Xiaofan Liang
Replication and Extension of Cram School Effect paper
Minerva Schools at KGI

OVERVIEW

Cram school is a unique cultural phenomenon in Asia which has appeared out of a growing competition about spots in good schools . Students attend cram school in their free time to either study ahead or catch up with classes. Liu (2012) studied the effect of cram school in Taiwan and the question of who go to cram school in his paper *Does Cram School matter? Who goes to Cram School? Evidence from Taiwan*. In this paper, I confirmed most of professor Liu's findings through my own replications, extended the investigation by testing the effect of cram school more rigorously through matching and furthered the discussion of who (whose children) goes to cram school by looking at the interactions between parents' education and economic status.

CRAM SCHOOL AND EDUCATION POLICY

The existence of cram school has exacerbated the vicious cycle of competition among children from a low age. Policymakers are challenged to decide on how to regulate cram school. On one hand, they need to know the effect of going to cram school on students' academic performance compared with those who didn't, especially when considered along with family income level and education level. Considered 85% of the population, according to the sample data, went to cram school, the demography of people implies something more social than just the desires to improve academic learning. Therefore, it's important to investigate both questions to help policymakers understand this phenomena and create alternative solutions.

LIU'S DISCUSSION AND MY REPLICATION

In his paper, Liu ran an overarching multivariate linear regression model to find the effect of different variables on academic performance, with regard to the first variable in each clusters

(see figure 1). Cram School attendance has shown a significant positive effect on academic performance. Other factors such as family income and education display a bell shape effect (small on tails, big on middle income and education level) on academic performance. Logistic model of cram school participation also shows similar a pattern of bell shape, implying that people with middle income or medium education level are more likely to send their kids to cram schools, relative to people with low income and low education levels.

Table 3
Multivariate regression analysis on performance.

	Model 1: general score		Model 2: math score	
	b	Beta	*b*	Beta
Constant	−0.804*		−0.771*	
Cram school attendance				
Hours/week	0.067*	0.290	0.071*	0.319
Hours squared	−0.004*	−0.184	−0.004*	−0.197
Family monthly income				
Less than 20,000	–		–	
20,000–50,000	−0.075*	−0.484	−0.073*	−0.492
50–100,000	0.022	0.140	0.019	0.129
100–150,000	0.058*	0.377	0.042	0.284
150–200,000	0.002	0.014	0.019	0.13
200,000+	−0.007	−0.047	−0.008	−0.051
Parents' education				
Middle school	–		–	
High school	0.220*	0.116	0.198*	0.109
Vocational College	0.502*	0.201	0.432*	0.181
University	0.723*	0.232	0.608*	0.204
Masters and PhD	0.846*	0.165	0.684*	0.139

Table 4
Results for bivariate logistic models of cram school participation.

	Attending/not attending		
	All students	Boy students	Girl students
Constant	−0.075	−0.281	−0.059
Family monthly income			
Less than 20,000 NT	–	–	–
20,001–50,000	−0.069	−0.087	−0.045
50,001–100,000	0.172*	0.164*	0.181*
100,001–150,000	0.233*	0.219*	0.256
150,001–200,000	−0.117	−0.103	−0.127
200,000+	−0.218	−0.190	−0.266
Parents' education			
Middle school	–	–	–
High school	0.217*	0.207*	0.226*
Vocational college	0.496*	0.468*	0.527*
University	0.476*	0.393*	0.576*
Masters and PhD	0.101	0.090	0.115
Father's occupation			
Unemployed	–	–	–
Teacher	0.640*	0.782*	0.499
Public worker	0.550*	0.666*	0.435

(Figure 1)

I decided to limit the variables used in my replication to "cram school participation", "cram school hours", "income" and "education" as they involved in the most important findings of Liu's paper. Since I used a smaller set of variables for the multivariate and logistic regression models and may have used different ways to clean the data, the regression coefficients I got for each variable are different from Liu's (see figure 3). Nonetheless, the general pattern remains the same: when other variables are fixed, attending cram school has a significant positive effect on academic performance which is increased by 2.99 units, relative to not attending cram school. However, my logistic model shows something slightly different from Liu's. The likelihood of

people sending their kids to cram school increases with income level while seem to decrease when education level grows, but the coefficient values for high education levels are not statistically significant.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	33.4027836	0.38527080	86.699495	0.000000e+00
cramschoolYN	2.9901153	0.25056731	11.933382	1.198282e-32
cramhours	-0.2102531	0.01747744	-12.029967	3.787311e-33
income_nt$20,000 - nt$49,999	2.8458286	0.31894361	8.922670	5.188488e-19
income_nt$50,000 - nt$99,999	4.8857346	0.33598087	14.541705	1.626863e-47
income_nt$100,000 - nt$149,999	6.1348224	0.43308832	14.165292	3.391359e-45
income_nt$150,000 - nt$199,999	6.3543410	0.60978030	10.420706	2.539832e-25
income_nt$200,000 or more	5.9829867	0.66864966	8.947865	4.137259e-19
edu_senior (vocational) high school	3.1882242	0.20839763	15.298755	2.351104e-52
edu_junior/science and technological college	6.4258316	0.29739603	21.606985	1.162727e-101
edu_university	9.3940223	0.37638766	24.958369	3.542467e-134
edu_graduate school	9.8372246	0.67545469	14.563856	1.183229e-47
edu_other	-1.6042857	1.09182498	-1.469362	1.417602e-01

