequence, CGAHW students had to answer adaptively-generated questions in the sequence in which they were generated.

---

teacher to grade different PBTHW for different students. Another, option was to generate fixed PBTHW for students but make it available to students on the computer (instead of paper-based format), to control for any additional student engagement/learning due to computer display of questions. However, we could not do it because of the availability of limited number of laptops in our study.

[8] Most contemporary adaptive-learning applications use multimedia tools to make the educational content more engaging and offer helpful hints on included concepts to guide students to learn these concepts in a gradual fashion. Since we seek to examine the effect of only offering personalized questions to students on their learning, we did not include these functionalities in our adaptive algorithm.

The students in the four sections of 7th grade were offered homework from a total of 25 different homework covering the 7th grade mathematics syllabus and likewise in the two sections of 8th grade from 25 different homework covering the 8th grade mathematics syllabus.

4.0 Data Description

We conducted our experiment in the two schools from October 2014 to April 2015. During this period, we collected data on each homework offered, such as unique homework number; topics covered in the homework; number of easy/hard/total questions offered in each PBTHW (number of easy/hard questions in a PBTHW were same for all students), number of easy/hard/total questions offered to CGAHW students (number of easy/hard questions offered to CGAHW students varied based on their performance), and number of correct easy/hard/total answers for each student in each homework.

Table 1: Summary of No. of completed homework

No. of completed HW	No of Students	Oct-Dec 2014			Jan-April 2015		
		Mean	Std. Dev.	*t-stats*	Mean	Std. Dev.	*t-stats*
PBTHW	118	6.99	0.81	2.38	9.52	2.69	1.68
CGAHW	114	6.62	1.46		8.91	2.83	

Students in different class sections were assigned different numbers of homework by their class teachers but both categories of students in a section received same numbers of homework. Also, some students did not complete all homework assigned to them. Therefore, we observe a variation in numbers of homework completed by students in our sample. The students were assigned homework on topics that were taught from October to December 2014 and then tested on those topics in a half-yearly exam at the end of December 2014. Thereafter, the students were assigned homework on further topics taught from January to March 2015 and tested on the entire course in a final summative exam in April 2015. The summary statistics of the mean number of CGAHW and PBTHW completed by students in each period is provided in Table 1. We find that on average, PBTHW students completed more homework than do their CGAHW counterparts.

Table 2: Distribution of No. of completed homework

No. of HW per student	HW type	No of Students	Percentile Value				
			0	25	50	75	100
Oct-Dec 2014	*PBTHW*	118	6	6	7	8	8
	CGAHW	114	1	6	7	8	9
Jan-April 2015	*PBTHW*	118	7	7	10	13	13
	CGAHW	114	1	7	7	12	13

In addition to the mean number of homework, in Table 2, we also find that CGAHW students completed fewer number of homework at all percentiles of their distribution than that of PBTHW students.

We further compare the total number of questions and the proportion of easy questions offered in the two categories of homework in Table 3. Since the total number of questions in corresponding homework of the two categories were kept the same, we find same total number of questions at all percentile values for the two categories of homework. But, we find that a higher proportions of easy questions and thus lower proportions of hard questions were offered in CGAHW as compared to PBTHW at all percentile values.

Table 3: Distribution of homework questions

No. of Qs	HW type	No. of student-HW observations	Percentile Value				
			0	25	50	75	100
All Qs	*PBTHW*	1949	12	20	20	26	30
	CGAHW	1770	12	20	20	26	30
Easy Qs	*PBTHW*	1949	0.22	0.5	0.6	0.8	1
	CGAHW	1770	0.10	0.6	0.88	1	1

4.1 Homework Performance

Students are likely to incorrectly answer homework questions on topics that they have not understood well during class room instructions, and adaptive software would, in turn, offer them more questions on those topics. Moreover, CGAHW allows gradual learning by first offering easy questions on a topic and offering hard questions only when easy questions on that topic have been correctly answered. In contrast, a PBTHW offers a fixed number of easy and hard questions on topics covered in that homework. Therefore, as compared to the corresponding PBTHW, students are likely to receive greater number (proportion) of total/easy questions on topics that they have not understood well and smaller number (proportion) of total/easy questions on topics that they have understood well in a CGAHW. Thus, we expect that an average CGAHW student will score lower and receive greater number (proportion) of easy questions as compared to PBTHW students.

We seek evidence of the above intuition in our data. In Table 3, we saw evidence of a greater number (proportion) of easy questions being offered in CGAHW as compared to PBTHW. In Table 4, we provide the distribution of percentage of correctly answered total questions and correctly answered easy questions separately for the two categories of students. It is evident from Table 4 that PBTHW students on average, correctly answer a higher percentage of easy as well as total questions as compared to their

CGAHW counterparts at all percentile values of their distribution. Thus the data support our intuition on adaptive generation of questions and student performance in CGAHW.

Table 4: distribution of homework scores on easy and total Qs

Homework Scores	HW type	No. of student-HW observations	Percentile Value				
			0	25	50	75	100
Percentage of correctly answered Total Qs	*PBTHW*	1949	0	40	64	80	100
	CGAHW	1770	0	25	40	60	100
Percentage of correctly answered Easy Qs	*PBTHW*	1949	0	38	67	88	100
	CGAHW	1770	0	25	43	67	100

5.0 Econometric Analysis and Results

Drawing from the general specification of an educational production function (Hanushek 1979), the cognitive achievement for a student *i* of category *j* in period *t* (A_{it}^{j}) is given as:

$$A_{it}^{j} = f(\sum F_{it}^{j}, \sum P_{it}^{j}, \sum S_{it}^{j}, I_{i}^{j}) \quad ; \quad \text{-- (1)}$$

where, $\sum F_{it}^{j}$ indicates the vector of family background influences on student *i* of category *j* cumulative to period *t*; $\sum S_{it}^{j}$ indicates the vector of school inputs for student *i* of category *j* cumulative to period *t*; $\sum P_{it}^{j}$ indicates the peer effect for student *i* of category *j* cumulative to period *t*; and I_{i}^{j} indicates the genetic endowed mental abilities of student *i* of category *j*. Similarly, the student's achievement in period (*t*-1) can be written as:

$$A_{i(t-1)}^{j} = f(\sum F_{i(t-1)}^{j}, \sum P_{i(t-1)}^{j}, \sum S_{i(t-1)}^{j}, I_{i}^{j}) \quad \text{-- (2)}$$

