Abstract Title Page
Not included in page count.

Title: Early RCT Findings for ELL Elementary Student Learning Outcomes After a Two-Year Pedagogical Intervention

Authors and Affiliations:

Dr. Pedro R. Portes, University of Georgia
Manuel S. Gonzalez Canche, University of Georgia
Robert Stollberg, University of Georgia

Background / Context:

The Instructional Conversation (IC) is a complete classroom pedagogy that focuses on teaching through small-group dialogue with students allowing responsive instruction for each student. The IC is anchored by cognitive-developmental theory and by four decades of multi-method quasi-experimental studies. One study (Saunders, 1999) delivered the IC intervention as a key component of an elementary language-arts program. This study found differences between control and experimental groups in terms of English language development and reading achievement. A subsequent study (Saunders & Goldenberg, 1999a) showed statistically significant effects of the IC on achievement. This study and others involving ELLs (see Slavin and Cheung, 2005) are methodologically limited; most are quasi-experiments that are generally underpowered and lack strong fidelity of implementation or counterfactual evidence (Portes, Gonzalez Canche & Stollberg, Forthcoming). In a subsequent WWC report of all published studies in literacy and language development of ELLs, the IC approach ranked highest in English Language development. These rankings far surpassed the average rank of any competing program (Institute for Educational Sciences, 2007). The study presented here describes the first research of an-IES funded, randomized efficacy trial of IC effects on standardized test outcomes for both ELL and Non-ELL students.

Theory of Change Framework

The elements of the IC are highly consistent with both the learning sciences literature (see Bransford et al., 2005; Duschl & Hamilton, 2011; Tokuhama-Espinosa, 2014) and cultural historical theory (Cole, 1996; Vygotsky, 1978, 1986). The intervention's theory of change predicts that once the critical components that define the IC model are well implemented, the responsive assistance generated for ELLs can significantly improve their academic achievement. The model is hypothesized to work mainly through the interaction between cognitive and second language development that is produced by conversational experiences framed around specific academic goals with contextualized and challenging instructional practices. For ELLs in particular, the assistance provided in focused small group learning experiences is pivotal. The IC model aims to promote higher-order thinking skills that contribute to reading comprehension and other areas through mediated learning activities.

ELL students require ample opportunity for output in order to acquire correct forms and fluency (Swain, l985). As learners dialogue with the teacher and peers about meaningful exercises, they are given time to practice English. This, in turn, promotes reading comprehension and related language skills that translate into competence in test performance areas worded in English. The IC also affords the teacher sufficient time to model standard forms of language and to respond to each student by offering the feedback, theorists argue, is necessary for a learner to "notice" correct forms of language (Krashen, 1985, DeKeyser, 1998).

Finally, the IC model addresses some important mediators that contribute to ELL engagement and development. Seixas and Peck (2004) among others have noted that a truly student-centered model involves teaching the ability to see and understand the world from a perspective outside our own. For Latino ELLs in particular, Boutakidis, Rodriguez and Knutson (2014) note a significant interaction between academic engagement and grade point average not found for sampled non-Latino students. Latino-serving non- Hispanic teachers tend to be more effective when they have an interest in and/or knowledge of their students' cultures (Moll & Arnot-Hopffer, 2005), one that the IC facilitates through contextualization in challenging conversational activities throughout the school year. Teachers serving Latino ELLs also report a

preference for using pedagogies that are useful and applicable to all learners over those which are reported as useful primarily for Latino and/or ELL populations (Rader-Brown & Howley, 2014; Adesope, Lavin, Thompson and Ungerleider, 2011).

Purpose / Objective / Research Question / Focus of Study:

The main research question for this study is: Do ELLs taught by teachers who implement the IC pedagogical model perform above controls? We aim to explore how well and for whom the pedagogy works if indeed effects are found. Our focus in this multi-cohort design (over three years) involves districts and schools where students are clustered.

The main questions are stated as three null hypotheses:

1. In general, experimental group students do not perform significantly different from controls in any content area exam.
2. Experimental ELL students taught by IC teachers do not perform significantly above ELL control students in our target content areas (English Language Arts, Reading) or other standardized content areas also examined.
3. Experimental non-ELL students do not perform significantly different from non-ELL control students in Language or other content areas.

