RIETI Discussion Paper Series 19-E-040

Does Computer-aided Instruction Improve Children's Cognitive and Non-cognitive Skills?: Evidence from Cambodia

ITO, Hirotake
Keio University

KASAI, Keiko
Keio Research Institute at SFC

NAKAMURO, Makiko
Keio University

RIETI
Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 19-E-040

May 2019

Does computer-aided instruction improve children's cognitive and non-cognitive skills?: Evidence from Cambodia[1]

Hirotake ITO

Graduate School of Media and Governance, Keio University

Keiko KASAI

Keio Research Institute at SFC

Makiko NAKAMURO

Faculty of Policy Management, Keio University

Abstract

This paper examines the causal effect of computer-aided instruction (CAI) on children's cognitive and non-cognitive skills. Closely working with the Cambodian government, we ran clustered-randomized controlled trials at five elementary schools near Phnom Penn for three months. Students were randomly assigned into 20 treatment classes which were allowed to use an app based on CAI instead of regular math classes during the intervention, or 20 control classes. Our empirical results drawn from these experiments suggest that average treatment effect on cognitive skills measured using several types of math achievement tests and IQ tests is positive and statistically significant. The effect is significantly large, especially as compared with prior literature examined in developing countries: the estimated coefficients on student achievements are 0.56-0.67 standard deviations and IQ scores 0.70 standard deviations even after controlling for demographic factors. Furthermore, it is found that that CAI-based app can increase the students' subjective expectation of being able to attend tertiary education in the future. However, there is no significant effect on non-cognitive skills, such as motivation.

Keywords: CAI, cluster-randomized controlled trial, non-cognitive skills
JEL classification: A20, I21

[1]This study is conducted as a part of the Project, "Research on the Improvement in Resource Allocation and Productivity among the Healthcare and Education Service Industries" undertaken at Research Institute of Economy, Trade and Industry (RIETI). This study is financially supported by Kakenhi Grant-in-Aid (ID:18H05314, Makiko Nakamuro). The authors is grateful for helpful comments and suggestions by Discussion Paper seminar participants at RIETI.

1 Introduction

The World Bank stated in its flagship report about ”learning crisis” (World Bank, 2017): a large fraction of students in developing countries missed the opportunity to acquire even foundational skills at school, e.g., how numbers work, which is required when buying and selling goods in market, handling household budgets, or doing transactions with back and other financial institutions, etc.

While lower-income countries are rapidly expanding the primary enrollments for past decades, many of them face substantial obstacles to keep away from this ”learning crisis”. First of all, the expansion of primary enrollments has kept higher pace than the expansion of school inputs, such as teachers and other school resources. It is concerned that a decline in per-capita input available leads to low quality of primary education. Secondly, hiring high-quality of teachers has been difficult in many countries because teachers are paid less than other comparably qualified professionals, particularly in urban areas. Thirdly, the huge gap between high- and low-achieving students brings teachers into more difficult situation to adjust their instruction in response to class composition.

Advanced technologies are able to offer promising ways to help with these situation in developing countries: although just simply installing computers to classrooms does not change students' learning, as shown in Barrera-Osorio and Linden (2009), well designed computer-assisted learning (CAL) may help students to access high-quality of instructions and materials even under the severe teacher shortage and learn at their own pace. However, despite the great expectation, the evidence on the effect of CAL is yet mixed. A computer-assisted learning in India was confirmed to vastly improve student performance, especially for the initially lowest achieving students (Linden, 2008), while ”One Laptop Per Child programs” in Peru and Uruguay had no impact on student reading or math achievements (Cristia et al., 2012; De Melo, Machado and Miranda, 2014).

This research is designed to rigorously estimate the causal impact of the CAI on students' cognitive and non-cognitive skills, closely working with the government of the Kingdom of Cambodia, Japan International Cooperation Agency (JICA), and Hanamaru-lab, one of the Japanese private companies, which provides the newly developed CAI apps, called ”Think!Think!”. The primary objective of ”Think!Think!” is to foster the foundational math skills for elementary school students.

To examine the effect of ”Think!Think!”, we ran clustered randomized-controlled trial and intervened approximately 1,500 students from G1 thorough G4 at five public elementary schools near Phnom Penh during May though August, 2018. Because each school has two classes in each grade, students are randomly assigned either into 20 treatment classes which are using ”Think!Think!” instead of taking regular math class during the three-months intervention, or 20 control classes.

Our empirical results drawn from this experiments suggest that average treatment effects on cognitive skills measured by several types of math achievement tests and IQ tests are positive and statistically significant. The effect size is significantly large, especially as compared with prior literature examined in developing countries: the preferred point estimates on student achievements are 0.56-0.67 standard deviations and IQ scores 0.70 standard deviations even after controlling for prior score at the baseline-survey, gender, grade, birth-of-month, parental education and schools' time-invariant characteristics. Furthermore, it is found that that CAI-based app can raise the students' subjective expectation to go to college in the future. However, there is no significant effect on non-cognitive skills, such as motivation and self-esteem.

The remainder of this paper proceeds as follows: Section I provides the literature review. Section II explained the research design, data and empirical specifications. Section III presents the main results on cognitive and non-cognitive skills. Section IV provides heterogenous effects in terms of students' characteristics. Section V concludes and discusses policy implications.