If the student's cognitive achievement in the previous period is assumed to be a sufficient statistics for the history of inputs till the prior period and the student's genetic endowed ability, we can combine (1) and (2) to write the educational production function in the value added form as:

$$A_{it}^{j} = f(F_{it}^{j}, P_{it}^{j}, S_{it}^{j}, A_{i(t-1)}^{j}, \varepsilon_{it}^{j}) \quad ; \quad \text{-- (3)}$$

where, F_{it}^{j} , P_{it}^{j}, and S_{it}^{j} are, respectively, the contemporaneous values of family background, peer effect, and school resources for student *i* of category *j* in period *t*. The term ε_{it}^{j} represents the idiosyncratic error. Assuming the arguments of (3) to be additively separable and their parameters to be non-age varying, as is generally assumed in the past literature (Todd and Wolpin 2003, Krueger 2000), the educational production function can be further simplified as:

$$A_{it}^{j} = \alpha F_{it}^{j} + \beta P_{it}^{j} + \delta S_{it}^{j} + \gamma A_{i(t-1)}^{j} + \varepsilon_{it}^{j} \qquad \text{-- (4)}$$

Next we apply the details of our experimental setup to simplify the above educational production function. In our setup, student *i* comes from the population of students in six sections of two residential schools. We randomly assign half of the students in each section to receive CGAHW and the remaining students to receive PBTHW (i.e., *j* = CGAHW, PBTHW). We measure the cognitive achievements of a student in mathematics A_{it}^{j} with her mathematics exam scores in academic year 2014-15 (denoted as $EScore_{it}$), as most of the economics of education literature considers the standardized test scores an appropriate measure of students' cognitive achievements (Hanushek 1979). The students' cognitive achievements during the experiment period is influenced by the contemporaneous inputs (family environment, school resources, peer effects) they receive in this period and their cognitive achievements in mathematics $A_{i(t-1)}^{j}$ prior to the experiment period, as reflected in their mathematics exam scores in academic year 2013-14 (denoted as $EScore_{i(t-1)}$). Prior studies on educational production function indicate that student baseline scores on similar tests at the beginning of study are a good control for omitted history of inputs and student genetic endowed capacity (Todd and Wolpin 2003, Hanushek 2003). The mathematics syllabus at the middle school level gradually builds up with the concepts learnt in 6th grade are built upon in the concepts learnt in 7th and 8th grades. Therefore, students' achievements in similar mathematics exams in the previous grades are determined by the history of all relevant inputs they received prior to the experiment period and their genetic endowed ability in mathematics.

As students resided in school dormitories in our experiment, the influence of parent/siblings and family environment on students' cognitive achievements were absent, i.e., $F_{it}^{CGAHW} = F_{it}^{PBTHW} = 0$. Even if we assume some effects of parent/siblings due to their occasional interaction with students during school holidays, these effects were likely to be similar for the two categories of students, due to the random assignment, i.e., $E(F_{it}^{CGAHW}) = E(F_{it}^{PBTHW})$. Next, as students were randomly assigned to the two categories in each section included in the experiment, there was no systematic difference in the peer effects for the two categories of students, i.e., $E(P_{it}^{CGAHW}) = E(P_{it}^{PBTHW})$. Moreover, the school resources such as class size, class time, books, other school facilities, and teachers for the two categories of students were similar. The only school resource that was different for the two categories of students was the offered homework. If we assume the additive separable influence of different school resources on students' cognitive achievements, the total effect of school resources on students receiving CGAHW would be $S_{it}^{CGAHW} = S_{it}^{PBTHW} + CGAHW_{it}$, where $CGAHW_{it}$ indicates the differential effect of CGAHW on students' cognitive achievements as compared to PBTHW. Therefore, equation (4) can be simplified to the following OLS specification:

$$EScore_{it} = \beta_0 + \beta_1.CGAHW_{it} + \gamma.EScore_{i(t-1)} + \varepsilon_{it} \quad ; \qquad \text{-- (5)}$$

where, $CGAHW_{it}$ is an indicator variable equal to one if student i is assigned CGAHW and zero otherwise. Since students were randomly assigned to either CGAHW or PBTHW, the error term in specification (A) is uncorrelated to the right-hand-side variables, $CGAHW_{it}$ and $EScore_{i(t-1)}$, and hence the coefficient estimates of specification (5) would be unbiased.[9] Overall, the coefficient estimate of $CGAHW_{it}$ in specification (5) would show the causal effect of CGAHW on students' cognitive achievements in mathematics.

5.1 Comparison of Prior Mathematical Abilities

We collected students' scores (percentage marks obtained) in the half-yearly and yearly summative mathematics exams in the academic year 2013-14, a year prior to the experiment period. These exams were paper-based and contained descriptive questions. The question papers for these exams were set by a central authority as per the common standards at secondary school level. The mathematics syllabus from 6^{th} to 8^{th} grades gradually builds up, such that 7^{th} grade mathematics topics utilize the concepts covered in 6^{th} grade mathematics topics as the starting point. Therefore, students' scores in exams at a grade level measure their mathematical abilities in concepts covered up to that grade level, and it is a sufficient statistics for the history of inputs (such as family environment, peers, and school resources) up to that grade level and their genetically endowed abilities. Table 5 shows the mean exam scores and the *t-statistics* for the difference in means for the two categories of students. The *t-value* of 0.78 indicates that the mean scores for the two categories of students are statistically indistinguishable.

Table 5: Comparison of previous year mean exam scores

Previous Year Exam Scores	No. of Students	Mean	St. Dev.	*t-value*
PBTHW	118	47.82	20.24	0.78
CGAHW	114	49.99	22.11	

We also compare the distributions of the prior year mean exam scores for the two categories of students in Figure 1, and find them to be similar over the whole distribution. Moreover, we find a wide variation in the prior year exam scores for the two categories of students, which we utilize in our subsequent analysis to examine the effect of CGAHW on students of different abilities.

[9] Omission of baseline achievement variable $EScore_{i(t-1)}$ should not affect the coefficient of $CGAHW_{it}$, as the CGAHW is randomly assigned to students. However, inclusion of $EScore_{i(t-1)}$in specification (5) increases the precision of coefficient of $CGAHW_{it}$.

