Data Science for Social Impact Final

Daniel Farnand, Armanda Lewis, Jonathan Mejia

This document was prepared in R markdown. To look at the code that generates it, either look at the Rmd file (also attached) or find it along with all the other relevant code and data at (github.com/dfarnand/OLPCReplicationProject)[https://github.com/dfarnand/OLPCReplicationProject].

1 Replicate Original Results

This section will replicate the original results published in the paper. Unfortunately, one of the tables can not be replicated here, as it was generated with individual educational data from the Peruvian government, from which students or teachers may have been identifiable if released. This table contained descriptive characteristics of the schools, however, and does not contain results or effects.

In order to generate the remaining results, the Stata do file `resultados.do` was run. This file loads the raw data in `docente.dta` and `estudiante.dta`, runs all statistic functions and models, and saves the appropriate results tab-separated format (although they are confusingly given the extension of `xls`).

1.1 School Characteristics

Table 1: Unavailable due to confidentiality risk

NA

1.2 Sample Size

Table 2: Student Sample Sizes

	2nd Grade	Followed Cohort	6th Grade	All	Interviewed Sample
Baseline (Nov 2008)	NA	1333	NA	NA	NA
Follow Interview (Nov 2010)	NA	1305	1314	NA	2619
Followup Tests (Nov 2010)	1419	1333	1339	4091	2672

1.3 Pretreatment Balance

Table 3: Pretreatment Balance (Followed Cohort)

	Treatment	Control	Adj. Difference (SE)	Obs
Math	-0.006	0	0.007 (0.091)	1333
Language	0.039	0	0.056 (0.092)	1335
Average Academic Achivement	0.016	0.006	0.025 (0.085)	1333
Overage	0.164	0.15	0.016 (0.023)	1335

	Treatment	Control	Adj. Difference (SE)	Obs
Female	0.494	0.51	0.001 (0.028)	1335
Native Lang Spanish	0.878	0.88	0.005 (0.023)	1335
Attended Preschool	0.736	0.71	0.02 (0.034)	1335

1.4 Followup Balance

Table 4: Balance at Followup (Interviewed Sample)

	Treatment	Control	Adj. Difference (SE)	Obs
Age	10.81	10.73	0.085 (0.054)	2619
Female	0.493	0.514	-0.012 (0.02)	2612
Native Lang Spanish	0.816	0.835	-0.005 (0.019)	2621
Number of Individuals in Household	5.655	5.552	0.087 (0.089)	2617
Number of Siblings in Household	3.039	2.966	0.028 (0.104)	2621
Father Attained More than Primary Education	0.376	0.392	-0.011 (0.024)	2620
Mother Attained More than Primary Education	0.216	0.232	-0.017 (0.021)	2621
Mother's Native Lang Spanish	0.678	0.654	0.033 (0.026)	2621
TV	0.655	0.662	-0.008 (0.029)	2618
Radio	0.806	0.801	0 (0.022)	2617
Cell Phone	0.304	0.374	-0.065 (0.033)	2615
Electricity	0.802	0.79	0.007 (0.031)	2618
Running Water	0.697	0.683	0.015 (0.036)	2618
Sewage	0.174	0.144	0.019 (0.026)	2617
Cement Floor	0.122	0.113	0.014 (0.017)	2620
Recieves Conditional Cash Transfer	0.343	0.298	0.036 (0.03)	2606
More than 5 Books	0.299	0.262	0.041 (0.027)	2617
Less than 15 Minutes from School	0.656	0.634	0.03 (0.029)	2619

1.5 Compliance to Treatment

Table 5: Treatment Compliance

	Treatment	Control	Adj. Difference (SE)	Obs
School Received Laptops	1	0.083	0.916 (0.02)	317
School has Electricity	0.971	0.944	0.023 (0.023)	316
School has Internet Access	0.01	0	0.009 (0.01)	317
Teacher Received Training	0.71	0.068	0.634 (0.028)	938

1.6 Effects on Computer Use

To evaluate the treatment effects, the authors of the paper use the multiple hypothesis testing adjustment detailed by Benjamini, Krieger, and Yekutieli (Benjamini et al 2006), specifically **Sharpened Two-Stage q-values**. Code was provided to calculate these values, however it unfortunately relies on a manual process, pasting in a vector of p-values with which it will calculate the resulting q vector. Attempts were made to replicate this with R, but it appears an implementation of this particular method has not yet been developed.