2 Literature Review

In prior literature, investments in computers at school is defined as either (i) Information Communication technology (ICT) or (ii) Computer-Aided Instruction (hereafter, CAI). CAI programs, which does not necessarily require an Internet connection, is becoming more broadly accepted in public schools for recent years. However, while several studies showed that well-designed CAI programs appeared to have strong and positive effects on lower grade students' math or science achievements, especially in developing countries, other studies found the insignificant effects on reading and language test scores. For example, Rouse and Krueger (2004) ran a large scale Randomized Controlled Trial (hereafter, RCT) with the computer software program, called ”Fast For Word”, for third grade to sixth grade students in the urban district of northwest United States. Their results showed that the effect of this program on the language acquisition and reading skills are small and statistically significant. On the other hand, Banerjee et al. (2007) presented the effect of the CAI program for fourth grade students in urban

India. Students who were randomly assigned into treatment schools increased their math achievements by 0.47 standard deviation, mostly due to the achievement gains from initially lower performing children at the distribution. Surprisingly, this positive effects remained even after the programs was terminated, although the magnitude has become smaller to about 0.10 standard deviation.

In the field of economics, investments in computers, the Internet, software and other technologies is placed in the context of education production function. As Bulman and Fairlie (2016) pointed out that the binding constraint in the model may be the amount of time available for instruction, which is regarded as one of the educational inputs. In other words, this trade-off between time spent using the computer in class and time spent for traditional instruction makes more difficult to answer the question of whether schools should choose the CAI programs relative to traditional instructions. However, literature, such as Rouse and Krueger (2004) and Banerjee et al. (2007) estimated the effect of supplemental education or remedial education with CAI programs outside of class.

To deal with this problem, Barrow, Markman and Rouse (2009) set up that middle school students in randomly assigned treatment classes are taught using CIA, while students in control classes were taught traditionally in class. This enables authors to compare the newly developed CAI program and traditional instructions under the limited resource and time constraints at school. Two-year experiment showed that treatment students significantly improved their math achievements at least 0.17 standard deviation than their counterparts. Carrillo, Onofa and Ponce (2011) conducted the similar experiment with Barrow, Markman and Rouse (2009) in Ecuador for elementary school students. Using CAI in class, instead of traditional instruction, helped to improve math performance, but not language acquisitions. Although many research suggested that computers can offer highly individualized instruction, allowing students to learn at their own pace, the evidence from more recent economic studies is mixed, depending on the subjects, grade, program contents, and the extent to which CAI is a substitute or a supplement to traditional instruction. Interestingly, the positive effects appears to be stronger in developing countries.

3 Methodology and Data

3.1 Background

Our study targets five public elementary schools around 10 km radius near Phnom Penh. Because these schools did not receive any aids or assistance from other development agencies during the period of our intervention, it is certainly ruled out the confounding factors from other external interventions. A majority of households around the schools are engaged in farming and fishing as income generating activities. Only small fraction of parents completed tertiary education. The location of these five schools are illustrated the below.

Figure1.

3.2 Baseline Survey

Prior to the intervention, we conducted the baseline survey in class during May 21 to May 25, 2018, with a full cooperation of teachers and staffs. The baseline survey included two sets of 40-minute achievement tests for G3 and G4 students respectively, 40-minute IQ tests for all students and approximately 20-minute surveys both for all students and parents.

In order to measure students' cognitive skills, the two sets of achievement tests were derived from the past exam of National Assessment Test (NAT) originally administered by Cambodia's Ministry of Education, Youth and Sports for G3 students and Trends in International Mathematics and Science Study (TIMSS) originally administered by The International Association for the Evaluation of Educational Achievement (IEA) for G4 students. Unfortunately, because there are any accessible standardized tests to measure the cognitive skills for younger students, we did not administer the achievement tests for G1 and G2 students.

Instead, we administered two sets of age-appropriate IQ tests implemented at the baseline survey. One of the IQ test is called "new Tanaka-B type intelligence test" (Tanaka, Okamoto and Tanaka (2003)), which has long been widely used in Japan and Asian countries to measure young children's cognitive skills. G3 and G4 students took more advanced version of test, which is slightly different from the test that G1 and G2 students took. Both of these Tanaka-B type intelligence tests were not only translated into the local language but also modified to appropriately describe the local situations (e.g.,illustrations of banknote, food, and people, etc).