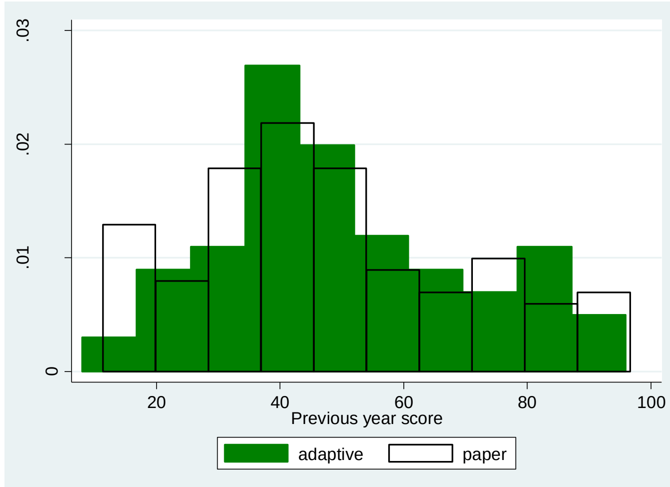

Figure 1: Histogram of previous year mean exam scores

5.2 Evidence of Personalization in Computer Generated Homework

So far we have shown that CGAHW students on average, received a greater number of easy questions, and obtained lower scores than that of their PBTHW counterparts. Among CGAHW students, students who obtained higher (lower) scores, i.e., correctly answered higher (lower) percentage of questions, got fewer (more) easy questions. In other words, the number (proportion) of easy questions offered to students in CGAHW should be negatively correlated with their homework scores. Since 114 students solve 50 different CGAHW, we need to control for the differences in unobserved characteristics of homework and students. Accordingly, we estimate the relationship between number (proportion) of easy questions offered and scores obtained in CGAHWs with the following fixed effects specification:

$$(NQeasy\ or\ PQeasy)_{ji} = \beta_i + \beta_j + \beta_1 \times HWScore_{ji} + \varepsilon_{ji} \quad ; \quad \text{------ (6)}$$

where, *i* denotes 114 CGAHW students and *j* denotes 50 different CGAHW assigned to these students. As different students completed different numbers of homework, we have a total of 1770 student-homework observations in our data (instead of 114x50 student-homework observations). Variable $HWScore_{ji}$ denotes the percentage of correct answers given by student *i* in homework *j*; $EScore_{i(t-1)}$ denotes the exam score of student *i* in academic year 2013-14; $NQeasy_{ji}$ and $PQeasy_{ji}$, respectively, denote the number and proportion of easy questions offered in CGAHW *j* to student *i*. Parameters β_j and β_i, respectively, denote the homework and student fixed effects. Thus, coefficient β_1 in specification (6) identifies the effect of students' homework scores on the number (proportion) of easy questions received in those homework. A negative and significant estimate for coefficient for *HWScore* (β_1) would provide evidence for the adaptive question generation in CGAHW.

Table 6: Effect of HW scores on No. of easy questions offered

	Prop. of Easy Questions			No. of Easy Questions		
	Coefficient	St. Err.	*t-value*	Coefficient	St. Err.	*t-value*
HW Score	-0.005***	0.000	-17.91	-0.054***	0.003	-17.94
Intercept	0.971***	0.032	29.71	10.03***	0.329	30.47
N	1770			1770		
HW Fixed effect	Yes			Yes		
Student Fixed effect	Yes			Yes		

∗∗∗, ∗∗, and ∗ denote statistically significant at α=0.01, 0.05, and 0.10 levels (two-sided test), respectively.

The coefficient estimates for specifications (6) are shown in Table 6. From Table 6, we find negative and highly significant estimates for (β_1), which provides support for adaptive generation of questions in CGAHW. The coefficient estimate of -0.005 for variable *HW Score* indicates that a 10 percent higher HW score would reduce the number of easy questions offered in a CGAHW by five percent.

5.3 Exam Performance

Next we compare the performances of the two categories of students in mathematics exams during the experiment period. During the experiment period, a half-yearly exam in December 2014, and a yearly summative exam in April 2015, were conducted. These exams were paper-based and contained descriptive questions. The questions in these exams were set by a central authority (not by the class teacher) as per the common standards of Andhra Pradesh Education Board at 7th and 8th grade levels. Therefore, these exams were akin to standardized test measuring the cognitive performance of students as per the required standards in mathematics at the secondary school level.

Table 7: Comparison of exam scores

HW type	N	Mean Exam Score	Std. Dev.	*t-value* for diff in means	Percentile Value				
					0	25	50	75	100
PBTHW	236	51.86	23.05	2.19	1	37	49	70	98
CGAHW	228	56.64	22.91		6	40	56	74	99

The mean exam scores (percentage marks obtained) by the two categories of students in these exams are reported in Table 7. We find that CGAHW students obtained statistically higher (significant at α=0.05) exam scores than that of their PBTHW counterparts. To rigorously test this fact, we use regression specifications in the following section.

We divide the students into terciles based on their previous year mean exam scores in mathematics, and categorize the students in these terciles as low, medium, and high mathematical ability students. The mean exam scores were below 38 for low ability students, between 38 and 53 for medium ability students, and above 53 for high ability students. The summary statistics of previous year exam scores for the two categories of students in each ability category are provided in columns 4 and 5 of Table 8.

Table 8: Summary statistics for students with different abilities

Student Type	HW Type	No of Students	Previous Year Exam Score		Mean HW Score	
			Mean	St. Dev	Mean	St. Dev
Low Ability	*PBTHW*	41	24.64	7.99	52.73	21.38
	CGAHW	33	27.38	8.03	33.26	12.98
Medium Ability	*PBTHW*	39	45.73	4.25	57.12	16.31
	CGAHW	42	45.22	4.15	41.90	15.54
High Ability	*PBTHW*	38	73.89	13.32	58.90	14.37
	CGAHW	39	72.79	11.82	46.32	14.34

In columns 6 of Table 8, we provide the mean homework scores (i.e. average of all homework scores for a student). We find that students of higher ability obtain higher mean homework scores, and PBTHW students obtain higher homework scores than that of CGAHW students in each ability category.