Even when using the Stata code, however, the exact values published in the paper were not obtained (though never so drastically as to change an interpretation of significant). For this reason, we opted to calculate the p value adjustment for this replication using the similar Benjamini & Hochberg (1995) method.

Table 6: Effects on Computer Access and Use (Interviewed Sample)

	Treatment	Control	Adj. Difference (SE)	q-value	Obs
School has computers (School Level)	0.986	0.55	0.418 (0.037)	2.059e-23	317
Number of computers per student at the school (School Level)	1.178	0.119	1.046 (0.049)	0	312
Student has a computer	0.874	0.091	0.781 (0.027)	0	2609
Computer access summary measure	2.716	0	2.71 (0.093)	0	2609
Used a computer last week	0.843	0.321	0.518 (0.041)	2.287e-28	2610
Used a computer at school last week	0.82	0.265	0.55 (0.042)	6.311e-30	2610
Used a computer at home last week	0.418	0.038	0.39 (0.03)	2.745e-29	2610
Used a computer in a private center last week	0.072	0.081	-0.008 (0.018)	0.7754	2610
Ever used internet	0.177	0.114	0.065 (0.023)	0.05118	2605
Computer use summary measure	0.901	-0.002	0.914 (0.064)	1.178e-34	2617

The p adjustment used here was slightly more conservative than that used in the original paper, leading the "ever used internet" variable to no longer be considered significant within an alpha of 0.05.

1.7 Laptop Use and Competence

Table 7: Laptop Use (from Logs), and Competancy (Percent correct responses in Interview)

	mean
No Sessions in Last Week	0.1585
One Session in Last Week	0.1539
Two Sessions in Last Week	0.1116
Three Sessions in Last Week	0.106
Four or more Sessions in Last Week	0.4699
Laptop Use at School	0.6265
Laptop Use at Home	0.3735
Proportion Standard Applications	0.4713
Proportion Games	0.1702
Proportion Music	0.1146
Proportion Programming	0.05666
Propotion Other	0.1871
Number of Students with Laptop Logs Extracted	0.81
Competancy: Basic operation	0.7874
Competancy: Journal application	0.5718
Competancy: Write application	0.6694

	mean
Competancy: Wikipedia application	0.6036
Competancy: Picture books	0.6329
Competancy: Stories	0.6575
Average competence	1962
Number of students Interviewed	1712

1.8 Effects on Learning Behavior and Non-cognitive Outcomes

Table 8: Effects on Learning Behaviors and Noncognitive Outcomes

	Treatment	Control	Adj. Difference (SE)	q-value	Obs
Enrollment (School Level)	55.87	56.87	-1.754 (2.346)	0.6251	312
Attendance	0.782	0.75	0.029 (0.018)	0.306	4735
Studied at home yesterday	0.84	0.808	0.039 (0.022)	0.257	2617
Studied at home one or more hours daily last week	0.642	0.637	0.006 (0.032)	0.897	2620
Read a book yesterday	0.358	0.431	-0.065 (0.03)	0.1932	2617
Read a book last week	0.782	0.81	-0.017 (0.027)	0.6734	2610
Learning behaviors summary measure	-0.033	-0.001	-0.017 (0.04)	0.7884	2621
Intrinsic Motivation Index	-0.089	0	-0.091 (0.064)	0.3645	2618
Self-perceived school competence index	-0.119	0	-0.141 (0.067)	0.1932	2616
Noncognitive outcomes summary measure	-0.105	-0.001	-0.116 (0.056)	0.1932	2614