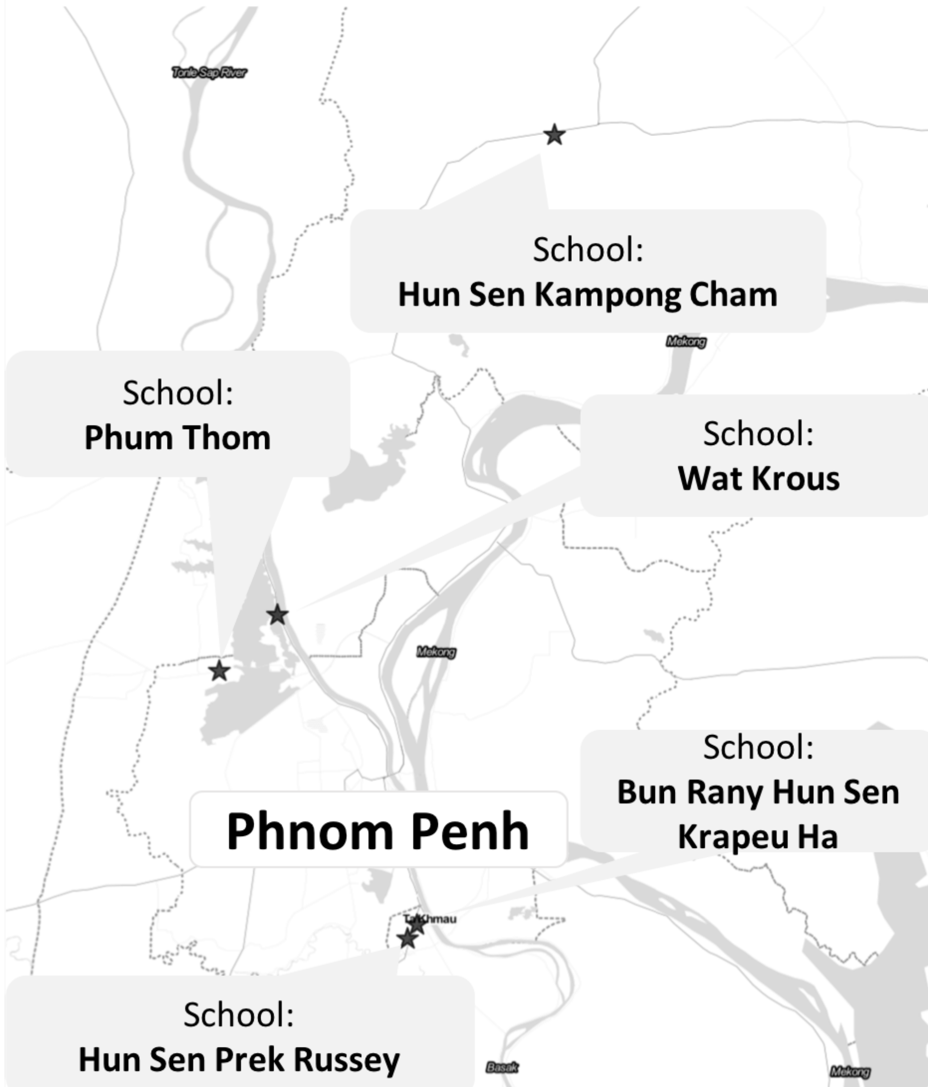

Figure 1: intervention schools

The other intelligence test is Goodenough Draw-a-Man (DAM) test (Goodenough, 1926). In this test, students were asked to complete drawings of a whole person(s) on a piece of paper for 10 to 15 minutes. The several examples of children's drawings collected at our baseline survey in Figure 4. Although this test has been criticized on the validity as a measurement of intelligence, more recent research suggested that this DAM scale is still effective to screen the lower range of intelligence for 5-12 years old children(Scott, 1981).

The survey for all G1-G4 students asked them to answer their demographic information, such as gender, grade, birth-of-month, hours of studying at home and subjective likelihood to go to college in the future, etc. Furthermore, the survey included a set of questionnaires to measure the non-cognitive skills, such as Rosenberg self-esteem scale (Rosenberg, 1965) and internal and external motivation for study (Sakurai and Takano, 1985). The survey for all parents asked to answer their socio-economic status, such as parents' educational backgrounds.

After three-month intervention, the end-line survey was then conducted between Aug 16 to Aug 25. We again administered the same sets of achievement tests, IQ tests, the questionnaires for students, especially focusing only on time-varying variables, such as the willingness to go to college and time spent studying at home.

3.3 Education app: "Think!Think"

The app called "Think!Think!" that we used in our intervention is originally developed by Hanamaru-lab. Ltd, taking an full advantage of its long-time experiences accumulated through running a great numbers of cramming schools for school-aged children nationwide in Japan. The objective of this app is to foster the foundational math skills for elementary school students. More specifically, this app incorporates an adaptive learning with its own original computer algorithm and automatically provides a large numbers of math problems, materials, and instructions in response to the proficiency level of each individual students at their own pace.

To be more comfortably used by elementary school students in Cambodia, " Think!Think! " was slightly modified to meet Cambodian curriculum standards and translated into the local language, Khmer. The tablet-PCs were allocated individually, while students were involved in "Think!Think!" in class. Moreover, CAI does not often require the additional teaching staff to manage a class. In our intervention, we placed three staffs who did not have any teaching experiences prior to the intervention to helped students on the technical matter and time management.