We saw that PBTHW students on average, completed higher number of homework. Since, students learn by doing homework, their exam scores are likely to be affected by the number of homework they complete. Therefore, we add the number of homework completed as a right-hand-side variable in our EPF specification (5) to examine the effect of CGAHW on students' exam scores:

$$EScore_{ik} = \beta_1 + \beta_2 \times CGAHW_i + \varepsilon_{ik} \quad ; \qquad \text{---- (7a)}$$

$$EScore_{ik} = \beta_1 + \beta_2 \times CGAHW_i + \beta_3 \times NHW_{ik} + \beta_4 \times EScore_{i(t-1)} + \varepsilon_{ik} \quad ; \qquad \text{---- (7b)}$$

where, *i* denotes students and *k* denotes exams (half-yearly, and yearly summative). Variable $EScore_{ik}$ indicates the exam scores for student *i* in exam *k*; $EScore_{i(t-1)}$ indicates the mean score obtained by student *i* in the previous year exams; NHW_{ik} indicates the number of homework completed by student *i* prior to exam *k*; and $CGAHW_i$ is an indicator variable equal to one if student *i* received CGAHW and zero if she received PBTHW. We run specifications (7a) and (7b) for the whole sample of students, and separately for low, medium, and high ability students. The results of these regressions are reported in Table 9. The standard errors are cluster corrected at the student level.

Table 9: Regression Results

Exam Scores	Coeff	St. Err.	t-value	Coeff	St. Err.	t-value
	All Students					
CGAHW	4.78**	2.18	2.19	4.28**	2.03	2.11
No. of HW				0.35*	0.19	1.76
Pr. Yr. Exam Scores				0.42***	0.05	8.48
Intercept	51.86***	1.53	33.83	27.59***	3.75	7.40
N	464			464		
	Low ability Students					
CGAHW	3.22	3.40	0.97	1.97	3.36	0.59
No. of HW				0.52	0.32	1.61
Pr. Yr. Exam Scores				0.53**	0.21	2.50
Intercept	40.90***	2.26	18.08	21.70***	7.01	3.10
N	148			148		
	Medium ability Students					
CGAHW	6.35*	3.38	1.88	6.59*	3.40	1.94
No. of HW				0.32	0.32	1.02
Pr. Yr. Exam Scores				0.17	0.42	0.41
Intercept	53.20***	2.32	22.98	41.34**	19.63	2.11
N	162			162		
	High ability Students					
CGAHW	1.91	3.81	0.49	2.80	3.80	0.74
No. of HW				0.17	0.37	0.47
Pr. Yr. Exam Scores				0.49***	0.16	3.10
Intercept	61.92***	2.82	21.96	23.26*	13.68	1.70
N	154			154		

∗∗∗, ∗∗, and ∗ denote statistically significant at α=0.01, 0.05, and 0.10 levels (two-sided test), respectively.

For specification (7b) on the full sample of students, we find a positive and significant estimate of 4.28 for *CGAHW*, which indicates that CGAHW students obtained 4.28 percent higher marks than that of PBTHW students. The estimated value of 4.28 translates into an eight percent improvement on the mean exam score of 51.86 for PBTHW students.[10]

We find a positive and significant estimate of *CGAHW* for only medium ability students in our split regressions on students of different abilities, both with and without controlling for previous year exam scores. Also, we find a higher magnitude of *CGAHW* coefficient for medium ability students (6.59) than that for low and high ability students (1.97 and 2.80, respectively). This suggests that the medium ability students primarily gain from receiving CGAHW but not the low and high ability students.

[10] We further added section fixed effects in specification (7a) and (7b) to account for differences in class room instructions by different teachers in different sections and found similar coefficient estimates (4.16) for CGAHW.

5.4 Effect of Personalized Learning on Exam Performance

So far we have estimated the treatment effect from an indicator treatment variable for CGAHW. Next, we provide empirical evidence that the improvement in students' exam scores are due to personalization they received in CGAHW. So, we need to measure the personalization received by students in CGAHW. Students receive personalization in two dimensions in CGAHW: (1) topics - more questions offered on topics in which students incorrectly answered questions and (2) level of difficulty - hard questions offered on a topic only when easy questions on that topic were correctly answered. Students receive a greater number of easy questions in their weak areas in CGAHW as compared to the PBTHW, as per their learning needs. Therefore, the difference in total number of easy questions offered in a CGAHW from that in the corresponding PBTHW captures the offered personalization in both dimensions in the CGAHW. We denote this value as *DfNQeasy*, and as per this definition it is zero for PBTHW. In Appendix C, we show through several examples that *DfNQeasy* value in a CGAHW captures personalization in both dimensions. We also show in Appendix C that if students understand the topics covered in the homework well, they would receive a greater (smaller) number of hard (easy) questions in CGAHW leading to a negative value of *DfNQeasy*. But this is as per the students' learning needs and hence the magnitude of *DfNQeasy* (and not its sign) measures the level of offered personalization in such CGAHW. Therefore, we use the absolute value of *DfNQeasy* (denoted as |*DfNQeasy*|) to measure the level of personalization offered in CGAHW. For 1770 student-CGAHW observations in our data, the *DfNQeasy* values at 0, 25, 50, 75, and 100 percentiles were -8, 0, 4, 6, and 16, respectively. This indicates that while *DfNQeasy* values are largely positive, about 20 percent of the CGAHWs had negative *DfNQeasy* values.

While a higher |*DfNQeasy*| value suggest a higher level of offered personalization, it may not necessarily mean higher learning. For instance, students may be offered a greater number of easy questions on a topic when they incorrectly answer the earlier homework questions on that topic. But, if students don't consult book and class notes to learn the relevant concepts, they may keep incorrectly answering the later homework questions on that topic, and, in turn, be offered more easy questions on that topic. In such cases, students receive higher levels of personalization (high |*DfNQeasy*| value) in the homework, but obtain lower homework scores. Therefore, higher offered personalization (i.e., higher |*DfNQeasy*| value) in a homework indicates higher personalized learning if student's score in the homework is considered. Using this insight, we identify the effect of personalized learning on students' exam scores by comparing the exam scores of students with similar homework scores but different |*DfNQeasy*| values in their homework. If personalized learning affects students' exam scores positively, students with higher |*DfNQeasy*| values in their homework should obtain higher exam scores.

