

**Abstract Title Page**  
*Not included in page count.*

**Title:** Teachers' Helpers: Experimental Evidence from Costa Rica on Computers for English Language Learning

**Authors and Affiliations:** Sarah Humpage (Mathematica Policy Research (presenter)) and Horacio Álvarez-Marínelli (Inter-American Development Bank).

## **Abstract Body**

*Limit 4 pages single-spaced.*

### **Background / Context:**

*Description of prior research and its intellectual context.*

Computers have taken an increasingly prominent role in education around the world in recent years in developed and developing countries alike. As developing country governments have turned their focus from increasing enrollment to improving the quality of education in their schools, many have made access to computers a key component to their strategies (Trucano, 2005). There is an emerging body of research now that shows that these strategies, which often come with a hefty price tag, have varied effects (Glewwe et al., forthcoming).

Rigorous evaluations have found that general computer use has had no or negative effects on math and language learning (Cristia et al., 2012; Malamud & Pop-Eleches, 2011; and Barrera-Osorio & Linden, 2009). Several studies have found that specific software packages have also failed to have positive effects (Barrow, Markman & Rouse, 2007; Angrist & Lavy, 2002; and Rouse & Krueger, 2004), while others have had positive effects on learning outcomes (Roschelle et al., 2010; Banerjee, Cole, Duflo & Linden, 2007; He, Linden & MacLeod, 2007; Rosas et al., 2003; and Campuzano et al., 2009).

This paper contributes to this literature by comparing the use of computer-assisted language learning software to traditional methods of English instruction in Costa Rica. We also compare two different software programs to one another. Because schools were randomly assigned to one of these two treatment groups or a control group, this research permits a rigorous comparison of the differential effects of two software programs, holding contextual factors constant.

### **Purpose / Objective / Research Question / Focus of Study:**

*Description of the focus of the research.*

Many developing countries have made English language learning a key component of their strategies to advance in the global economy (Pinon & Haydon, 2010). Costa Rica is one of these countries. This paper evaluates the effectiveness of technology as a tool to support learning English as a foreign language in primary schools in Costa Rica. By randomly assigning two software packages, DynEd English and Imagine Learning, we are able to identify differences in causal effects between the packages.

This paper addresses the following research questions: First, what is the impact of each of the two English language learning software programs on test scores, as compared to a teacher alone? Second, what is the magnitude of the effect of each program compared to the other? Third, do these effects vary by school-level baseline performance, students' baseline test scores or gender? This paper contributes to the literature by evaluating the effectiveness of computers in an area where computers may provide a critical support to teachers in a curricular area (in this case, English) in which they are likely to have limited skills and, more generally, to the literature on technology's causal effects on learning.

### **Setting:**

*Description of the research location.*

This research takes place in the Alajuela province of Costa Rica. Costa Rica has one of the most effective education systems in Latin America, as is evidenced by its performance on regional tests, which is above average for the region, even after accounting for the country's income and

expenditure per pupil (PREAL, 2009). Nonetheless, the country's ability to excel in English language instruction is limited by a limited supply of teachers qualified to teach English. The Costa Rican government is interested in the viability of computer-assisted language learning methods to address this short supply.

Alajuela is immediately north of Costa Rica's capital city, San Jose. With a level of urbanization, literacy rate and unemployment similar to the national average, the province is similar to the nation as a whole (INEC, 2013). The schools participating in the study are distributed throughout Alajuela province.

### **Population / Participants / Subjects:**

*Description of the participants in the study: who, how many, key features, or characteristics.*

The sample is comprised of 866 students who were enrolled in 3<sup>rd</sup> grade in the first year of the two-year study; the sample is approximately evenly divided among the DynEd software group, the Imagine Learning group and a control group. The sample suffered significant attrition. Restricted to students who are not missing any test score data, the sample falls to 498 students/ however, there is no significant difference in attrition by treatment group.

### **Intervention / Program / Practice:**

*Description of the intervention, program, or practice, including details of administration and duration.*

Schools assigned to the control group continued teaching English as they had in previous years according to Ministry of Education guidelines. Schools assigned to either of the treatment groups received a laptop and headset for every third grade student and either the DynEd or Imagine Learning software. The intervention lasted two years, from 2010 to 2011. In the first year, students in the treatment groups used computers every day for English instruction, while in the second year, they used the computers three days a week, and worked with their teachers the other two days. Students in the control group worked with their English teacher every day both years.