1.9 Effects on Academic Achievement and Cognitive Outcomes

Table 9: Effects on Academic and Cognitive Outcomes

	Overall Adj. Difference (SE)	Overall q-value	Interviewed Adj. Difference (SE)	Interviewed q-value
Math	0.052 (0.06)	0.5927	0.11 (0.066)	0.2879
Language	-0.029 (0.055)	0.7166	-0.003 (0.058)	0.9756
Academic Achievement Summary	0.011 (0.053)	0.8893	0.053 (0.056)	0.5535
Raven's Progressive Matrices	0.106 (0.055)	0.228	0.065 (0.061)	0.5267
Verbal Fluency Test	0.127 (0.088)	0.3645	0.119 (0.094)	0.4265
Coding Test	0.078 (0.095)	0.5951	0.1 (0.099)	0.5451
Cognitive Skills Summary	0.103 (0.058)	0.257	0.094 (0.061)	0.3233
Academic Achievement/Cognitive Skill Summary	0.065 (0.049)	0.4108	0.075 (0.05)	0.3527

1.10 Heterogeneous Effects on Academic Achievement and Cognitive Outcomes

Table 10: Heterogeneous Effect on Academic Achievement (Note that Standard Error is given in parenthesis, and q value is given is square brackets

	Second Grade	Followed Cohort	Sixth Grade	Female	Male	Low Baseline Score	High Baseline Score
Math	-0.074	0.011	0.202	0.024	0.079	-0.085	0.166
	(0.09)	(0.082)	(0.075)	(0.066)	(0.067)	(0.077)	(0.098)
	[0.595]	[0.926]	[0.064]	[0.794]	[0.452]	[0.505]	[0.282]
Language	-0.085	-0.061	0.054	-0.016	-0.038	-0.046	-0.017
	(0.086)	(0.074)	(0.068)	(0.061)	(0.066)	(0.077)	(0.076)
	[0.545]	[0.595]	[0.611]	[0.853]	[0.698]	[0.684]	[0.885]
Math/Language	-0.08	-0.027	0.128	0.003	0.021	-0.067	0.076
Summary	(0.082)	(0.071)	(0.062)	(0.057)	(0.061)	(0.071)	(0.082)
measure	[0.545]	[0.793]	[0.193]	[0.976]	[0.802]	[0.554]	[0.554]
Raven's	0.185	-0.029	0.155	0.088	0.112	0.06	0.142
Progressive	(0.077)	(0.074)	(0.07)	(0.065)	(0.06)	(0.079)	(0.073)
Matrices	[0.134]	[0.793]	[0.193]	[0.402]	[0.249]	[0.623]	[0.228]
Verbal	0.133	0.147	0.089	0.097	0.164	0.133	0.24
fluency test	(0.102)	(0.101)	(0.097)	(0.094)	(0.091)	(0.109)	(0.132)
	[0.411]	[0.364]	[0.554]	[0.536]	[0.255]	[0.44]	[0.255]
Coding test	0.042	0.13	0.069	0.069	0.105	-0.047	0.237
	(0.102)	(0.107)	(0.104)	(0.099)	(0.098)	(0.125)	(0.122)
	[0.788]	[0.44]	[0.664]	[0.658]	[0.526]	[0.793]	[0.228]
Cognitive	0.12	0.08	0.105	0.085	0.125	0.047	0.205
Summary	(0.069)	(0.065)	(0.066)	(0.063)	(0.06)	(0.069)	(0.086)
measure	[0.257]	[0.44]	[0.309]	[0.402]	[0.193]	[0.664]	[0.138]
Overall	0.041	0.035	0.112	0.052	0.081	-0.001	0.151
Summary	(0.065)	(0.058)	(0.054)	(0.053)	(0.052)	(0.061)	(0.074)
Measure	[0.673]	[0.684]	[0.193]	[0.545]	[0.323]	[0.985]	[0.193]

In general the results replicated well, with a minor exception and the required change in p value adjustment. If recommendations could be made to the researchers, it would be that they functionalize their code better to not require copying and pasting of values. Furthermore, as even following their procedure the exact q values were not duplicated, it would help if they could explain if extra steps are needed.

2 New Analyses

2.1 Redoing Analyses with Stratum and School-Level random Intercepts

The text mentions looking at the data as a multilevel model by Strata and "Randomization Group" (i.e. school), however the Stata code shows a slightly different model, a normal regression using clustered standard error. Though this is likely to yield similar results, redoing the analysis while allowing for random intercepts is relevant to confirming the authors findings.