As students received different levels of personalization in different homework, the total personalization received by student *i* that would affect her performance in exam *k*, denoted by $|Personalization|_{ik}$, was computed as $\sum|DfNQeasy|_{ik}$ / NQ_k. Where, $\sum|DfNQeasy|_{ik}$ is the sum of $|DfNQeasy|$ values for all homework that student *i* completed prior to exam *k*, and NQ_k is the sum of total numbers of questions in all homework offered to students in their grade level prior to exam *k*. Thus, value of $|Personalization|$ variable for a CGAHW student is effectively the difference in proportion of easy questions received in all CGAHW students than that her PBTHW counterpart, and for a PBTHW students is zero. Likewise, we computed the $Personalization_{ik}$ variable for student *i* for exam *k* based on $\sum DfNQeasy_{ik}$ values for her homework (instead of $\sum|DfNQeasy|_{ik}$). We also computed the mean scores of all homework that student *i* completed prior to exam *k* and denote it as $MHWScore_{ik}$. The summary statistics of these variables are reported in Table 10.

Table 10 indicates a high variation in personalization variable across CGAHW students. The personalization variable is zero for PBTHW students and thus not shown in Table 10. The negative values of $Personalization_{ik}$ indicate lower proportion of easy questions (and hence more hard questions) offered in CGAHW as compared to the corresponding PBTHW. Similar to the findings in earlier sections, we find that the mean homework scores for CGAHW students are lower than that of PBTHW students over the entire distribution.

Table 10: Distribution of Aggregate Personalization and HW Scores

	HW Type	No. of Students-exam Observations	Percentile Values				
			0	25	50	75	100
|Personalization|	CGAHW	228	0.00	0.20	0.26	0.31	0.54
Personalization	CGAHW	228	-0.21	0.09	0.17	0.26	0.40
MHWScore	PBTHW	236	0	47.5	56.9	67.3	95.7
	CGAHW	228	0	29.4	40.5	51.4	89.5

Next we examine the effect of varied levels of offered personalization and personalized learning in CGAHW on students' exam scores with the following set of specifications:

$$EScore_{ik} = \beta_1 + \beta_2 \times |Personalization|_{ik} + \varepsilon_{ik} \quad ; \qquad \text{----- (8a)}$$

$$EScore_{ik} = \beta_1 + \beta_2 \times |Personalization|_{ik} + \beta_3 \times NHW_{ik} + \beta_4 \times MHWScore_{ik} + \beta_5 \times EScore_{i(t-1)} + \varepsilon_{ik} \quad ; \qquad \text{----- (8b)}$$

where, *i* denotes students and *k* denotes exams (half-yearly and yearly summative). All other variables have the same meaning as mentioned earlier in the paper. The coefficient estimate of $|Personalization|$ in specifications 8(a) and 8(b), respectively, indicates the effect of offered personalization on exam scores without and with controlling for the mean homework scores. Coefficient estimates for these specifications are reported in Table 11.

Table 11 – Estimates for effect of levels of Personalization on Exam Scores

Exam Scores	Coeff. Est.	St. Err.	t-value	Coeff. Est.	St. Err.	t-value	Coeff. Est.	St. Err.	t-value
	All Students						CGAHW Students		
|Personalization|	14.07^{*}	7.63	1.84	31.07^{***}	7.73	4.02	37.21^{*}	19.33	1.93
MHWscore				0.29^{***}	0.07	4.37	0.53^{***}	0.11	4.99
No of HW				0.12	0.19	0.63	-0.36	0.31	-1.14
Pr. Yr. Exam Scores				0.37^{***}	0.05	7.39	0.31^{***}	0.08	3.9
Intercept	52.47^{***}	1.47	35.81	16.67^{***}	4.60	3.62	14.21^{**}	6.90	2.06
N	464			464			228		

∗∗∗, ∗∗, and ∗ denote statistically significant at α=0.01, 0.05, and 0.10 levels (two-sided test), respectively.

We find a positive and significant estimate of personalization variable (|*Personalization*|) in specification 8(a), which indicates that students obtain higher exam scores with higher offered personalization in their homework. After controlling for the mean homework scores in specification 8(b), we find a higher and more significant coefficient estimate for |*Personalization*|, which indicates that when students learn from offered personalization in their homework (and thus obtain similar scores in CGAHW as by their PBTHW counterparts), they obtain higher exam scores. We also estimated specification 8(b) on the sample of CGAHW students only, and find a positive and significant estimate for the personalization variable. This indicates that for the same mean HW scores, CGAHW students receiving higher personalization obtain higher exam scores. These results indicate that if students learn from offered personalization in CGAHW, they obtain higher scores in exams.

Table 12 – Effect of Personalized Learning on students of different ability categories

Exam Scores	Coeff. Estimate	St. Err.	t-value	Coeff. Estimate	St. Err.	t-value	Coeff. Estimate	St. Err.	t-value
	Low ability Students			Medium ability Students			High ability Students		
|Personalization|	24.80^{*}	13.03	1.90	44.19^{***}	12.91	3.42	21.29	13.14	1.62
MHWscore	0.25^{***}	0.10	2.64	0.37^{***}	0.11	3.33	0.25	0.16	1.51
No. of HW	0.36	0.31	1.17	0.06	0.31	0.21	-0.06	0.37	-0.15
Previous Yr. Score	0.41^{*}	0.21	1.95	0.01	0.42	0.03	0.42^{**}	0.17	2.49
Intercept	13.24	8.23	1.61	31.00	20.28	1.53	16.77	13.30	1.26
N	148			162			154		

∗∗∗, ∗∗, and ∗ denote statistically significant at α=0.01, 0.05, and 0.10 levels (two-sided test), respectively.

Students, who score high in CGAHW, are offered fewer easy (more hard) questions, and thus the *Personalization* variable for them may have negative value. We separately estimated specification (8a) with the right-hand-side variable as *Personalization* instead of |*Personalization*| for only those students who have a *Personalization* value of less than or equal to zero. Thus, we identify the difference in exam scores of students doing well in CGAHW (negative *Personalization* value) from that of students receiving PBTHW (*Personalization* =0). We find a coefficient estimate of -192.91 for variable *Personalization* (significant at $\alpha = 0.01$ level), which indicates that higher negative values of *Personalization* would lead to a higher exam scores.[11] In other words, CGAHW students, who receive a greater number of hard questions in the homework, score higher in exams than PBTHW students. Thus, personalized offerings of both hard and easy questions positively affect students' performance in exams.