(Figure 2: My Multivariate Model Regression Coefficients)

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.82844	0.06816	12.155	< 2e-16	***
income_nt$20,000 - nt$49,999	0.65898	0.07672	8.590	< 2e-16	***
income_nt$50,000 - nt$99,999	1.07615	0.08634	12.464	< 2e-16	***
income_nt$100,000 - nt$149,999	1.19116	0.12827	9.287	< 2e-16	***
income_nt$150,000 - nt$199,999	1.23261	0.19549	6.305	2.88e-10	***
income_nt$200,000 or more	1.27922	0.21840	5.857	4.70e-09	***
edu_senior (vocational) high school	0.27104	0.05981	4.532	5.84e-06	***
edu_junior/science and technological college	0.52085	0.09951	5.234	1.66e-07	***
edu_university	0.03572	0.11576	0.309	0.757651	
edu_graduate school	-0.62282	0.18079	-3.445	0.000571	***
edu_other	-0.37244	0.26719	-1.394	0.163339	

(Figure 3: My Logistic Model Regression Coefficients)

Liu (2012) explains the bell shape regression coefficients as richer families are more likely to send their children to more expensive private tutoring programs rather than cram schools. However, the phrasing of the original survey question for cram school participation did include private tutoring classes (Chang, 2004). Therefore, it's not enough to interpret the data by just looking at income or education variables independently while ignoring their interrelated nature.

EFFECT OF CRAM SCHOOL UNDER MATCHING

To confirm Liu's finding, I calculated the direct effect of cram school on academic performance by matching people who went to cram school and those who didn't and shared the same family income and education background and subtract their academic performance. GenMatch suggested the weights of matching for income and education variables and gave an estimated effect of 3.2 of cram school attendance on academic performance (see Figure 4). The worse balance of matching among all variables is 0.317 (see Figure 5), which is the highest despite I had tried to lump variables together. This is not a strong balance to support the estimate, but the estimate is aligned with the replications in multivariate regression model, which implies that the assumption of linear model is on track.

```
Estimate...  3.2071
SE.........  0.13811
T-stat.....  23.221
p.val......  < 2.22e-16

Original number of observations..............  12366
Original number of treated obs...............  10560
Matched number of observations...............  10560
Matched number of observations  (unweighted).  10560
```

(Figure 4)

```
Before Matching Minimum p.value: < 2.22e-16
Variable Name(s): w1p515_num w1p104_num  Number(s): 1 2

After Matching Minimum p.value: 0.31731
Variable Name(s): w1p515_num  Number(s): 1
```

(Figure 5)

WHO GOES TO CRAM SCHOOL ON SOCIAL MAP

To examine the accuracy of Liu's logistic model, I turned the model into a predictive machine and cross validated its out-of-sample test error. I first split the samples in half training set and half test set. The confusion matrix told me that the model is 85.7% accurate. I further cross validated the test error through k-fold cross validation which gave me 76.8% accuracy when predicting outside samples (see code in Appendix). This information further support the generalization of conclusions derived from this sample and logistic model.

According Bourdieu (1985), the sociologists that raised the concept of social space, people's behaviors and mindsets are framed by their economic capital (e.g. income) and cultural

capital (e.g. education). Liu's finding only showed the effects of education or income independently, so I run the logistic model again with an interaction term of education and income (see Figure 6 in Appendix) and reorganized the statistically significant outcomes on a social map (see Figure 7 in Appendix). It's clear that people who are just below the average education line and people in the fourth quadrants (high income low edu) are most motivated to send their kids to cram schools to compensate the cultural capital they themselves are lacking. In contrast, those who are already educated are less motivated, but still very likely to send their kids to cram school to retain and transmit the cultural capitals they have. This indicates which social group will be impacted the most if the government would change the policy on cram schools. For further research, we can find a economic indicator as a dependent variable and see how these clusters will move on economic scale (x-axis) relative to their movement on education scale (y-axis). For example, on what economic level would people start prioritizing their kids' education (e.g. cram school attendance) than income benefits.