Setting:

This efficacy study takes place across 12 school districts in the state of Georgia with moderate ELL concentration (10-48%). Teachers were recruited using various communications and local meetings to explain the study design and incentives for the two-year commitment required.

Population / Participants / Subjects:

In this study, 121 teachers completed the trial across three randomized cohorts. In the two cohorts for whom data is available, we analyzed data from 1527 students nested in the randomly assigned classrooms during the efficacy year following a year of professional development as described below. Principals assigned teachers to students in ways that varied by district and school characteristics, so that some teachers were the sole study participants in their schools and the others were in schools with other control or experimental teachers. Documentation of this variation was collected and is being evaluated. .

Following WWC standards, attrition is defined as the ratio of those teachers who dropped out of the study to the total number of randomized teachers, which here was 1.1%. As shown in available attrition analyses (that account for differential attrition across treated and control status), even when applying a rigorous standard, this study does not suffer from significant attrition bias. Although the present study is a cluster RCT, the assessment of attrition at the subcluster (i.e., individual within a cluster, in this case student within a classroom) level was not necessary, given that there were no students who decided to stop participation due to control or treatment status.

The student sample was a near-even split in ELL status (761 non-ELL, 766 ELL). The total sample of 3rd and 5th grade students employed in our analyses is split by experimental condition, grade and gender. Most of the ELL students were receiving English for Speakers of Other Languages (ESOL) services or had recently exited into regular classrooms where they

were monitored by teachers Table 1 shows descriptives of the student sample as they pertain to CRCT scores.

Research Design:

The research design is a clustered randomized controlled trial at the classroom level and designed to avoid the most common confounding factors. One factor is single unit study (WWC, 2014), where there is only one school taking part of the study. Another example is when "the characteristics of the units in each group differ systematically in ways that are associated with the outcomes." (WWC, 2004, p. 19). In the latter case, if all teachers participating in the IC would have had attained graduate degrees, whereas their control counterparts would only have lesser degrees. Following WWC standards, the outcome variables are standardizes student scores from the Georgia Criterion-Referenced Competency Tests (CRCT), standardized tests given to both 3rd and 5th grade students across five content areas (English Language Arts, Reading, Math, Science, and Social Studies. For Cohort 3 data yet to be released, the CRCT content tests are being replaced by the Georgia Milestones Assessment System) which eliminates one of the CRCT content areas and has scores yet to be reported.

Data Collection and Analysis:

Our results concern student outcomes - particularly those for ELLs (n= 766) - for the first two cohorts of students across content areas (mainly Reading and English Language Arts but also Math, Science, and Social Studies), Data regarding student standardized achievement were gathered from participating school districts for two of three cohorts of teachers (n = 121), of which the experimental group (n =59) implemented the IC.[1]

All the findings regarding test outcomes use standardized Z-scores for each grade level (given that these standardized scores are not scaled across grade level), the measure of each student's test score with the mean subtracted, standardized in number of standard deviations by grade level, then aggregated for the final analyses. Afterwards, both data sets were aggregated to test the research questions. The analytic strategy focused on OLS analyses of standardized CRCT content areas by ELL status, condition, and gender. We include all relevant CRCT content areas where impact is shown beyond ELA and Reading (Math, Science and Social Studies). Our statistical model is as follows:

$$Y_i = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon_i (1)$$

where X_1 is a dummy variable indicating ELL status, X_2 represents student gender, X_3 indicates whether the student was in 3rd or 5th grade, and ϵ_i the error term[2].

Intervention / Program / Practice:

The IC model was implemented over a two year period per cohort. Experimental teachers were first instructed in organizing classroom practices and in using the Center for Research on Education, Diversity & Excellence (CREDE) standards for pedagogical delivery. The intervention consisted of a year of professional development for experimental teachers including

[1] Power estimates indicate that this study is not underpowered, See Figure 1 caption for details.

[2] Multilevel models that account for the nested nature of the data are being currently fitted and will incorporate teacher level indicators in the final analysis of all three cohorts. Preliminary results taken from these models (null models) indicate that ELL students have and interclass correlations coefficient (ICC) of .30, whereas their non-ELL counterparts have an ICC of practically zero. This justifies the inclusion of multilevel analyses for the ELL sample.

two full-day fall renewals[3]. Experimental group teachers were also provided an initial intensive week of PD and then received coaching feedback throughout the practice years. The IC teachers and controls were monitored and assessed throughout both the practice and efficacy years. Controls were also provided incentives to continue teaching as they would normally. Both groups were required to complete logs and to be videotaped during small group teaching lessons.