The following tables are based on tables 6, 8, and 9 (both above and in the original paper). In general the model used was the following:

$$\text{Outcome}_{ijk} = b_0 + b_1 \text{Treatment}_{ijk} + \zeta_{jk} + \eta_k + \epsilon_{ijk}$$

Where ζ_{jk} refers to the School-specific random effect ($\zeta_{jk} \sim N(0, \sigma_\zeta^2)$) and η_k is the Strata-specific random effect ($\eta_k \sim N(0, \sigma_\eta^2)$).

The exception is school-level outcomes (marked where they occur). For these, there was obviously no need to control for school-level effects, therefore the model was somewhat simplified.

$$\text{Outcome}_{jk} = b_0 + b_1 \text{Treatment}_{ijk} + \eta_k + \epsilon_{jk}$$

Where η_k is the Strata-specific random effect, $\eta_k \sim N(0, \sigma_\eta^2)$.

Table 11: Effects on Computer Access and Use (Interviewed Sample)

	Adj. Difference	Standard Error	q Value
School has computers (School Level)	4.039	0.6218	3.724e-10
Number of computers per student at the school	1.059	0.04709	0
Student has a computer	0.7737	0.02844	0
Computer access summary measure	2.686	0.0987	0
Used a computer last week	0.5298	0.03825	0
Used a computer at school last week	0.5626	0.03942	0
Used a computer at home last week	0.3794	0.0386	0
Used a computer in a private center last week	-0.007662	0.01735	0.719
Ever used internet	0.05731	0.0242	0.07391
Computer use summary measure	0.9027	0.06969	0

With this model, the difference between the treatment and control in terms of "Ever having used a computer" is much smaller. This provides further validity to the comparison of the effects on the treatment group later.

Table 12: Effects on Learning Behaviors and Noncognitive Outcomes

	Adj. Difference	Standard Error	q Value
Enrollment (School-Level)	-1.484	2.33	0.6299
Attendance	0.003478	0.006486	0.6877
Studied at home yesterday	0.0327	0.02449	0.2994
Studied at home one or more hours daily last week	0.00863	0.03329	0.8425
Read a book yesterday	-0.06497	0.03098	0.1327
Read a book last week	-0.01983	0.02875	0.6094
Learning behaviors summary measure	-0.01899	0.0417	0.719
Intrinsic Motivation Index	-0.09506	0.06416	0.2643
Self-perceived school competence index	-0.1406	0.07042	0.1414
Noncognitive outcomes summary measure	-0.1169	0.05862	0.1414

A few effect were found to be negative, but not significantly so, so we can conclude its most likely the result of pure measurement error on zero effect.

Table 13: Effects on Academic and Cognitive Outcomes (Interviewed Sample)

	Adj. Difference	Standard Error	q Value
Math	0.1179	0.07021	0.2146

	Adj. Difference	Standard Error	q Value
Language	-0.004407	0.0595	0.941
Academic Achievement Summary	0.05723	0.05888	0.4268
Raven's Progressive Matrices	0.08892	0.0626	0.2685
Verbal Fluency Test	0.1281	0.1033	0.3238
Coding Test	0.1239	0.1084	0.3515
Cognitive Skills Summary	0.1123	0.06712	0.2146
Academic Achievement/Cognitive Skill Summary	0.08916	0.05521	0.2149