CONCLUSION

My replication and extension of Liu's paper showed that there are indeed positive effect of going to cram school on students' academic performance. People, especially those who have limited education background but are well-off rely on this venue heavily and desperately seek to improve their kids' cultural capital as a compensation for their own lack of education. A single decision to ban the cram school will be inappropriate in this context. The key is to diversify the existing repertoire of cultural capitals and attach its meaning to different non-academic talents and self-cultivations rather than a single indicator as education. More qualitative research needs to be done to find the alternative solutions.

REFERENCE

Bourdieu, P. (1985). The social space and the genesis of groups. *Social Science Information*, *24*(2), 195-220. http://dx.doi.org/10.1177/053901885024002001

Chang, Ly-yun. (2004). Taiwan Education Panel Survey: *Wave1(2001) Junior High School Students (Public version)（C00124_2）* [Data file]. Available from Survey Research Data Archive, Center for Survey Research, Research Center for Humanities and Social Sciences, Academia Sinica Web site: https:// srda.sinica.edu.tw. doi: 10.6141/TW-SRDA-C00124_2-1

Chang, Ly-yun. (2004). *Taiwan Education Panel Survey: Wave1 (2001) Junior High School Parents (Public version)（C00124_8）* [Data file]. Available from Survey Research Data

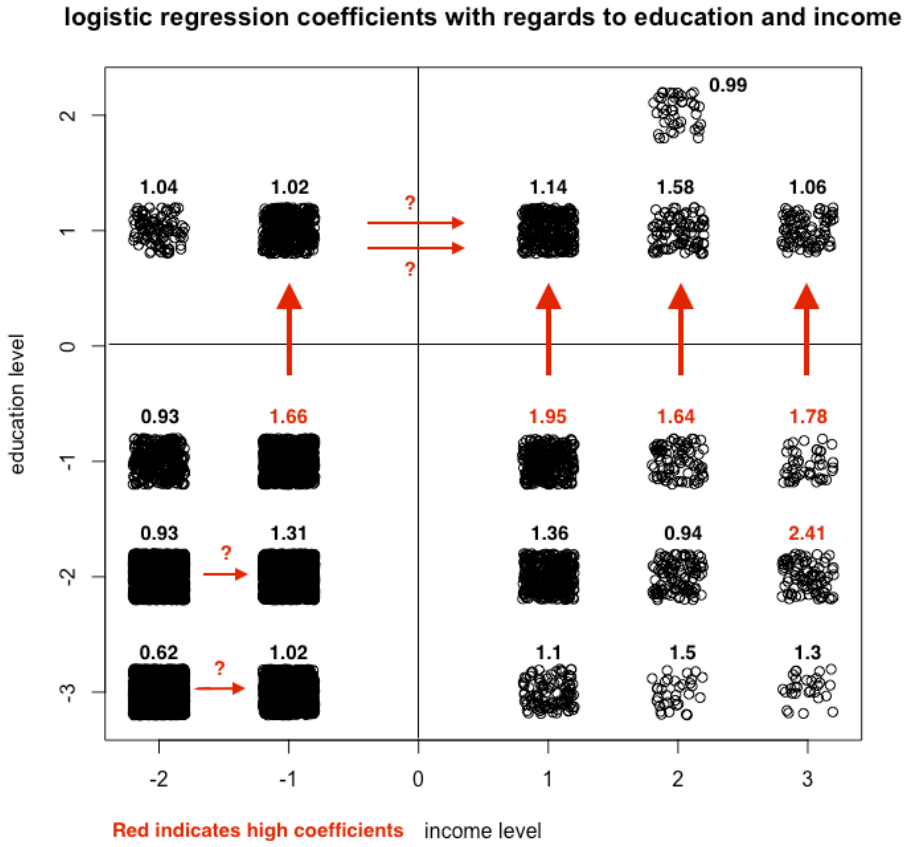
Archive, Center for Survey Research, Research Center for Humanities and Social Sciences, Academia Sinica Web site: https:// srda.sinica.edu.tw. doi: 10.6141/TW-SRDA-C00124_8-1.

Liu, J. (2012). Does cram schooling matter? Who goes to cram schools? Evidence from Taiwan. *International Journal Of Educational Development*, *32*(1), 46-52. http://dx.doi.org/10.1016/j.ijedudev.2011.01.014