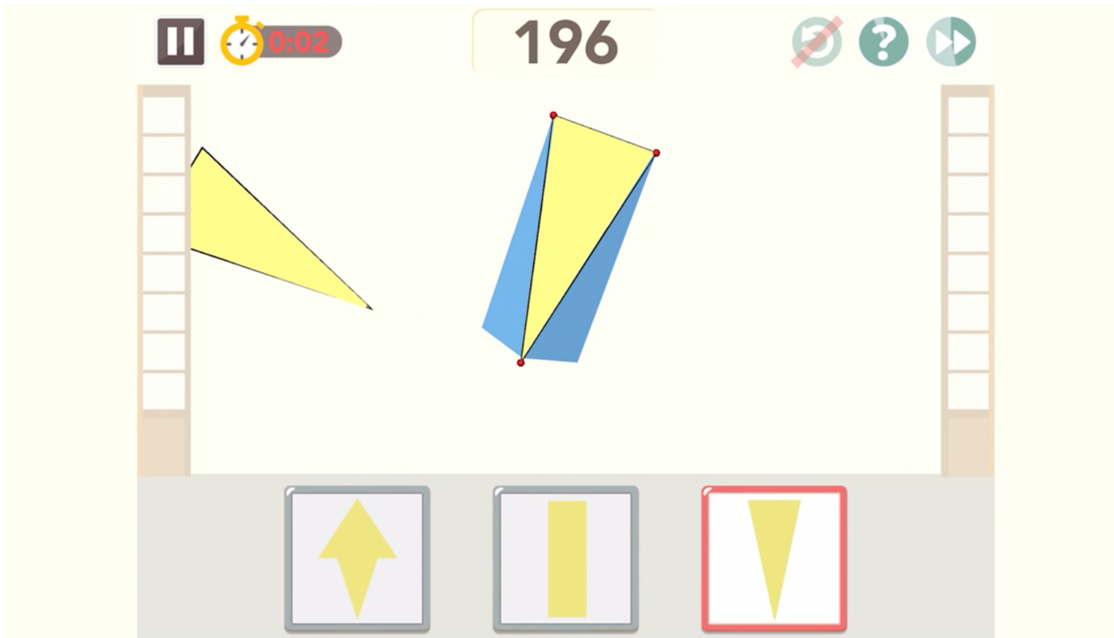

Figure 2: Sample problem

3.4 Cluster-Randomized Controlled Trial

Nevertheless to say, if we allow students to access CAI-based app, ”Think!Think!”, based on their own preference, students who often use the app probably perform better to begin with. Students who have sought to access higher quality of education, including the exposure to new technology, are not as enthusiastic to study, on average, as those who never did. Random assignment of the access to CAI-based app solves this selection bias.

In our experiments, approximately 1,500 from G1 though G4 students at five public elementary schools are supposedly assigned either into treatment group who are using ”Think!Think! instead of regular math class, or control group. Students in treatment classes used ”Think!Think” for approximately 30 minutes everyday. Peer effect may be one of the major threat to the internal validity of this experiment and the interactions among students may violate the stable unit treatment value assumption (SUTVA). To avoid this situation, besides the fact that cluster-randomized controlled trial (hereafter, cluster-RCT) is more common in education as suggested in literature, we randomized students within intact classrooms, rather than individual students within them.[1]

Because each school has two classes at each grade, We thus randomly assigned one of the classes at each grade and each school into treatment, or control otherwise (See Figure 3 for the cluster-randomization). This setting made us to create 20 treatment classes (with 807 students) and 20 control classes (with 829 students) across five schools. However, there is still a potential concern that students in treatment classes would talk to their friends in other control classes at the same school about what they have learned. To reduce the risk of spillovers, we are not allowed treatment students to access "Think!Think!" out of classes. They are neither allowed to bring the tablet PCs to their own place of residence.

The intervention is implemented during the period between May 25 and Aug 15. Despite the short-run intervention, it seems that students are very organized and enthusiastic to do "Think"Think" for this three months.

4 Econometric specification and Results

4.1 Econometric specification

To identify the causal effect of using "Think!Think!", we run OLS regression of the following model generates an unbiased estimate of the intent-to-treat effect and the effect of randomly assigning students to be taught using CAI on various outcomes in question. Our equation of interest is:

$$Y_{i,j,t} = \alpha + \beta T_{i,j,t} + \gamma Y_{i,j,t-1} + X_{i,j,t}\delta + \varepsilon_{i,j,t}$$

where $Y_{i,j,t}$ is outcome variables of individual student i, school j at time t. $T_{i,j,t}$ in equation (1) is the exogenous variation in the access to CAI and the key independent variable of interest. $X_{i,j,t}$ is the vector of controls, while $\varepsilon_{i,j,t}$ is the idiosyncratic error term.

The crucial identifying assumption in this empirical model is that the relationship between the exposure to CAI-based app and students' unobserved ability is orthogonal to the error term, conditional on the controls. Under this assumption, the estimate of β in equation (1) can be interpretable as the causal impact of the introduction to CAI-based app on outcomes.

[1] However, as pointed out by (Imbens, and Wooldridge, 2009), it is technically difficult to separate out the direct effect of the intervention on individual from the indirect effect of the peer on the individual.

4.2 Variables Defined

As shown in Table1, the outcome variables of interest denoted by $Y_{i,j,t}$ is defined as follows: both achievement tests are converted to T-scores with mean 50 and standard deviation 10. Table1 shows a balance check performed for the baseline survey. It is found that there is no statistically significant difference in NAT between G3 students assigned into treatment and control classes, although G4 students in treatment classes performed much better in TIMSS than ones in control classes.

The other outcome variables are defined as IQ test scores: The results of Tanaka-B type IQ test and Draw-a-Man test are converted into Mental Age (measured as level of intellectual performance, MA) and are then defined IQ scores as MA over chronological age (biological age, CA) multiplied by 100. According to the descriptive statistics, a mean of Tanaka-B type IQ score is 78.612 with 13.451 standard deviation and a mean of DAM type IQ score is 0.692 with 0.207. While there doesn't exist statistically significant difference in Tanaka-B type IQ score but is in DAM type score.

As other set of outcome variables, the measures of non-cognitive skills are coded as the mean of a set of questionnaires specific to self-esteem and motivation. The self-esteem scale is slightly higher for treatment students, while motivation scale is balanced between two-group students. The willingness to go to college is measured on a 3-point scale (1=not likely through 3 very likely) based on student subjective expectations. The original questionnaire of the hours spent studying at home is the response category, ranging from one (=not at all) through six (=more than four hours). We set the minimum to zero and maximum to four and then took the median value for categories between two (=less than 30 minutes) and five (=two to three hours).

The key independent variable of interest denoted by $T_{i,j,t}$ is a dummy variable coded as one if students are assigned into treatment class, zero otherwise.