The two software programs are similar in many respects. Both use multi-media content and present material that is appropriate to students' individual levels. Whereas DynEd presents all its material only in English, the Imagine Learning software introduces early concepts in the student's native language. Key characteristics of each program are presented in Table 1.

### **Research Design:**

*Description of the research design.*

This is a randomized controlled trial. The study implementation team (distinct from the analysis team) first identified a sample of primary schools that met key criteria for inclusion: access to electricity, having an English teacher who was not involved in any other pilot, and having a minimum of five students enrolled in the third grade. After conducting the original randomization, however, the implementation team learned that numerous schools did not fulfill these criteria. Unfortunately, the team's response was not consistent across cross. Schools that did not have an English teacher in the control group were replaced with new schools that did have an English teacher, but schools without an English teacher in either treatment group were left in the sample. This introduced a systematic difference between the two groups; the treatment groups, which included schools without English teachers, were smaller, more rural and had less prior experience with English on average.

To address this problem, to estimate the treatment effect of either software, as compared to the control group, we restrict the sample for analysis to the sample of schools from the original randomization that did have English teachers; this smaller sample does not suffer from the

systematic differences that the full sample does. In this restricted sample, all schools have English teachers. We use the full sample, however, when comparing the two software programs to one another, as the criteria for inclusion were applied in the same way to these two groups. This sample includes schools with and without English teachers.

In the restricted sample, there are no significant differences in baseline test scores between the treatment and control groups, although there is a significant difference in the percent of the sample that is female (DynEd has fewer girls). In the full sample, which is only used to compare the two treatment groups, there is no significant difference in baseline test scores between the two groups, but there is a difference in gender. Baseline characteristics for the full and restricted samples are presented in Table 2.

### **Data Collection and Analysis:**

*Description of the methods for collecting and analyzing data.*

Program effects were measured as changes in student scores on the Woodcock-Muñoz Language Survey-Revised (WMLS-R). Students took this test in three rounds of data collection: at the beginning and end of third grade (the 2010 school year), and at the end of fourth grade (the 2011 school year). This test is a norm-referenced, standardized instrument that measures language proficiency in reading, writing, listening and comprehension. The instrument has strong concurrent validity with other standardized tests that measure oral language (the IDEA Proficiency Test and the Language Assessment Scale), intelligence (Wechsler Adult Intelligence Scale) and academic achievement (Wide Range Achievement Test and Woodcock-Johnson III Tests of Achievement) (Woodcock et al., 2005). The test includes picture vocabulary, verbal analogies, understanding directions, and story recall subtests, generating scores for each of these subtests as well as an oral language score, which combines items from the other subtests that are relevant to oral language skills. With the exception of gender, data on student characteristics are not available. Test scores are standardized based on control group baseline scores, so treatment effects have an effect size interpretation.

To address baseline differences, a difference in difference model is used to estimate the treatments' effects on English language proficiency at the end of the first year (round 2) and second year (round 3) of the study.

The difference in difference model controls for time-invariant differences among the two treatment groups and the control group, as well as common time trends that are found in both the treatment groups and the control group. This is seen in equation (1), where  $Test_{ijt}$  is the test score for student  $i$  in school  $j$  in time  $t$ ,  $t$  is a time dummy variable indicating whether the observation is post-treatment (in this case, post-treatment could be for round 2 or round 3),  $T_j$  indicates whether the student is in a school that is in the treatment group (this could be either DynEd or Imagine Learning),  $T_j*t$  interacts the treatment and time dummies, and  $\varepsilon_{ijt}$  is a mean-zero error term for individual  $i$  in school  $j$  and time  $t$ . The coefficient on the interaction of treatment and time indicator,  $\beta_3$ , represents the estimated treatment effect.

$$Test_{ijt} = \beta_0 + \beta_1 t + \beta_2 T_j + \beta_3 T_j * t + \varepsilon_{ijt} \quad (1)$$

This equation is estimated for effects on test score growth from baseline to round 2 and baseline to round 3, comparing each treatment group to the control group as well as to one another. Standard errors are clustered at the school level for all estimates.