So far we have compared the exam scores of CGAHW students receiving different levels of personalization in their homework with that of PBTHW students after controlling for their mean homework scores. However, we find that in general CGAHW students obtain lower homework scores at all percentile values of the homework scores distribution (see Table 4). So we examine that at how much lower mean homework scores as compared to their PBTHW counterparts, CGAHW students still benefit from higher offered personalization in CGAHW. For this analysis, we artificially bumped up the homework scores of CGAHW students by 10 percent at a time and then compare the exam performance of the two categories of students in specification (8b). So with a 10 percent bump in homework scores of CGAHW students, we estimate the effect of personalization received by CGAHW students by comparing their exam scores with that of PBTHW students, who obtained 10 percent higher mean homework score. We find a positive and significant estimate of |*Personalization*| in specification (8b) with forty percent bump in mean homework scores of CGAHW students, which indicates that CGAHW students do better in exams with offered personalization than their PBTHW counterparts even with forty percent lower mean homework scores.

We further examine how the effect of levels of personalization on students' exam scores differ across students of different abilities. For this analysis, we estimate specifications (8b) on low, medium, and high ability students separately, and report the estimated coefficients in Table 12. We find similar estimates, in sign and significance, for the personalization variable (|*Personalization*|) in Table 12 to the estimates for CGAHW indicator variable in Table 9. In Table 12, we see that after controlling for mean homework scores, the coefficient for personalization variable are significant for both low and medium ability students. This means that for similar homework scores, low and medium ability CGAHW students, who are offered higher levels of personalization in homework, obtain higher exam scores than that of their PBTHW counterparts. In other words, when higher levels of personalization results in personalized learnings for low and medium ability students, they obtain higher exam scores. Interestingly, we do not

[11] The magnitude of coefficient for personalization variable is due to comparison of exam scores of only high ability CGAHW students with all PBTHW students in this specification.

find a significant effect of level of personalization on the exam scores of high ability students even after controlling for their mean homework scores. This suggest that the high ability students have already learned concepts from class room instructions and thus offering them personalized versus fixed homework does not make any difference in their learning. Overall, these results suggest that the personalized homework could reduce the achievement gap between high and low ability students.

6.0 Conclusion

We conducted a randomized field experiment on a group of residential schools to examine the effect of personalized learning through computer-generated adaptive homework on students' mathematics exam scores. We offered randomly selected half of the students CGAHW in mathematics and the remaining students were offered PBTHW. Besides the differences in their offered homework, the two groups of students had access to the same mathematics class teachers, school resources, allotted study time, and peer group. Moreover, being a residential school, the students' educational achievement were not influenced by their family environment (parents and siblings) during the study period. We found that while the CGAHW students on average, obtained lower homework scores than their PBTHW counterparts, they obtained 4.28 percent higher exam scores than their PBTHW counterparts. We computed a measure of aggregate personalization received by students in homework and show that students, who received higher levels of personalization in homework, obtained higher exam scores. Thus, we provide direct empirical evidence of higher "Learning by Doing" from personalization in CGAHW. We also found that student with prior mathematical abilities in low and middle categories benefit from the CGAHW.

Our study adds to the growing literature on effect of ICT on educational production. Specifically, we measure the causal effect of offering personalized homework questions to students on their cognitive achievements. To achieve this, we intentionally did not utilize other features of ICT that could enhance student learning. We also controlled for other confounding factors that could affect student learning through the institutional arrangements and randomization in our field setup. Thus, we show that merely offering personalized questions to students through an in-house designed and developed ICT application can substantially increase their cognitive achievements. Other features of ICT, such as interactively offering helpful clues to students in solving questions, can be easily added to our adaptive homework software, to further enhance student learning. Overall, our results have important implications for remedying the quality of education delivery in resource-strapped education systems. Our results are particularly relevant for the educational systems in developing countries that grapple with large class sizes, lower quality and number of teachers, and limited school resources. Moreover, our findings that

CGAHW helps improve the cognitive achievements of the low and medium ability students but not high ability students is policy relevant, because it suggests that personalized homework could reduce the achievement gap between high and low ability students.

We acknowledge that our present study has several limitations, but that provides opportunities of future research to address them. First, our limited sample size, even though randomly drawn from a large student population of economically poor families, is a limitation. The large estimated effect size of personalized homework could be due to the poor quality of class room instructions in the classes in our experiment. In schools with better class room instructions, the scope for improvement in students' cognitive achievements through personalized homework may be limited. Thus, while our central results that personalized homework improves students' cognitive achievements is generalizable, but the estimated effect size may vary across different field setups. Second, our results on effect of personalized homework in mathematics may not generalize to other subjects, such as language arts and literature. Further studies are required to better understand the effect of computer-generated adaptive homework in other subjects. One promising extension of the present research is to examine the effect of additional ICT functionalities with personalized homework on students' cognitive achievements. An examples of such additional ICT functionality is offering students illustrative visual clues to help them solve questions on a concept, and in the process enhance their learning on that concept. In a similar vein, ICT-mediated peer discussions among students while solving homework questions could also help in their learning. It would be interesting to design such a system and measure its efficacy in improving students' cognitive achievements.

References

Angrist, Joshua D., and Victor Lavy. 1999. "Using Maimondides' rule to estimate the effect of class size on scholastic achievement." *Quarterly Journal of Economics* 114(2) pp. 533-575.

Anzai, Y. and Herbert A. Simon, 1979. The Theory of Learning by Doing, *Psychological Review*, 86(2), pp. 124-140.

Arrow, Kenneth J. 1962. The Economic Implication of Learning by Doing. *The Review of Economic Studies*, 29(3), pp. 155-173.

Banerjee, A., E. Duflo, E., S. Cole, and L. Linden, 2007. Remedying Education: Evidence from Two Randomized Experiments in India, *Quarterly Journal of Economics*. 122(3) pp. 1235-1264.

Becker, William E. and Walstad, William B. 1990. Data Loss from Pretest to Posttest as a Sample Selection Problem. *Review of Economics and Statistics*, February 72(1), pp. 184-88.