Findings / Results:

With respect to the general hypothesis, the results indicate that experimental students do perform significantly different from controls in one content area exam (see Table 2 through 6). More specifically, treated students performed better in English Language Arts than their control peers for the entire sample and across cohorts ($\beta = .120$, $p<.05$).

With respect to the main research question (hypothesis 2), as shown in Tables 7 through 11, the models indicate that ELL students taught by IC teachers did perform significantly above ELL control students in English Language Arts, Mathematics, and Science. In addition, treated ELL girls in 5th grade outperformed their female control counterparts. In terms of coefficient magnitudes, treated ELL students outperformed their control counterparts by .25 ($p. <.001$) in English Language Arts. It is worth noting that further analyses disaggregated by gender and grade level indicate that most of this gain is driven by the improvement of 5th grade female ELL students ($\beta = .342$, $p <.001$). The coefficients in Math and Science were 0.147 and 0.139, respectively ($p < .10$). There was no difference in CRCT performance attributable to grade level only (without gender) in this sample.

Finally, the last hypothesis compares the performance of Non-ELL students across treatment and control statuses. The results showed that, based on the analyses of two-cohorts of data, there are no differences between IC experimental and control non- ELL students in any content areas.

Conclusions:

Overall, the IC had positive effects mainly on ELL students in the predicted language and related outcomes. There were no apparent effects of the IC on non-ELL students which raises important questions based on the literature reviewed here suggesting an omnibus effect. Our final report will include additional student data and analysis on school effects as part of a hierarchical linear model.

Our findings show that ELL students' reading and other areas of academic performance improved through the IC approach. ELL students' learning potential can be assisted by teaching practices sensitive to language and literacy development and these IC discussions in general are still not generally found in most regular elementary schools (see Lawrence, Crosson, Blagoev & Snow, 2015). The increase in language arts and hypothesized higher order thinking seem to drive or fan higher performance in math and science areas perhaps through better comprehension of test questions keyed to student learning objectives In sum, pedagogical models that challenge ELLs to articulate their thinking appears to promote learning allowing teachers to identify and utilize the competence and cultural resources that ELLs bring to the classroom. Further research is needed to determine the nature of differential (IC) impacts on ELL and other groups.

[3] Details of the coaching and intensive PD program are available upon request in reports that document fidelity of implementation, counterfactual evidence, and studies of videos contrasting IC and control teachers.

References

Adesope, O.O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research, 80*(2), 207-245.

Boutakidis, I.P., Rodríguez, J.L., Miller, K.K., & Barnett, M. (2014). Academic engagement and achievement among Latina/o and non-Latina/o adolescents. *Journal of Latinos and Education, 13*(1), 4-13.

Bransford, J., Barron, B., Pea, R.D., Meltzoff, A., Kuhl, P., Bell, P., . . . Reeves, B. (2005). Foundations and opportunities for an interdisciplinary science of learning. *The Cambridge handbook of the learning sciences*, 39-77.

Cole, M. (1996). Interacting minds in a life-span perspective: A cultural-historical approach to culture and cognitive development. *Interactive minds*, 59-87.

DeKeyser, R. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. *Focus on form in classroom second language acquisition*, 42-63.

Dörnyei, Z. (2001). *Motivation strategies in the language classroom*: Ernst Klett Sprachen.

Dörnyei, Z., & Schmidt, R.W. (2001). *Motivation and second language acquisition* (Vol. 23): Natl Foreign Lg Resource Ctr.

Duschl, R., & Hamilton, R. (2011). Learning science. *Handbook of research on learning and instruction*, 78-107.

Ellis, N.C. (2008). The dynamics of second language emergence: Cycles of language use, language change, and language acquisition. *The Modern Language Journal, 92*(2), 232-249.

Ercikan, K. (2003). Are the English and french versions' of the third international mathematics and science study administered in canada comparable? Effects of adaptations. *International Journal of Educational Policy Research and Practice, 4*, 55-75.

Gardner, R.C., Tremblay, P.F., & Masgoret, A. (1997). Towards a full model of second language learning: An empirical investigation. *The Modern Language Journal, 81*(3), 344-362.

Institute for Education Sciences. (2007). Instructional conversations and literature logs.