As controls denoted by $x_{i,j,t}$, the demographic variables, such as gender, age, and parental educational backgrounds, are substantially equal between treatment and control students. The variable on parental education represents the highest level of education of either one of the parents. Note that those information is retrieved from the parental survey conducted at the same time with student survey. However, unlike the student survey administered at schools during the class, because the response rate of parental survey is approximately 85 percent.

Even though the observable characteristics are balanced between two groups, it is the fact that the several outcome variables,such as the achievement score for

G4 students, DAM type IQ score, and self-esteem scale, are not substantially comparable at the baseline survey. Because the lack of balance can occur by chance even when randomization is carried out correctly and the chance of achieving balance when we randomize at the group level is increasing as the sample size increases, we may not worry too much if we see the imbalance in four out of 15 variables. However, even though schools randomized the change in class composition every year, it may still happen to make imbalance between treatment and control groups due to drop-out or absence at the day of the baseline survey. We thus control for the demographic variables we use for the balance check in order to make as much as possible "pure" comparison.

It may be possible that the average treatment effects depends on particular students' sub-groups of interest. For example, if boys are more familiar with the computer related equipment, the effect may be stronger for boys than girls. This kind of heterogeneous effect must be very important for policy makers in considering how to tailor the needs of particular subgroups. We will discuss this point in section 4.

4.3 Results

Effect on cognitive skill

Our analysis begins with the estimated effect of CAI on student achievements. The coefficients estimated by OLS are reported in Table 2 with the the heteroskedasticity-robust standard errors and clustering at the school level. Our primary interest is presented in the first raw in Table, the estimated coefficient of the access to "Think!Think" on NAT for G3 and TIMSS for G4.

Basically, Model I contols only for the prior achievement score at the baseline survey, and Model II controls basic demographic controls, such as gender, grade, birth-of-month, parental education and school time-invariant fixed effects, in addition to the prior test score.

The results clearly show that the estimated coefficients on standardized test scores has a positive and statistically significant at a 0.1 percent level (Table 2). The estimated coefficients for the sample of G3 students indicate the exogenous exposure to CAI app raises average test scores by about 0.67 if expressed in units of standard deviations by grade.

Adding demographic controls in Model II neither change the magnitude of coefficients across specifications nor improve the precision of our estimates by explaining the variation in test scores. Once We includes the interaction term

	ALL	Treatment (A)	Control (B)	Difference (A)-(B)
Achievement Test (NAT, G3)	50.000 (10.000)	49.227 (9.573)	50.765 (10.375)	1.538
Achievement Test (TIMSS, G4)	50.000 (10.000)	51.921 (9.213)	48.068 (10.407)	-3.853***
IQ Test (Tanaka-B)	78.612 (13.451)	78.432 (13.131)	78.795 (13.777)	0.363
IQ Test (Draw-a-man)	0.692 (0.207)	0.678 (0.206)	0.705 (0.207)	0.027**
Self-esteem	2.762 (0.549)	2.726 (0.596)	2.794 (0.502)	0.068**
Motivation	0.656 (0.142)	0.652 (0.150)	0.660 (0.133)	0.008
Willingness to go to college	2.410 (0.771)	2.342 (0.809)	2.467 (0.734)	0.125***
Hours of studying at home	114.858 (121.313)	112.384 (115.812)	117.056 (126.032)	4.672
Gender (male=1, woman=0)	0.525 (0.500)	0.530 (0.499)	0.519 (0.500)	-0.011
Age	8.485 (1.553)	8.501 (1.573)	8.470 (1.535)	-0.031
Highest parental education				
College or Graduate school	0.017 (0.129)	0.012 (0.110)	0.021 (0.145)	0.009
High school	0.340 (0.474)	0.353 (0.478)	0.328 (0.470)	-0.026
Junior high school	0.222 (0.416)	0.218 (0.413)	0.227 (0.419)	0.009
Elementary school	0.164 (0.370)	0.160 (0.367)	0.167 (0.374)	0.007
no education(ref)	0.001 (0.035)	0.002 (0.049)	0.000 (0.000)	-0.002
Birth of Month				
Jan. - Mar.	0.228 (0.420)	0.219 (0.414)	0.238 (0.426)	0.019
Apr.-Jun.	0.240 (0.427)	0.258 (0.438)	0.223 (0.417)	-0.035*
Jur.-Sep.	0.243 (0.429)	0.251 (0.434)	0.236 (0.425)	-0.014
Oct.-Dec.	0.264 (0.441)	0.254 (0.436)	0.273 (0.446)	0.019

Note: ***, **, and * represent 0.1%, 1%, and5% significance level, respectively.

Table 1: Descriptive statistics and balance test

and tests whether to exist the heterogenous effects in terms of gender, grade, and parental education, we find small point estimates on nearly all interaction terms, and difference among these coefficients do not support the hypothesis of significant heterogenous effects on test scores. The results will be provided upon request.

The results drawn from our empirical evidence is consistent with expectations for the sample of G4 sample (Table 2): The access to CAI app improves standardized test score by 0.56 standard deviations per three-month exposure. Adding controls tends to increase the point estimates and decrease the standard errors of these estimates. By the same token, we do not find any significant and heterogenous effects on test scores in terms of gender, grade and parental education.