2.2 Analysis of Laptop Use and Competance

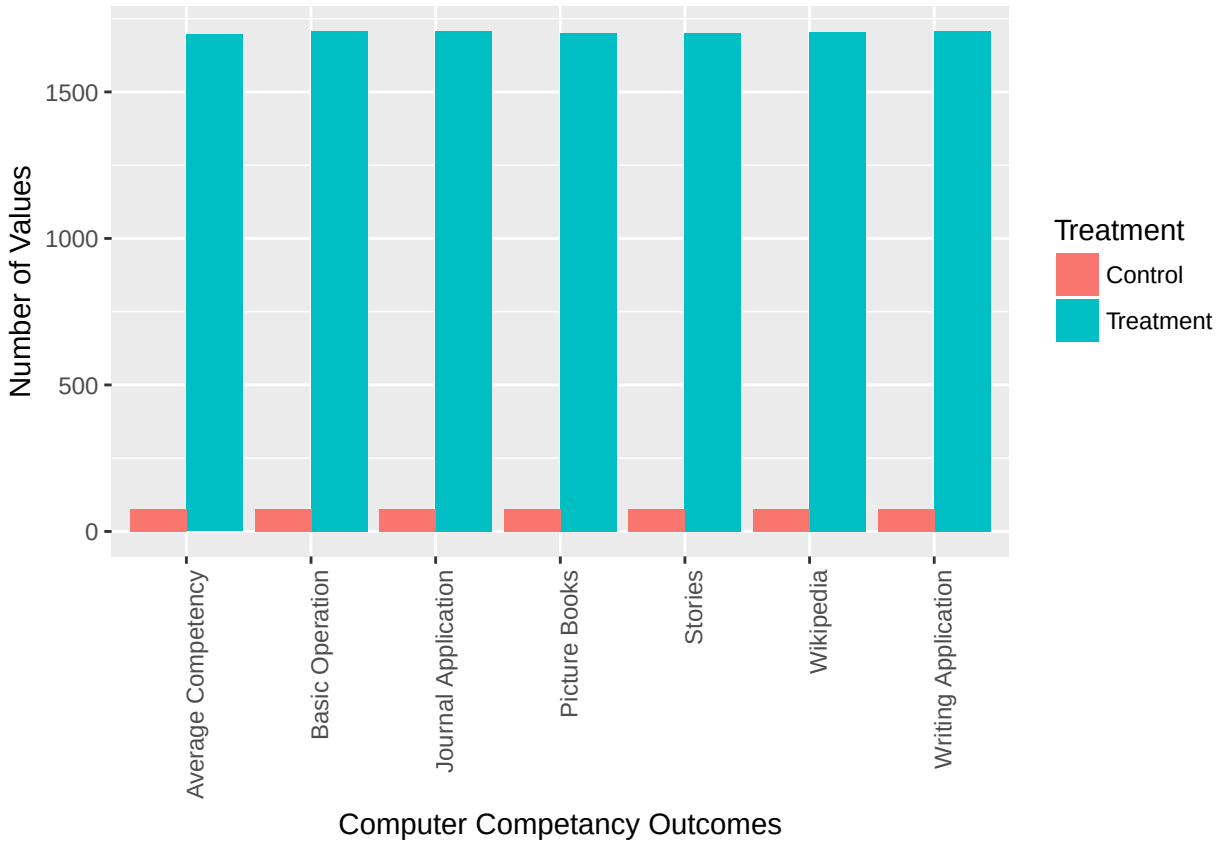
Table 7 in the paper refers to measure of laptop use and competence, however these were limited to the treatment group. The reason given was that the OLPC laptops use a specialized Linux operating system that differs in its interface from that of a windows computers. The researchers argue that comparing such skills between the two groups would be an invalid comparison. However data was collected for these competencies for at least some of the control groups, therefore we will examine the results when applying the same multilevel model to this data. This will include the competencies (*Panel B*), but not the indicators of laptop use, as these data were extracted from the logs of the laptops, so the control group will obviously not provide any observations.

Table 14: Effect on Computer Competencies

	Adj. Difference	Standard Error	q Value
Basic Operation	0.1069	0.04045	0.008805
Journal Application	0.2359	0.08164	0.00683
Writing Application	0.2423	0.0622	0.0004433
Wikipedia	0.2944	0.0774	0.0004433
Picture Books	0.2194	0.07856	0.00683
Stories	0.2122	0.07622	0.00683
Average Competency	0.2198	0.05573	0.0004433

These all appear to be significantly greater competencies in the treatment group as compared with the control. However, the discussion of whether this comparison (a competency of opening Wikipedia on an OLPC laptop, compared with on a Windows computer, for example) is actually valid at all. Furthermore, the sample size of the control group with non-missing data in these categories is problematic.

```
## Loading required package: ggplot2
```



In future studies it might be recommendable to attempt to measure more general computer skills, such as typing, or locating a new application that could be more comparable between the treatment and control. The fact that such things were not included in the published paper is unfortunate, as it missed an opportunity to examine the potentially positive results of receiving the laptop: computer skills which are certainly useful in modern job markets.

3 References

- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B 57*, 289-300.