## **Findings / Results:**

*Description of the main findings with specific details.*

Table 3 presents standardized test scores by wave for each group for the full sample and the restricted sample. Tables 4a, 4b and 4c present main effects. Panel A in each table represents the treatment effect at the end of the first year, while Panel B presents the treatment effect at the end of the second year. The coefficient on the time variable ( $t$  in the tables) represents the change in test scores for students in the control group, while the coefficient on the treatment interacted with the time variable represents the treatment effect. These tables show that DynEd has significant positive effects compared to the control group, while Imagine Learning has no positive effects. When comparing DynEd against Imagine Learning, DynEd has a significantly greater effect on several subtests in the first and second years. Imagine does not have a significantly greater effect on any subtest. DynEd's effects are greater in the first year of the study when students work with the software five days a week than in the second year, when they work with the software three days a week.

Tables 5 through 7 present the results of subgroup analysis to test whether treatment effects varied for schools with lower average baseline test scores, students with lower baseline test scores, or for girls. In these tables, the coefficient on the treatment interacted with time and the subgroup (e.g.,  $t \cdot \text{DynEd} \cdot \text{Low}$ ) represent the difference in treatment effect for the subgroup as compared to the effect for those that are not members of the subgroup. For example, the coefficient on  $t \cdot \text{DynEd} \cdot \text{Low}$  represents the difference in DynEd's effect for low students and DynEd's effect for higher scoring students. The subgroup analysis reveals that Dyned's effects did not vary significantly by schools' baseline test scores, although the effect was significantly greater on one subtest in the second year for students with low baseline test scores. In contrast, Imagine Learning's effects were significantly lower in low-scoring schools and for low-scoring students on some subtests in both years. Neither software's effects varied significantly by gender, but Dyned's advantage over Imagine is significantly greater for girls on some subtests in the second year.

## **Conclusions:**

*Description of conclusions, recommendations, and limitations based on findings.*

The main finding of this research is that academic software can be an effective learning tool, but that this depends on the software. Previous research has already shown that technology can be effective in some cases and ineffective in others. One of this paper's contributions is to show that these heterogeneous effects are not simply the product of using technology in different contexts (although that is likely to be important as well). By randomly assigning two different software packages to students in similar schools, this research has shown that the type of technology used matters, holding other factors constant. Furthermore, technology's effectiveness also depends on student characteristics like baseline abilities and gender.

This research also contributes to the literature that shows that technology *can* be an effective teaching aid. Students working with the DynEd software learned 53% more than students in the control group, who worked with teachers alone. Future research should explore the cost implications of using computers and software to teach English or other subjects in which teachers may lack skills, as compared to alternative solutions, such as expanded teacher training.

## Appendices

*Not included in page count.*

### Appendix A. References

- Angrist, J. & Lavy, V. (2002). New Evidence on Classroom Computers and Pupil Learning. *The Economic Journal* 112 (October), 735-765.
- Angrist, J. & Pischke, J. (2009). *Mostly Harmless Econometrics*. Princeton: Princeton University Press.
- Banerjee, A., Cole, S., Duflo, E. & Linden, L. (2007). Remedying Education: Evidence from Two Randomized Experiments in India. *The Quarterly Journal of Economics* 122(3), 1235-1264.
- Barrera-Osorio, F. & Linden L. (2009). The Use and Misuse of Computers in Education. *World Bank Policy Research Working Paper 4836*, Impact Evaluation Series No. 29.
- Barrow, L., Markman, L. & Rouse, C. (2007). Technology's Edge: The Educational Benefits of Computer-Aided Instruction. *Federal Reserve Bank of Chicago Working Paper 2007-17*.
- Bruhn, M. & McKenzie, D. (2009). In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics* 1(4), 200-232.
- Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009). Effectiveness of Reading and Mathematics Software Products: Findings From Two Student Cohorts—Executive Summary (NCEE 2009-4042). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- DynEd International, Inc. (2013). "First English Program Overview." Accessed online at <http://www.dyned.com/us/products/firstenglish/> on April 10, 2013.
- Duflo, E. (2004). "Scaling Up and Evaluation" in *Accelerating Development*, edited by Francois Bourguignon and Boris Pleskovic. Oxford, UK and Washington, DC: Oxford University Press and World Bank.
- Duflo, E., Glennerster, R., and Kremer, M. (2008). Using Randomization in Development Economics Research: A Toolkit. *Handbook of Development Economics, Vol. 4*: 3895-3962.
- Ferrando, M., Machado, A., Perazzo, I., & Haretche, C. (2010). Una primera evaluación de los efectos del Plan Ceibal en base a datos de panel. Mimeograph accessed online at [http://www.ccee.edu.uy/ensenian/catsemecnal/material/Ferrando\\_M.Machado\\_A.Perazzo](http://www.ccee.edu.uy/ensenian/catsemecnal/material/Ferrando_M.Machado_A.Perazzo)