Belo., R., P. Ferreira, R. Telang.(2014). Broadband in School: Impact on Student Performance. *Management Science* 60(2). 265-282.

Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York. 1966. Equality of Educational Opportunity. Washington, D.C. U.S. Government Printing Office.

Fain Paul. 2014. Learning to Adapt. Inside Higher Ed. June 13 2104 (available at https://www.insidehighered.com/news/2014/06/13/profits-lead-way-adaptive-learning-becomes-more-popular)

Hanushek, Eric A. 1979. "Conceptual and empirical issues in the estimation of educational production functions." *Journal of Human Resources* 14(3) pp. 351-388.

Hanushek, Eric A. 1992. "The trade-off between child quantity and quality." *Journal of Political Economy* 100(1) pp.84-117.

Hanushek, Eric A. 1999. "Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects." *Educational Evaluation and Policy Analysis* 21(2) pp.143-163.

Hanushek, Eric A. 2003. "The failure of input-based schooling policies." *Economic Journal* 113 (485) pp. F64-F98.

Hanushek, Eric A. 2007. Education Production Function, Palgrave Encyclopedia.

Hanushek, Eric A., and Dennis D. Kimko. 2000. "Schooling, labor force quality, and the growth of nations." *American Economic Review* 90(5) pp.1184-1208.

Hanushek, Eric A., Steven G. Rivkin, and Lori L. Taylor. 1996. "Aggregation and the estimated effects of school resources." *Review of Economics and Statistics* 78(4) pp. 611-627.

Krueger, Alan B. 1999. "Experimental estimates of education production functions." *Quarterly Journal of Economics* 114(2) pp. 497-532.

Krueger, Alan, and Cecilia Rouse, 2004 “Putting Computerized Instruction to the Test: A Randomized Evaluation of a ‘Scientifically-based’ Reading Program,” *Economics of Education Review*, pp. 323–338.

Lazear, Edward P. 2003. "Teacher incentives." *Swedish Economic Policy Review* 10(3):179214.

Leuven, Edwin, Mikael Lindahl, Hessel Oosterbeek, and Dinand Webbink, 2004 “The Effect of Extra Funding for Disadvantaged Pupils on Achievement,” IZA Discussion Paper No. 1122, 2004.

Machin, Stephen, Sandra McNally, and Olmo Silva, 2007 “New Technology in Schools: Is there a Payoff?” *The Economic Journal* 117(522) pp. 1145-1167

Mincer, Jacob. 1970. "The distribution of labor incomes: a survey with special reference to the human capital approach." *Journal of Economic Literature* 8(1) pp.1-26.

Mulligan, Casey B. 1999. "Galton versus the human capital approach to inheritance." *Journal of Political Economy* 107(6 pt. 2) pp. S184-S224.

Murnane, Richard J., John B. Willett, Yves Duhaldeborde, and John H. Tyler. 2000. "How important are the cognitive skills of teenagers in predicting subsequent earnings?" *Journal of Policy Analysis and Management* 19 (4) pp. 547-568.

NMC Horizon Report: 2015 Higher Education Edition available at http://cdn.nmc.org/media/2015-nmc-horizon-report-HE-EN.pdf

OECD. (2015), Students, Computers and Learning: Making the Connection, PISA, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264239555-en

Psacharopoulos, George, and Harry A. Patrinos. 2004. "Returns to investment in education: a further update." *Education Economics* 12(2) pp. 111-134.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, schools, and academic achievement." *Econometrica* 73(2) pp.417-458.

Sosin Kim, Betty J. Blecha, Rajshree Agarwal, Robin L. Bartlett and Joseph I. Daniel 2004. Efficiency in the Use of Technology in Economic Education: Some Preliminary Results. *The American Economic Review* 94(2) Papers and Proceedings of the One Hundred Sixteenth Annual Meeting of the American Economic Association San Diego, CA, January 3-5, 2004 pp. 253-258.

Todd, P.E. and Wolpin, KI. 2003. On the specification and estimation of the production function for cognitive achievement, *Economic Journal*, 113, pp. F3-33.

Water, John K. 2014. The Great Adaptive Learning Experiment. *Campus Technology*, April 16 (available at https://campustechnology.com/articles/2014/04/16/the-great-adaptive-learning-experiment.aspx)

World Bank Report 2005. Expanding Opportunities and Building Competencies for Young People A New Agenda for Secondary Education. Washington, DC: World Bank. © World Bank. http://siteresources.worldbank.org/EDUCATION/Resources/278200-1099079877269/547664-1099079967208/Expanding_Opportunities_Secondary.pdf)

World Bank Report. 2009. Secondary Education in India: Universalizing Opportunities, January 2009, Human Development Unit, South Asia region. Washington, DC: World Bank. © World Bank. http://datatopics.worldbank.org/hnp/files/edstats/INDstu09b.pdf)

Verspoor, Adriaan M. 2008. At the Crossroads Choice for Secondary Education in Sub-Saharan Africa. Washington, DC: World Bank. © World Bank. https://openknowledge.worldbank.org/handle/10986/6537

Zimmer, T. 2014. Rethinking Higher Ed: A Case for Adaptive Learning. *Forbes*, October 22, 2014 (available at http://www.forbes.com/sites/ccap/2014/10/22/rethinking-higher-ed-a-case-for-adaptive-learning/#2715e4857a0bc68286462936)

APPENDIX

Appendix A: Mathematics Syllabus

7th grade topics- Ratio applications, Exponents, Data handling – mean, median, and mode, Properties of triangles, Number classifications (integers and real numbers), Area and perimeter of Two-dimensional and three-dimensional geometric figures, Solving algebraic expressions and equations.

8th grade topics- Radicals (square roots and cube roots), Laws of exponents and powers, Frequency distribution tables and graphs, Area of Plane Figures, Linear equation in one variable and two variables, Ratios, proportions and percentages, Solving algebraic expressions, Factorization, Surface area and volume of Two-dimensional and three-dimensional geometric figures.