Krashen, S.D. (1985). *Inquiries & insights: Second language teaching: Immersion & bilingual education, literacy*: Alemany Press Hayward, CA.

Lawrence, J.F., Crosson, A.C., Paré-Blagoev, E.J., & Snow, C.E. (2015). Word generation randomized trial discussion mediates the impact of program treatment on academic word learning. *American Educational Research Journal*, 0002831215579485.

Lee, O., & Anderson, C.W. (1993). Task engagement and conceptual change in middle school science classrooms. *American Educational Research Journal, 30*(3), 585-610.

Moll, L.C., & Arnot-Hopffer, E. (2005). Sociocultural competence in teacher education. *Journal of teacher education, 56*(3), 242.

Paris, S.G., & Paris, A.H. (2001). Classroom applications of research on self-regulated learning. *Educational psychologist, 36*(2), 89-101.

Pintrich, P.R., Marx, R.W., & Boyle, R.A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational research, 63*(2), 167-199.

Pintrich, P.R., & Schunk, D.H. (2002). *Motivation in education: Theory, research, and applications*: Prentice Hall.

Portes, P.R., Canche, M.G., & Stollberg, R. (2015). Assessing teacher quality through video analyses: From professional development to fidelity of implementation. *Forthcoming*.

Portes, P.R., & Salas, S. (2011). *Vygotsky in 21st century society: Advances in cultural historical theory and praxis with non-dominant communities*: ERIC.

Rader-Brown, L., & Howley, A. (2014). Predictors of the instructional strategies that elementary school teachers use with English language learners. *Teachers College Record, 116*, 1-34.

Saunders, W. (1999). Improving literacy achievement for English learners in transitional bilingual programs. *Educational Research and Evaluation, 5*(4), 345-381.

Saunders, W., & Goldenberg, C. (1999). The effects of instructional conversations and literature logs on the story comprehension and thematic understanding of English proficient and limited English proficient students.

Seixas, P., & Peck, C. (2004). Teaching historical thinking. *Challenges and prospects for Canadian social studies*, 109-117.

Slavin, R.E., & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Educational Research, 75*(2), 247-284.

Swain, M. (1985). Large-scale communicative language testing: A case study. *New directions in language testing*, 35-46.

Tokuhama-Espinosa, T. (2014). *Making classrooms better: 50 practical applications of mind, brain, and education science*: WW Norton & Company.

Vygotsky, L. (1978). Mind in society: The development of higher mental process: Cambridge, MA: Harvard University Press.

Vygotsky, L.S. (1986). Thought and language (rev. Ed.): Cambridge, MA: MIT Press.

What Works Clearinghouse. (2014). WWC procedures and standards handbook (version 3.0). Princeton, NJ: US department of education, institute of education sciences.

Table 1: Descriptives by Condition

Aggregate Sample

Content Area	N	Mean	St. Dev.	Min	Max
ELA	1,483	829.2	47.2	273	972
Math	1,479	832	61.9	276	990
Reading	1,492	829	50.1	286	932
Science	1,492	827.4	38.6	722	970
Social Studies	1,490	827	31.1	742	950

	All Control Students					All Treatment Students				
Content Area	N	Mean	St. Dev.	Min	Max	N	Mean	St. Dev.	Min	Max
ELA	679	826.8	48	273	930	804	831.1	46.3	280	972
Math	678	830.3	63.8	284	990	801	833.5	60.2	276	990
Reading	682	828.2	55.7	286	920	810	829.8	44.9	321	932
Science	682	826.1	37.4	727	945	810	828.4	39.6	722	970
Social Studies	681	826.1	30.5	742	933	809	827.7	31.6	753	950

	All Non-ELL Students					All ELL Students				
Content Area	N	Mean	St. Dev.	Min	Max	N	Mean	St. Dev.	Min	Max
ELA	735	830.8	58	273	972	748	827.5	33.2	283	930
Math	733	831.6	74.5	276	990	746	832.4	46.3	304	959
Reading	742	832.8	57.6	286	932	750	825.3	41.1	301	920
Science	742	832.1	38.4	740	970	750	822.7	38.3	722	970
Social Studies	740	830.8	30	759	933	750	823.2	31.7	742	950