\_I.y\_Vernengo\_A.%282010%29.Evaluacion\_de\_impacto\_del\_Plan\_Ceibal.pdf on June 10, 2013.

- Fernald, L., Gertler, P. & Neufeld, L. (2008). Role of cash in conditional cash transfer programmes for child health, growth and development: an analysis of Mexico's *Oportunidades*. *The Lancet* 371: 828-37.
- Fiszbein, A., Schady, N. & Ferreira, F. (2009). *Conditional Cash Transfers: Reducing Present and Future Poverty*. Washington, DC: The World Bank.
- Fudenberg, D., & Levine, D-K. (2006). A dual-self model of impulse control. *American Economic Review* 96: 1449-1476.
- Glewwe, P., Hanushek, R., Humpage, S., & Ravina, R. (Forthcoming). School Resources and Educational Outcomes in Developing Countries: A Review of the Literature From 1990 To 2010 in *Education Policy in Developing Countries*. Paul Glewwe, editor. Chicago, United States: University of Chicago Press.
- Glewwe, P. & Kremer, M. (2006). "Schools, Teachers and Education Outcomes in Developing Countries." In: E. Hanushek and F. Welch, editors. *Handbook of the Economics of Education*. Amsterdam, The Netherlands: Elsevier.
- Glewwe, P., Kremer, M. & Moulin, S. (2009). Many Children Left Behind? Textbooks and Test Scores in Kenya. *American Economic Journal: Applied Economics* 1 (1), 112-135.
- Greene, W. (2003). *Econometric Analysis*. New Jersey, Pearson Education. Fifth Edition.
- He, F., Linden, L. & MacLeod, M. (2007). Helping Teach What Teachers Don't Know: An Assessment of the Pratham English Language Learning Program. New York, United States: Columbia University. Mimeograph.
- Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica* 47 (1), 153-161.
- Imagine Learning, Inc. (2013). Program Overview. Accessed online at <http://www.imaginelearning.com/school/ProgramOverview.html> on April 10, 2013.
- Imbens, G. & Angrist, J. (1994). The Identification and Estimation of Local Average Treatment Effects. *Econometrica* 62(2), 467-76.
- Instituto Nacional de Estadísticas y Censos (INEC). (2013). Datos del País. Accessed online at <http://www.inec.go.cr/Web/Home/pagPrincipal.aspx> on July 6, 2013.
- Instituto Nacional de Estadística e Informática (INEI). (2007). Censos Nacionales 2007: XI de Población y VI de Vivienda.

- Lee, D.S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *Review of Economic Studies* 76(3), 1071-1102.
- Leuven, E., Lindhal, M., Oosterbeek, H. & Webbink, D. (2004). The Effect of Extra Funding for Disadvantaged Pupils on Achievement. *IZA Discussion Paper Series No. 1122*.
- Linden, L. (2008). Complement or Substitute? The Effect of Technology on Student Achievement in India. New York, United States: Columbia University. Mimeograph.
- Malamud, O., and C. Pop-Eleches. (2011). Home Computers and the Development of Human Capital. *Quarterly Journal of Economics* 126: 987-1027.
- Ministerio de Educación Pública (MEP). (2013). Número de Instituciones y Servicios Educativos en Educación Regular, Dependencia Pública, Privada y Privada Subvencionada. Accessed online at [http://www.mep.go.cr/indica\\_educa/cifras\\_instituciones2.html](http://www.mep.go.cr/indica_educa/cifras_instituciones2.html) on July 6, 2013.
- Malamud, O., & Pop-Eleches, C. (2011). Home Computers and the Development of Human Capital. *Quarterly Journal of Economics* 126: 987-1027.
- Manski, C.F. (1989). Schooling as experimentation: A reappraisal of the postsecondary dropout phenomenon. *Economics of Education Review* 8 (4), 305-312.
- Mullainathan, S. (2005). "Development Economics Through the Lens of Psychology" in *Annual World Bank Conference in Development Economics 2005: Lessons of Experience*, edited by Francois Bourguignon and Boris Pleskovic. Oxford, UK and Washington, DC: Oxford University Press and World Bank.
- One Laptop Per Child Foundation. (2013). "One Laptop Per Child Map". Website accessed May 16, 2013 at [laptop.org/map](http://laptop.org/map).
- Organization for Economic Cooperation and Development (OECD). (2013). DAC Members' Net Official Development Assistance in (2011). Accessed online at <http://www.oecd.org/dataoecd/31/22/47452398.xls> on July 10, 2013.
- Partnership for Educational Revitalization in the Americas (PREAL). (2009). How Much Are Latin American Children Learning? Highlights from the Second Regional Student Achievement Test (SERCE). Washington, DC, United States: Inter-American Dialogue.
- Penuel, W. (2006). Implementation and Effects of One-to-One Computing Initiatives: A Research Synthesis. *Journal of Research on Technology in Education* 38(3), 329-348.
- Perkins, D., Radelet, S., Lindauer, D. and Block, S. 2013. *Economics of Development*. New York, United States: W.W. Norton & Company. Seventh Edition.