Looking at the results in Table 3, the estimated coefficient on the Tanaka-B type IQ score is positive and statistically significant at 0.1 percent level. We calculate the effect size on the IQ score from Model I of Table 2 and obtain a coefficient of 0.70 standard deviation (=9.415/13.451). The estimated coefficient remains constant after controlling for demographic characteristics in Model II. However, the coefficient of the DAM score, regardless of specifications, are not statistically significant. Taken as a whole, the magnitude in cognitive skills appears to be very large, as compared with evidence suggested by previous literature that intervened for at least one year.

By using kernel density estimation, we obtain the probability density function both of achievement test scores and IQ scores to compare in score distributions between before and after the three-month intervention (Figure 3, 4 and 5, respectively). Although the difference in the DAM score for the entire sample, and even interaction term with grade, are not statistically significant, younger students seemed to be slightly improved.

Effect on non-cognitive skills

We then repeated the above protocol with a set of non-cognitive skills as outcomes. Unlike the results of cognitive skills, we do not find any significant effect for non-cognitive skills, measured by motivation and self-esteem (Table 4). However, it is clear that the estimated coefficient on willingness to go to school is positive and statistically significant at a 0.1 percent level, indicating students who used CAI app during the class are more likely to believe that they would take more advanced education in the future. The coefficient remains constant after controlling for demographic characteristics in Model II and suggests that the heterogeneous effects in terms of gender, grade and parental education does not exist. Although the estimated coefficient do not support the hypothesis of a positive effect of CAI-based app on non-cognitive skills, the estimated probability density function suggest the sign of slight change in younger grade.

Dependent Variable	National Asssessment		TIMSS	
	Model1	Model2	Model1	Model2
Treatment	6.771*** (0.579)	6.791*** (0.633)	5.642*** (1.029)	5.847*** (1.102)
Baseline Score	✓	✓	✓	✓
Control		✓		✓
Observations	331	285	314	278
Adjusted R2	0.718	0.702	0.215	0.212

Note: ***, **, and * represent 0.1%, 1%, and 5% significance level, respectively.

Table 2: Effect of treatment: cognitive skills

Dependent Variable	IQ		Draw A Man	
	Model1	Model2	Model1	Model2
Treatment	9.415*** (0.715)	9.186*** (0.753)	0.003 (0.008)	-0.004 (0.009)
Baseline Score	✓	✓	✓	✓
Control		✓		✓
Observations	1226	1042	1060	898
Adjusted R2	0.461	0.489	0.255	0.288

Note: ***, **, and * represent 0.1%, 1%, and 5% significance level, respectively.

Table 3: Effect of treatment: IQ

5 Conclusion

This paper examines the causal effect of computer-aided instruction (CAI) on children's cognitive and non-cognitive skills. Closely working with the Cambodian government, we ran clustered-randomized controlled trial at five elementary schools near Phnom Penh for three months. Students were randomly assigned into 20 treatment classes which were allowed to use the app based on CAI instead

Dependent Variable	Motivation		Self Esteem	
	Model1	Model2	Model1	Model2
Treatment	-0.0001 (0.008)	0.0001 (0.008)	0.008 (0.026)	0.005 (0.028)
Baseline Score	✓	✓	✓	✓
Control		✓		✓
Observations	876	743	1016	866
Adjusted R2	0.297	0.361	0.057	0.121

Note: ***, **, and * represent 0.1%, 1%, and 5% significance level, respectively.

Table 4: Effect of treatment: non-cognitive skills

Dependent Variable	Study Time(minutes)		Willigness to go to college	
	Model1	Model2	Model1	Model2
Treatment	-11.982 (7.366)	-12.860 (7.946)	0.135*** (0.049)	0.133** (0.053)
Baseline Score	✓	✓	✓	✓
Control		✓		✓
Observations	1247	1031	920	790
Adjusted R2	0.021	0.048	0.040	0.034

Note: ***, **, and * represent 0.1%, 1%, and 5% significance level, respectively.

Table 5: Effect of treatment: study input

of regular math class during the intervention, or 20 control classes. Our empirical results drawn from this experiments suggest that average treatment effect on cognitive skills measured by several types of math achievement tests and IQ tests is positive and statistically significant. The effect size is significantly large, especially as compared with prior literature examined in developing countries: the estimated coefficients on student achievements are 0.56-0.67 standard deviations and IQ scores 0.70 standard deviations even after controlling for demographic factors. Furthermore, it is found that that CAI-based app can raise the students' subjective expectation to go to college in the future. However, there is no significant effect on non-cognitive skills, such as motivation and self-esteem. However, this clustered-randomized controlled trial have hold only 3 months, so whether the effect stays or not is a very important and interesting point. Although, we have to pay attention about Hawthorn effect and John-Henry effect and other possible effects carefully. Due to the sort-term experiment, we should investigate that the app "Think!Think!" has as tremendous ability to improve students' scores on math in the short-term as for a long term.

References

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

Barrera-Osorio, Felipe, and Leigh L Linden. 2009. "The use and misuse of computers in education: Evidence from a randomized controlled trial of a language arts program." *Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (JPAL). www. leighlinden. com/Barrera-Linden*, 20.

Barrow, Lisa, Lisa Markman, and Cecilia Elena Rouse. 2009. "Technology's edge: The educational benefits of computer-aided instruction." *American Economic Journal: Economic Policy*, 1(1): 52–74.

Bulman, George, and Robert W Fairlie. 2016. "Technology and education: Computers, software, and the internet." In *Handbook of the Economics of Education*. Vol. 5, 239–280. Elsevier.

Carrillo, Paul E, Mercedes Onofa, and Juan Ponce. 2011. "Information technology and student achievement: Evidence from a randomized experiment in Ecuador."