	Control Non-ELL Students					Treatment Non-ELL Students				
Content Area	N	Mean	St. Dev.	Min	Max	N	Mean	St. Dev.	Min	Max
ELA	333	831.1	56.2	273	908	402	830.5	59.5	280	972
Math	333	832.9	75.9	284	990	400	830.6	73.4	276	990
Reading	335	835.1	57	286	915	407	830.9	58.2	321	932
Science	335	833.3	37.2	740	945	407	831.1	39.4	740	970
Social Studies	334	830.4	28.7	759	933	406	831.1	31.1	765	933

	Control ELL Students					Treatment ELL Students				
Content Area	N	Mean	St. Dev.	Min	Max	N	Mean	St. Dev.	Min	Max
ELA	346	822.7	38.2	283	930	402	831.7	27.7	761	930
Math	345	827.8	49.4	304	959	401	836.4	43.1	742	959
Reading	347	821.5	53.7	301	920	403	828.6	25.4	765	920
Science	347	819.1	36.3	727	923	403	825.7	39.7	722	970
Social Studies	347	821.9	31.6	742	922	403	824.4	31.8	753	950

TABLE 2 – Treatment Effects on ELA - Aggregate Sample

	Dependent variable :		
	ELA		
	(1)	(2)	(3)
Treatment	0.119**	0.119**	0.120**
	(0.052)	(0.052)	(0.052)
Female		0.270***	0.271***
		(0.051)	(0.052)
3rd Grade			0.012
			(0.052)
Constant	−0.065*	−0.204***	−0.211***
	(0.038)	(0.046)	(0.054)
Observations	1,478	1,478	1,478
R^2	0.004	0.022	0.022
Adjusted R^2	0.003	0.021	0.020
Residual Std. Error	0.998 (df = 1476)	0.989 (df = 1475)	0.990 (df = 1474)
F Statistic	5.240** (df = 1; 1476)	16.460*** (df = 2; 1475)	10.985*** (df = 3; 1474)
Note :			*p<0.1; **p<0.05; ***p<0.01

TABLE 3 – Treatment Effects on Math - Aggregate Sample

	Dependent variable :		
	Math		
	(1)	(2)	(3)
Treatment	0.037	0.037	0.037
	(0.052)	(0.052)	(0.052)
Female		0.112**	0.112**
		(0.052)	(0.052)
3rd Grade			0.004
			(0.052)
Constant	−0.020	−0.078*	−0.080
	(0.039)	(0.047)	(0.054)
Observations	1,477	1,477	1,477
R^2	0.0003	0.003	0.003
Adjusted R^2	−0.0003	0.002	0.001
Residual Std. Error	1.000 (df = 1475)	0.999 (df = 1474)	0.999 (df = 1473)
F Statistic	0.509 (df = 1; 1475)	2.574* (df = 2; 1474)	1.717 (df = 3; 1473)
Note :			*p<0.1; **p<0.05; ***p<0.01

TABLE 4 – Treatment Effects on Reading - Aggregate Sample

	Dependent variable :		
	Reading		
	(1)	(2)	(3)
Treatment	0.049	0.049	0.049
	(0.052)	(0.052)	(0.052)
Female		0.164***	0.164***
		(0.052)	(0.052)
3rd Grade			0.007
			(0.052)
Constant	−0.026	−0.111**	−0.115**
	(0.038)	(0.047)	(0.054)
Observations	1,482	1,482	1,482
R^2	0.001	0.007	0.007
Adjusted R^2	−0.0001	0.006	0.005
Residual Std. Error	1.000 (df = 1480)	0.997 (df = 1479)	0.997 (df = 1478)
F Statistic	0.866 (df = 1; 1480)	5.429*** (df = 2; 1479)	3.622** (df = 3; 1478)

Note : *p<0.1; **p<0.05; ***p<0.01

TABLE 5 – Treatment Effects on Science - Aggregate Sample

	Dependent variable :		
	Science		
	(1)	(2)	(3)
Treatment	0.034	0.034	0.034
	(0.052)	(0.052)	(0.052)
Female		0.056	0.056
		(0.051)	(0.052)
3rd Grade			0.003
			(0.052)
Constant	−0.019	−0.047	−0.049
	(0.038)	(0.046)	(0.054)
Observations	1,510	1,510	1,510
R^2	0.0003	0.001	0.001
Adjusted R^2	−0.0004	−0.0003	−0.001
Residual Std. Error	1.000 (df = 1508)	1.000 (df = 1507)	1.000 (df = 1506)
F Statistic	0.435 (df = 1; 1508)	0.800 (df = 2; 1507)	0.534 (df = 3; 1506)