- Pinon, R. & Haydon, J. (2010). English Language Quantitative Indicators: Cameroon, Nigeria, Rwanda, Bangladesh and Pakistan. A custom report compiled by Euromonitor International for the British Council.
- Rosas, R., Nussbaum, M., Cumsille, P., Marianov, V., Correa, M., Flores, P. et al. (2003). Beyond Nintendo: design and assessment of educational video games for first and second grade students. *Computers and Education* 40, 71-94.
- Roschelle, J., Shechtman, N., Tatar, D., Hegedus, S., Hopkins, B., Empson, S. et al. (2010). Integration of Technology, Curriculum, and Professional Development for Advancing Middle School Mathematics: Three Large-Scale Studies. *American Educational Research Journal* 47 (4), 833-878.
- Rouse, C. & Kreuger, A. (2004). Putting Computerized Instruction to the Test: A Randomized Evaluation of a “Scientifically Based” Reading Program. *Economics of Education Review* 23(4), 323-338.
- Sianesi, B. (2001). Implementing Propensity Score Matching in STATA. Prepared for the UK Stata Users Group, VII Meeting. London.
- Stanton, B. (2004). Assessment of Relevant Cultural Considerations is Essential for the Success of a Vaccine. *Journal of Health, Population and Nutrition* 22 (3), 286-92.
- Trucano, M. (2005). Knowledge Maps: ICT in Education. Washington, DC: infoDev / World Bank.
- United Nations. (2013). United Nations Millennium Development Goals. Accessed online at <http://www.un.org/millenniumgoals/> on July 10, 2013.
- United Nations Children’s Fund (UNICEF). (2012). Levels and Trends in Child Mortality: Report 2012.
- Villarán, V. (2010). “Evaluación Cualitativa del Programa Una Laptop por Niño: Informe Final.” Lima, Peru: Universidad Peruana Cayetano Heredia. Mimeographed document.
- Wang, S.J., Middleton, B., Prosser, L., Bardon, C.G., Spurr, C.D., Carchidi, et al.(2003). A Cost-Benefit Analysis of Electronic Medical Records in Primary Care. *The American Journal of Medicine* 114 (5), 397-403.
- Woodcock, R. W., Muñoz-Sandoval, A. F., Ruef, M., & Alvarado, C. F. (2005). *Woodcock Muñoz Language Survey–Revised*. Itasca, IL: Riverside.
- Wooldridge, J. (2002). Inverse probability-weighted M estimators for sample selection, attrition, and stratification. *Portuguese Economic Journal* 1: 117-139.

World Bank. (2012). Data retrieved from World Development Indicators Online database on October 15, 2012.

World Bank. (2013). Data retrieved from World Development Indicators Online database on July 9, 2013 .

## Appendix B. Tables and Figures

Not included in page count.