Cristia, Julian, Pablo Ibarrarán, Santiago Cueto, Ana Santiago, and Eugenio Severín. 2012. "Technology and child development: Evidence from the one laptop per child program."

De Melo, Gioia, Alina Machado, and Alfonso Miranda. 2014. "The impact of a one laptop per child program on learning: Evidence from Uruguay."

Goodenough, Florence Laura. 1926. "Measurement of intelligence by drawings."

Imbens, Guido W., , and Jeffrey M. Wooldridge. 2009. "Recent developments in the econometrics of program evaluation." 47: 5–86.

Linden, Leigh L. 2008. *Complement or substitute?: The effect of technology on student achievement in India.* InfoDev Working Paper, Columbia University.

Rosenberg, Morris. 1965. *Society and the adolescent self-image.* Princeton university press.

Rouse, Cecilia Elena, and Alan B Krueger. 2004. "Putting computerized instruction to the test: a randomized evaluation of a "scientifically based" reading program." *Economics of Education Review*, 23(4): 323–338.

Sakurai, Shigeo, and Seijun Takano. 1985. "A new self-report scale of intrinsic versus extrinsic motivation toward: learning for children." *Tsukuba psychological research*, 7: 43–54.

Scott, Linda. H. 1981. "Measuring intelligence with the Goodenough-Harris drawing test." 89: 483.

Tanaka, Kanichi, Kenroku Okamoto, and Hidehiko Tanaka. 2003. *The New Tanaka B Intelligence Scale.* Kaneko shobo.

World Bank. 2017. *World Development Report 2018: Learning to Realize Education's Promise.* The World Bank.