Note : *p<0.1; **p<0.05; ***p<0.01

TABLE 6 – Treatment Effects on Social Studies - Aggregate Sample

	Dependent variable :		
	Social Studies		
	(1)	(2)	(3)
Treatment	0.045	0.045	0.045
	(0.052)	(0.052)	(0.052)
Female		−0.021	−0.021
		(0.052)	(0.052)
3rd Grade			0.0004
			(0.052)
Constant	−0.024	−0.013	−0.014
	(0.038)	(0.047)	(0.054)
Observations	1,502	1,502	1,502
R^2	0.0005	0.001	0.001
Adjusted R^2	−0.0002	−0.001	−0.001
Residual Std. Error	1.000 (df = 1500)	1.000 (df = 1499)	1.000 (df = 1498)
F Statistic	0.749 (df = 1; 1500)	0.461 (df = 2; 1499)	0.307 (df = 3; 1498)
Note :			*p<0.1; **p<0.05; ***p<0.01

TABLE 7 – Treatment Effects on ELA - ELLs Only

	Dependent variable :		
	ELA		
	(1)	(2)	(3)
Treatment	0.263***	0.251***	0.250***
	(0.071)	(0.070)	(0.070)
Female		0.344***	0.342***
		(0.069)	(0.069)
3rd Grade			−0.061
			(0.070)
Constant	−0.268***	−0.438***	−0.404***
	(0.052)	(0.062)	(0.073)
Observations	748	748	748
R^2	0.018	0.050	0.051
Adjusted R^2	0.017	0.047	0.047
Residual Std. Error	0.963 (df = 746)	0.949 (df = 745)	0.949 (df = 744)
F Statistic	13.902*** (df = 1; 746)	19.440*** (df = 2; 745)	13.208*** (df = 3; 744)
Note :			*p<0.1; **p<0.05; ***p<0.01

Table 8 – Treatment Effects on Math - ELLs Only

	Dependent variable :		
	Math		
	(1)	(2)	(3)
Treatment	0.155**	0.149*	0.147*
	(0.076)	(0.076)	(0.076)
Female		0.187**	0.186**
		(0.076)	(0.076)
3rd Grade			−0.067
			(0.076)
Constant	−0.154***	−0.246***	−0.208***
	(0.056)	(0.067)	(0.079)
Observations	752	752	752
R^2	0.006	0.014	0.015
Adjusted R^2	0.004	0.011	0.011
Residual Std. Error	1.039 (df = 750)	1.035 (df = 749)	1.035 (df = 748)
F Statistic	4.176** (df = 1; 750)	5.171*** (df = 2; 749)	3.706** (df = 3; 748)

Note : *p<0.1; **p<0.05; ***p<0.01

Table 9 – Treatment Effects on Reading - ELLs Only

	Dependent variable :		
	Reading		
	(1)	(2)	(3)
Treatment	0.105*	0.098	0.098
	(0.060)	(0.060)	(0.060)
Female		0.219***	0.220***
		(0.060)	(0.060)
3rd Grade			0.047
			(0.060)
Constant	−0.237***	−0.345***	−0.372***
	(0.044)	(0.053)	(0.063)
Observations	747	747	747
R^2	0.004	0.022	0.023
Adjusted R^2	0.003	0.019	0.019
Residual Std. Error	0.821 (df = 745)	0.814 (df = 744)	0.815 (df = 743)
F Statistic	3.049* (df = 1; 745)	8.260*** (df = 2; 744)	5.711*** (df = 3; 743)

Note : *p<0.1; **p<0.05; ***p<0.01

TABLE 10 – Treatment Effects on Science - ELLs Only

	Dependent variable :		
	Science		
	(1)	(2)	(3)
Treatment	0.144**	0.139*	0.139*
	(0.071)	(0.071)	(0.071)
Female		0.129*	0.128*
		(0.071)	(0.071)
3rd Grade			−0.040
			(0.071)
Constant	−0.195***	−0.259***	−0.236***
	(0.053)	(0.063)	(0.075)
Observations	761	761	761
R^2	0.005	0.010	0.010
Adjusted R^2	0.004	0.007	0.006
Residual Std. Error	0.983 (df = 759)	0.981 (df = 758)	0.982 (df = 757)
F Statistic	4.057** (df = 1; 759)	3.686** (df = 2; 758)	2.559* (df = 3; 757)