**Table 1: Characteristics of DynEd and Imagine Learning Softwares**

	<b>DynEd</b>	<b>Imagine Learning</b>
Overall approach	<ul style="list-style-type: none"> <li>• “Non-linear,” blended approach</li> <li>• Level of lessons adapt to each student’s learning style and pace</li> <li>• Type of lessons adapts to students’ learning style</li> </ul>	<ul style="list-style-type: none"> <li>• “Linear” teaching approach</li> <li>• Lessons adapt to each students’ learning style and pace</li> </ul>
Skill focus	<ul style="list-style-type: none"> <li>• Listening, speaking, reading, writing</li> <li>• Focus on listening and speech recognition</li> </ul>	<ul style="list-style-type: none"> <li>• Sight word vocabulary</li> <li>• Decoding new words</li> </ul>
Activities	<ul style="list-style-type: none"> <li>• Learning English songs</li> <li>• Games</li> <li>• Watching videos of dialogue</li> <li>• Playback of student speech compared to correct pronunciation</li> </ul>	<ul style="list-style-type: none"> <li>• Learning English songs</li> <li>• Video game-like presentation of activities</li> <li>• Writing in journals</li> <li>• Recording conversations</li> </ul>
Supplementary materials	<ul style="list-style-type: none"> <li>• Worksheets</li> </ul>	<ul style="list-style-type: none"> <li>• None</li> </ul>
Teacher training	<ul style="list-style-type: none"> <li>• Teacher training online</li> </ul>	<ul style="list-style-type: none"> <li>• None</li> </ul>
Feedback for teachers on student performance	<ul style="list-style-type: none"> <li>• Teachers track progress online</li> <li>• Students take placement and mastery tests</li> </ul>	<ul style="list-style-type: none"> <li>• Teachers track progress online</li> </ul>
Use of native language	<ul style="list-style-type: none"> <li>• None – immersion approach</li> </ul>	<ul style="list-style-type: none"> <li>• “Fade” approach explains early vocabulary and concepts in Spanish</li> </ul>
Average time spent per week	<ul style="list-style-type: none"> <li>• 67 minutes</li> </ul>	<ul style="list-style-type: none"> <li>• 127 minutes</li> </ul>

Sources: DynEd website, Imagine Learning website. Source for time spent per week: project documentation.

**Table 2: Baseline Equivalence by Group**

Panel A: Full Sample						
Full sample	Control	DynEd	Imagine	Dyned - Control	Imagine - Control	DynEd - Imagine
Female	0.539 (0.500)	0.398 (0.491)	0.527 (0.501)	-0.141** (0.056)	-0.012 (0.057)	-0.130** (0.063)
Picture Vocabulary	0.073 (1.005)	-0.175 (0.895)	-0.286 (1.044)	-0.248 (0.182)	-0.359* (0.190)	0.111 (0.217)
Verbal Analogies	0.090 (1.044)	-0.115 (0.883)	-0.251 (0.699)	-0.205 (0.199)	-0.341* (0.176)	0.136 (0.153)
Understanding Directions	0.213 (0.936)	-0.158 (1.025)	-0.163 (0.992)	-0.371* (0.200)	-0.376* (0.200)	0.005 (0.234)
Story Recall	-0.012 (0.964)	-0.127 (1.077)	-0.224 (1.074)	-0.116 (0.184)	-0.212 (0.224)	0.097 (0.207)
Oral Language	0.127 (0.992)	-0.189 (0.943)	-0.287 (0.944)	-0.316 (0.204)	-0.414* (0.209)	0.098 (0.222)
Number of students	167	166	165	333	332	331
Number of schools	23	23	24	46	47	47
Panel B: Restricted Sample <sup>a</sup> (Only Schools with English Teachers; No Replacement Schools)						
Full sample	Control	DynEd	Imagine	Dyned - Control	Imagine - Control	DynEd - Imagine
Female	0.542 (0.500)	0.364 (0.483)	0.477 (0.502)	-0.178*** (0.063)	-0.065 (0.060)	-0.112 (0.067)
Picture Vocabulary	0.043 (1.032)	0.048 (0.877)	-0.095 (0.966)	0.005 (0.217)	-0.138 (0.198)	0.143 (0.231)
Verbal Analogies	0.153 (1.063)	-0.022 (0.940)	-0.181 (0.651)	-0.174 (0.247)	-0.334 (0.214)	0.160 (0.193)
Understanding Directions	0.134 (0.928)	-0.018 (1.012)	0.085 (0.778)	-0.151 (0.240)	-0.049 (0.188)	-0.103 (0.242)
Story Recall	0.052 (0.921)	0.013 (1.023)	0.008 (0.978)	-0.039 (0.197)	-0.043 (0.250)	0.004 (0.232)
Oral Language	0.120 (0.991)	0.003 (0.934)	-0.053 (0.752)	-0.117 (0.244)	-0.173 (0.206)	0.056 (0.225)
Number of students	131	118	86	249	217	204
Number of schools	16	13	10	29	26	23