Appendix:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.86072	0.08006	10.751	< 2e-16	***
Int$20,000 - nt$49,999.junior high school or less	0.61884	0.09653	6.411	1.45e-10	***
Int$50,000 - nt$99,999.junior high school or less	1.01781	0.12854	7.918	2.41e-15	***
Int$100,000 - nt$149,999.junior high school or less	1.09612	0.30692	3.571	0.000355	***
Int$150,000 - nt$199,999.junior high school or less	1.47465	0.60992	2.418	0.015615	*
Int$200,000 or more.junior high school or less	1.25954	0.61623	2.044	0.040959	*
Iless than nt$20,000.senior (vocational) high school	0.18073	0.15005	1.204	0.228409	
Int$20,000 - nt$49,999.senior (vocational) high school	0.92948	0.10017	9.279	< 2e-16	***
Int$50,000 - nt$99,999.senior (vocational) high school	1.30733	0.10629	12.300	< 2e-16	***
Int$100,000 - nt$149,999.senior (vocational) high school	1.36158	0.19513	6.978	3.00e-12	***
Int$150,000 - nt$199,999.senior (vocational) high school	0.94209	0.28994	3.249	0.001157	**
Int$200,000 or more.senior (vocational) high school	2.41011	0.59363	4.060	4.91e-05	***
Iless than nt$20,000.junior/science and technological college	0.16890	0.37699	0.448	0.654142	
Int$20,000 - nt$49,999.junior/science and technological college	0.93437	0.17243	5.419	6.00e-08	***
Int$50,000 - nt$99,999.junior/science and technological college	1.66322	0.15010	11.081	< 2e-16	***
Int$100,000 - nt$149,999.junior/science and technological college	1.95394	0.26955	7.249	4.20e-13	***
Int$150,000 - nt$199,999.junior/science and technological college	1.63798	0.43217	3.790	0.000151	***
Int$200,000 or more.junior/science and technological college	1.77834	0.60295	2.949	0.003184	**
Iless than nt$20,000.university	-0.45526	0.53309	-0.854	0.393110	
Int$20,000 - nt$49,999.university	1.04024	0.30794	3.378	0.000730	***
Int$50,000 - nt$99,999.university	1.02201	0.17298	5.908	3.46e-09	***
Int$100,000 - nt$149,999.university	1.14301	0.19476	5.869	4.39e-09	***
Int$150,000 - nt$199,999.university	1.58782	0.40201	3.950	7.83e-05	***
Int$200,000 or more.university	1.05887	0.33264	3.183	0.001457	**
Iless than nt$20,000.graduate school	-13.42678	324.74371	-0.041	0.967020	
Int$20,000 - nt$49,999.graduate school	0.52557	1.12090	0.469	0.639151	
Int$50,000 - nt$99,999.graduate school	0.42448	0.32343	1.312	0.189381	
Int$100,000 - nt$149,999.graduate school	0.48735	0.26343	1.850	0.064312	.
Int$150,000 - nt$199,999.graduate school	0.98511	0.44653	2.206	0.027376	*
Int$200,000 or more.graduate school	0.28441	0.44127	0.645	0.519234	
Iless than nt$20,000.other	0.23789	0.52257	0.455	0.648939	
Int$20,000 - nt$49,999.other	-0.27293	0.40245	-0.678	0.497655	
Int$50,000 - nt$99,999.other	0.69742	0.55591	1.255	0.209641	
Int$100,000 - nt$149,999.other	11.70534	187.49088	0.062	0.950219	
Int$150,000 - nt$199,999.other	11.70534	324.74371	0.036	0.971247	
Int$200,000 or more.other	0.23789	1.15747	0.206	0.837161	

(Figure 7: Logistic Model Regression Coefficients with the interaction term)



(Figure 8: Logistic regression coefficients on social map)

Details of Replication

Original Data	Taiwan Education Panel Survey https://srda.sinica.edu.tw/group/scigview_en/2/13
Study data	See the codebook
Sample size	(cleaned) 12366
Code	https://gist.github.com/anonymous/b3709a35afe025cd2823e164c3a466a0
Matching Covariates	W1p515 (income), w1p104 (education), I(w1p515*w1p104). All of them are factors that have many subfactors.

Codebook

Variable	Meaning	Property

w1nright	Numbers of problems answered right in analytical ability test.	Y (Continuous)
w1s1131	Cram School Participation (the summer before grade 7.)	X1 (Y/N)
w1s4284	Cram School Participation during primary school	X1 (Y/N)
w1s108a	Hours spent in Cram School	X2 (Continuous)
w1p515	Family total Income	X3 (Factors)
X3.1.w1p515.	Less than nt$20,000	X3.1
X3.2.w1p515.	nt$20,000 - nt$49,999	X3.2
X3.3.w1p515	nt$50,000 - nt$99,999	X3.3
X3.4.w1p515	nt$100,000 - nt$149,999	X3.4
X3.5.w1p515	nt$150,000 - nt$199,999	X3.5
X3.6.w1p515	nt$200,000 more	X3.6
w1p104	Parent’s highest education	X4 (Factors)
X4.1.w1p104	Junior high school or less	X4.1
X4.2.w1p104	Senior (Vocational) high school	X4.2
X4.3.w1p104	junior/science and technology college	X4.3
X4.4.w1p104	University	X4.4
X4.5.w1p104	Graduate	X4.5
X4.6.w1p104	Other	X4.6