Note : *p<0.1; **p<0.05; ***p<0.01

TABLE 11 – Treatment Effects on Social Studies - ELLs Only

	Dependent variable :		
	Social Studies		
	(1)	(2)	(3)
Treatment	0.066	0.065	0.064
	(0.074)	(0.074)	(0.074)
Female		0.036	0.034
		(0.074)	(0.074)
3rd Grade			−0.050
			(0.074)
Constant	−0.153***	−0.171***	−0.142*
	(0.054)	(0.066)	(0.078)
Observations	755	755	755
R^2	0.001	0.001	0.002
Adjusted R^2	−0.0003	−0.001	−0.002
Residual Std. Error	1.017 (df = 753)	1.017 (df = 752)	1.017 (df = 751)
F Statistic	0.786 (df = 1; 753)	0.511 (df = 2; 752)	0.493 (df = 3; 751)

Note : *p<0.1; **p<0.05; ***p<0.01

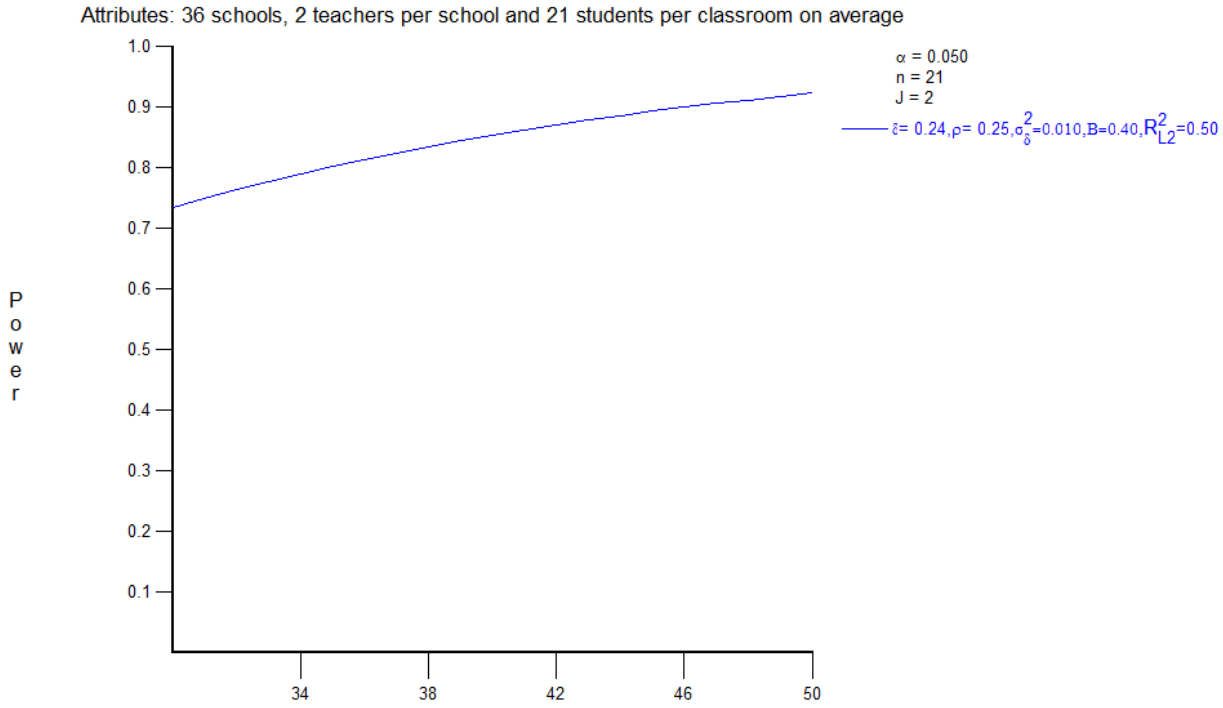

Figure 1. Power estimates for cohorts included. This figure indicates that even with two cohorts of students the study presented has enough power to detect differences across treated and control groups. More specifically, our estimations indicate that the total number of sites realized were 36, which as can be appreciated in the figure is associated with a power of .8, assuming an effect size of .24.