All variables have been standardized by the standard deviation and mean values from wave 1. The sample is restricted to individuals that are not missing test score data for any of the three waves. For means, standard deviations are presented in parentheses. For differences in means, standard errors are presented in parentheses and are adjusted for school-level clustering. \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$ .

**Table 3: Mean Test Scores by Group and Time Period**

Panel A: Full Sample			
	Control n=167	DynEd n=166	Imagine n=165
Panel A.1: Baseline			
Picture Vocabulary	0.073	-0.175	-0.286
Verbal Analogies	0.090	-0.115	-0.251
Understanding Directions	0.213	-0.158	-0.163
Story Recall	-0.012	-0.127	-0.224
Oral Language Composite	0.127	-0.189	-0.287
Panel A.2: End of Year One			
Picture Vocabulary	0.740	0.962	0.442
Verbal Analogies	0.216	0.030	-0.001
Understanding Directions	0.640	0.877	0.317
Story Recall	0.693	0.513	0.525
Oral Language Composite	0.726	0.796	0.400
Panel A.3: End of Year Two			
Picture Vocabulary	1.094	1.157	0.775
Verbal Analogies	0.471	0.222	0.215
Understanding Directions	1.011	1.044	0.644
Story Recall	1.095	0.915	1.033
Oral Language Composite	1.160	1.083	0.824
Panel B: Restricted Sample			
	Control n=131	DynEd n=118	Imagine n=86
Panel B.1: Baseline			
Picture Vocabulary	0.043	0.048	-0.095
Verbal Analogies	0.153	-0.022	-0.181
Understanding Directions	0.134	-0.018	0.085
Story Recall	0.052	0.013	0.008
Oral Language Composite	0.120	0.003	-0.053
Panel B.2: End of Year One			
Picture Vocabulary	0.676	1.017	0.587
Verbal Analogies	0.231	0.014	0.167
Understanding Directions	0.582	0.938	0.431
Story Recall	0.724	0.575	0.627
Oral Language Composite	0.693	0.849	0.565
Panel B.3: End of Year Two			
Picture Vocabulary	0.984	1.274	1.049
Verbal Analogies	0.446	0.359	0.509
Understanding Directions	0.950	1.104	0.941
Story Recall	1.019	0.934	1.220
Oral Language Composite	1.072	1.193	1.163

All test scores are standardized by the restricted sample control group's baseline test scores. Both samples are restricted to the sample of children with test score data for all three rounds.

**Table 4a: Effects of Dyned vs. Control**

Panel A: End of Year One (n=249)					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	0.043 (0.128)	0.153 (0.187)	0.134 (0.132)	0.052 (0.155)	0.120 (0.160)
t	0.632*** (0.133)	0.078 (0.209)	0.449*** (0.110)	0.673*** (0.209)	0.573*** (0.135)
Dyned	0.005 (0.217)	-0.174 (0.247)	-0.151 (0.240)	-0.039 (0.197)	-0.117 (0.244)
Dyned*t	0.337* (0.177)	-0.043 (0.276)	0.507*** (0.179)	-0.110 (0.249)	0.273 (0.195)
R-squared	0.153	0.010	0.135	0.111	0.128
Panel B: End of Year Two (n=249)					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	0.043 (0.128)	0.153 (0.187)	0.134 (0.132)	0.052 (0.155)	0.120 (0.160)
t	0.941*** (0.111)	0.294* (0.164)	0.816*** (0.097)	0.967*** (0.211)	0.952*** (0.146)
Dyned	0.005 (0.217)	-0.174 (0.247)	-0.151 (0.240)	-0.039 (0.197)	-0.117 (0.244)
Dyned*t	0.284 (0.196)	0.087 (0.249)	0.306 (0.186)	-0.046 (0.237)	0.237 (0.198)
R-squared	0.256	0.032	0.225	0.226	0.253

Test scores are standardized restricted sample control group baseline test score means and standard deviations. This analysis uses the restricted sample of students without missing test score data at schools with English teachers from the original randomization. Standard errors, adjusted for school-level clustering, are presented in parentheses. \* p<.1; \*\* p<.05; \*\*\* p<.01.