A Effect of Treatment: Estimated PDF function

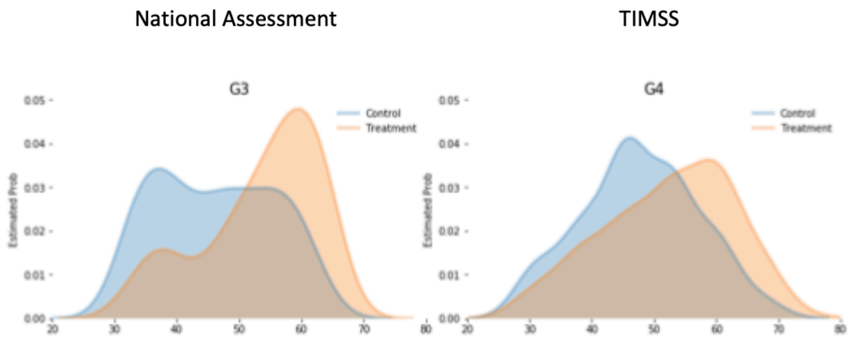

Figure 3: National assessment score and TIMSS

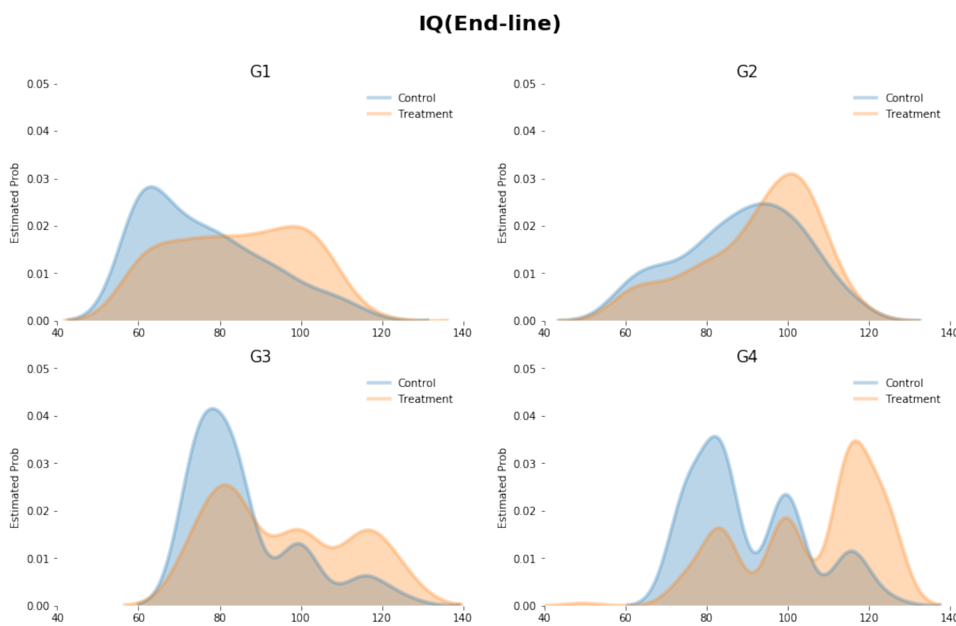

Figure 4: IQ

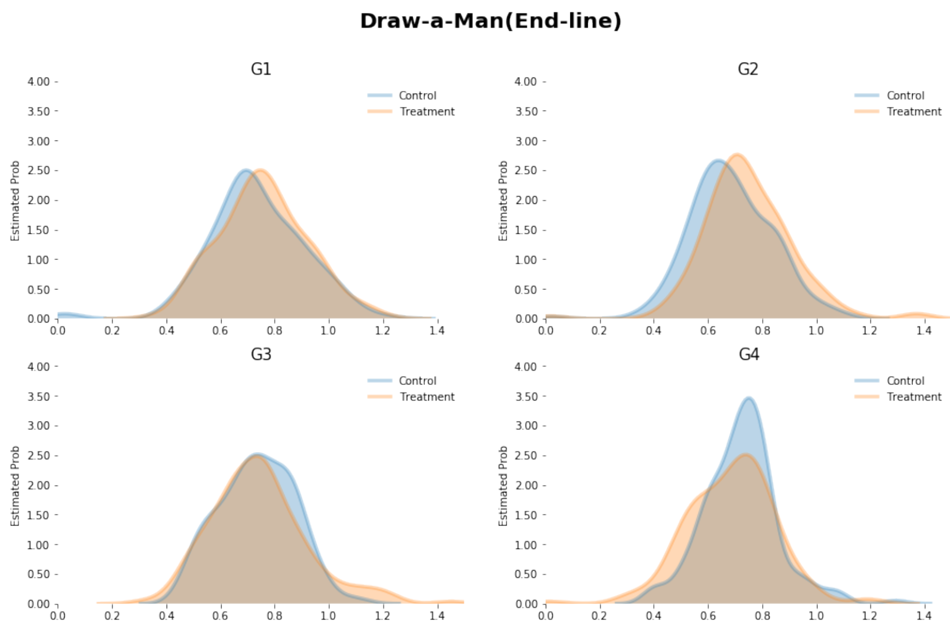

Figure 5: Draw a man test

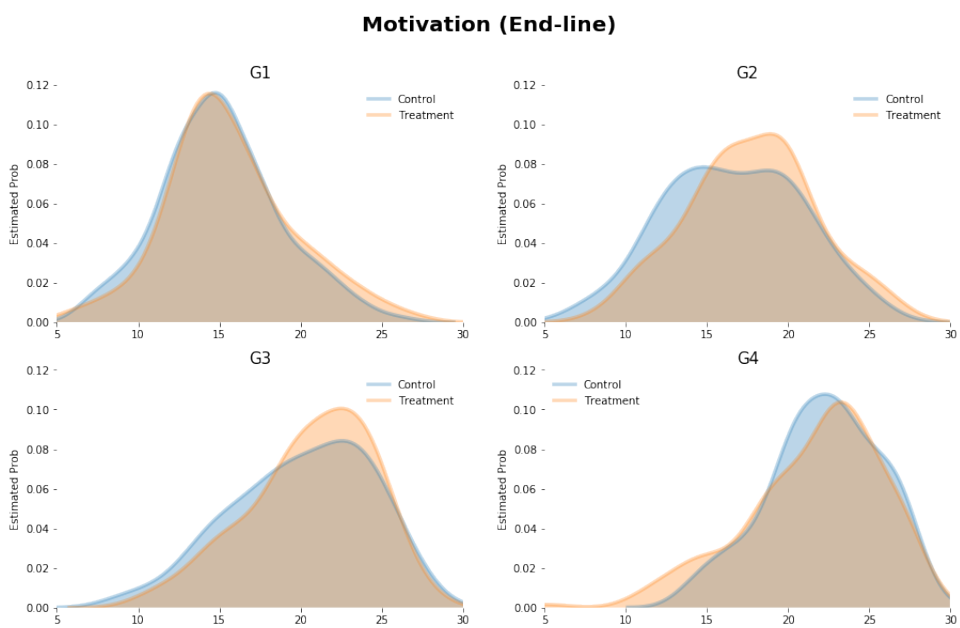

Figure 6: Motivation

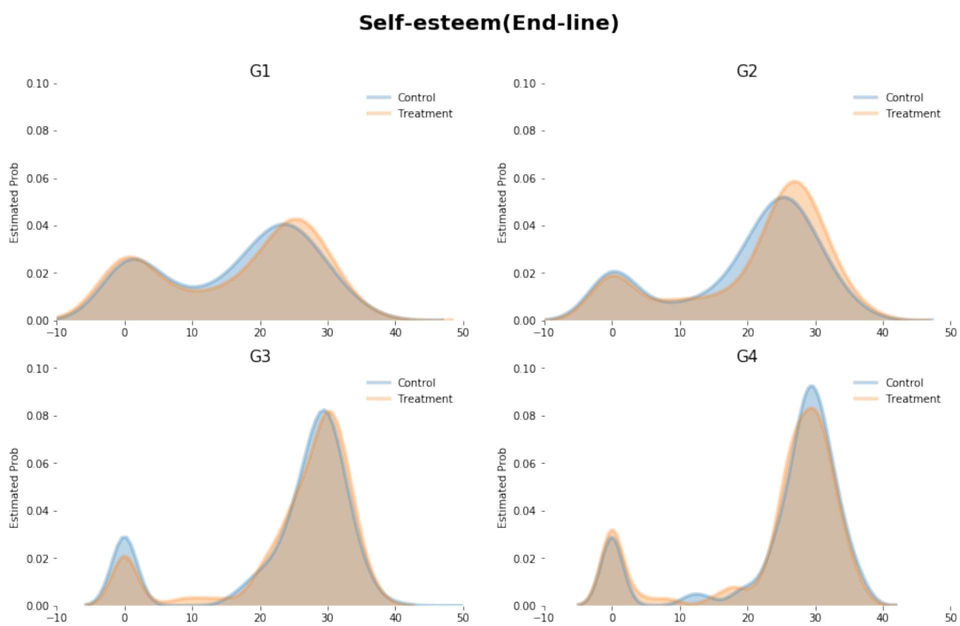

Figure 7: Self-Esteem