Appendix B: Adaptive Algorithm

Details of the adaptive algorithm are as below

- First half of the total number of questions in a homework, called base questions, are divided among the topics covered in the homework in the same ratio as a corresponding PBTHW. Say if k_j questions are to be drawn from a topic j out of the total N base questions, such that $\sum_j k_j = N$. Then, each base question can be drawn from a multinomial distribution where the probability of drawing a question from a topic j is $p_j = k_j / N$, such that, $\sum_j p_j = 1$. The base questions are drawn in the easy category. If a student answered all easy base questions on a topic correctly, she is additionally offered one hard base question on that topic. For example, a homework has total 10 questions covering two topics T1 and T2. Then each of the 5 base questions is drawn from a multinomial distribution (in case of two topics it reduces to Bernoulli trial), say with equal probability ½ from T1 or T2 under the easy category. If a student answered all easy base questions on T1 correctly, she will be offered an additional hard base question on T1. The basic idea is to offer a minimum number of questions from each topic to discover student's weak areas before the adaptive generation of questions.
- Based on the performance of student in base questions, the remaining questions are adaptive generated. Thereafter, the remaining questions are drawn in ratio of the proportion of incorrect answers on base questions on a topic. Each adaptive question is drawn from a multinomial distribution where the probability of drawing from a topic is the ratio of proportion of incorrect answers from that topic to the sum of proportion of incorrect answers on all topic. For instance, if a student incorrectly answers 1 out of 3 questions (33%) from T1 and 2 out of 2 (100%) from T2,

each of the remaining 5 questions in the homework will be drawn with a probability of 33/133 from T1 and a probability of 100/133 from T2.

- The decision to serve easy or hard question in the adaptive phase is based on the student's performance on base questions. A correct easy question fetches +1 point, a correct hard question fetches +2 point, and each incorrect easy or hard question fetches -1 point. Based on this scoring criteria, a topic-wise running counter of scores is maintained. At the end of base questions, if the running score on a topic is greater than or equal to +2, a hard question on that topic is generated, otherwise an easy question is generated on that topic. During the adaptive question generation, the topic-wise running score is updated as per the student's performance in each adaptively generated question. The updated score on a topic determines the difficulty level of the next adaptively generated question on that topic. In the above example, the student's score in T1 was +1-1-1=-1 and in T2 was -1-1=-2 at the end of base questions. So one easy question will be generated from each of these topics, and based on the student's performance on these questions, the running score will be updated in each topic. Once the student's score goes beyond +2 on a topic, she will be offered a hard question on that topic.

Appendix C: Level of Personalization in CGAHW

The difference in number of easy questions in a CGAHW from that of its corresponding PBTHW on the same topics – denoted as *DfNQeasy* – captures offered personalization in both dimensions, topics and the level of difficulty of questions. We illustrate this with an example of homework that has a total 12 questions covering two topics, T1 and T2 in Table C. In a PBTHW, the questions are equally divided between these two topics such that, four easy and two hard questions are served from each topic. Thus, the student received a total of eight easy questions in PBTHW.

For CGAHW on the same topics, we consider three students. The first student understands neither T1 nor T2 from the class room instructions. The student is first offered three easy base questions from each topic and she incorrectly answers them all. Thereafter, the remaining six questions are equally divided between the two topics, such that she is offered three easy questions from each topic. In this case, the student receives personalization on the difficulty level of questions but not on topics. The *DfNQeasy* value captures the higher number of easy questions offered, albeit on all topics equally, in CGAHW.

The second student understands T1 well but not T2. She may correctly answer three easy base questions on T1 and incorrectly answer three easy questions on T2. The student will be offered an additional hard base question on T1, which she correctly answers. In the adaptive phase, the student will be offered the remaining five questions from T2, say four easy questions and one hard question. In this case, the student receives a higher number of questions in her weak area (8 questions from T2) and a smaller number of questions in her strong area (4 questions on T1). Moreover, the student received fewer easy questions on T1 and a greater number of easy questions on T2. The *DfNQeasy* value in this case captures the personalization in both dimensions, a greater (smaller) total number of questions in weak (strong) areas and a greater (smaller) number of easy questions in weak (strong) areas.

Table C: Levels of personalization in different cases

<table>
<tr><td rowspan="4"></td><td colspan="12">CGAHW</td></tr>
<tr><td colspan="4">Student understands neither T1 nor T2</td><td colspan="4">Student understands T1 but not T2</td><td colspan="4">Student understands both T1 and T2</td></tr>
<tr><td colspan="2">T1</td><td colspan="2">T2</td><td colspan="2">T1</td><td colspan="2">T2</td><td colspan="2">T1</td><td colspan="2">T2</td></tr>
<tr><td>E</td><td>H</td><td>E</td><td>H</td><td>E</td><td>H</td><td>E</td><td>H</td><td>E</td><td>H</td><td>E</td><td>H</td></tr>
<tr><td rowspan="2">Discovery phase</td><td>3</td><td>0</td><td>3</td><td>0</td><td>3</td><td>0</td><td>3</td><td>0</td><td>3</td><td>0</td><td>3</td><td>0</td></tr>
<tr><td colspan="4"></td><td>0</td><td>1</td><td colspan="2"></td><td>0</td><td>1</td><td>0</td><td>1</td></tr>
<tr><td rowspan="3">Adaptive phase</td><td>3</td><td>0</td><td>3</td><td>0</td><td colspan="2" rowspan="3"></td><td>2</td><td>0</td><td>0</td><td>2</td><td>0</td><td>2</td></tr>
<tr><td colspan="4" rowspan="2"></td><td>2</td><td>0</td><td colspan="4" rowspan="2"></td></tr>
<tr><td></td><td>1</td></tr>
<tr><td>Total Qs</td><td>6</td><td>0</td><td>6</td><td>0</td><td>3</td><td>1</td><td>7</td><td>1</td><td>3</td><td>3</td><td>3</td><td>3</td></tr>
<tr><td>DfNQeasy</td><td colspan="4">12 - 8 = 4</td><td colspan="4">10 - 8 = 2</td><td colspan="4">6 - 8 = -2</td></tr>
</table>

The third student understands both T1 and T2 from the class room instructions. The student is first offered three easy base questions from each topic and she correctly answers them all. She will be offered an additional hard base question from both T1 and T2, which she would correctly answer. Thereafter, in the adaptive phase, the student is offered the remaining four questions equally from the two topics, and we assume that the student answer them all correctly. In this case, the student receives personalization on the difficulty level of questions but not on topics. The negative *DfNQeasy* value captures the smaller number of easy questions offered, albeit on all topics equally, in CGAHW as per the student needs.