**Table 4b: Effects of Imagine Learning vs. Control**

Panel A: End of Year One (n=217)					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	0.043 (0.129)	0.153 (0.188)	0.134 (0.132)	0.052 (0.155)	0.120 (0.161)
t	0.632*** (0.133)	0.078 (0.209)	0.449*** (0.110)	0.673*** (0.210)	0.573*** (0.135)
Imagine	-0.138 (0.198)	-0.334 (0.215)	-0.049 (0.188)	-0.043 (0.251)	-0.173 (0.206)
Imagine*t	0.049 (0.152)	0.270 (0.282)	-0.102 (0.159)	-0.054 (0.286)	0.045 (0.166)
R-squared	0.100	0.021	0.049	0.125	0.097
Panel B: End of Year Two (n=217)					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	0.043 (0.129)	0.153 (0.188)	0.134 (0.132)	0.052 (0.155)	0.120 (0.161)
t	0.941*** (0.111)	0.294* (0.165)	0.816*** (0.097)	0.967*** (0.212)	0.952*** (0.146)
Imagine	-0.138 (0.198)	-0.334 (0.215)	-0.049 (0.188)	-0.043 (0.251)	-0.173 (0.206)
Imagine*t	0.203 (0.189)	0.397 (0.235)	0.040 (0.161)	0.245 (0.308)	0.263 (0.193)
R-squared	0.208	0.059	0.176	0.300	0.245

Test scores are standardized restricted sample control group baseline test score means and standard deviations. This analysis uses the restricted sample of students without missing test score data at schools with English teachers from the original randomization. Standard errors, adjusted for school-level clustering, are presented in parentheses. \* p<.1; \*\* p<.05; \*\*\* p<.01.

**Table 4c: Effects of Dyned vs. Imagine Learning**

<b>Panel A: End of Year One (n=331)</b>					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	-0.102 (0.158)	-0.120 (1.000)	0.003 (0.161)	-0.022 (0.160)	-0.070 (0.162)
t	0.726*** (0.092)	0.287** (0.138)	0.466*** (0.090)	0.698*** (0.136)	0.696*** (0.086)
Dyned	0.111 (0.217)	0.156 (0.176)	0.005 (0.228)	0.090 (0.193)	0.099 (0.225)
Dyned*t	0.409*** (0.140)	-0.120 (0.225)	0.540*** (0.157)	-0.101 (0.171)	0.302* (0.155)
R-squared	0.231	0.017	0.157	0.115	0.193
<b>Panel B: End of Year Two (n=331)</b>					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	-0.102 (0.158)	-0.120 (0.0996)	0.003 (0.161)	-0.022 (0.160)	-0.070 (0.162)
t	1.059*** (0.097)	0.535*** (0.121)	0.784*** (0.090)	1.171*** (0.141)	1.125*** (0.085)
Dyned	0.111 (0.217)	0.156 (0.176)	0.005 (0.228)	0.090 (0.193)	0.099 (0.225)
Dyned*t	0.270* (0.158)	-0.148 (0.217)	0.383** (0.166)	-0.201 (0.168)	0.163 (0.147)
R-squared	0.289	0.052	0.225	0.284	0.298

Test scores are standardized restricted sample control group baseline test score means and standard deviations. This analysis uses the full sample of students without missing test score data. Standard errors, adjusted for school-level clustering, are presented in parentheses. \* p<.1; \*\* p<.05; \*\*\* p<.01.

**Table 5a: Effects of Dyned vs. Control for Low-Performing Schools**

<b>Panel A: End of Year One (n=249)</b>					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	0.334* (0.179)	0.575** (0.236)	0.454*** (0.153)	0.329** (0.130)	0.527*** (0.190)
t	0.613*** (0.157)	-0.271 (0.304)	0.241*** (0.057)	0.362* (0.190)	0.318** (0.130)
Dyned	0.084 (0.240)	-0.347 (0.322)	-0.078 (0.252)	-0.119 (0.154)	-0.128 (0.248)
t*Dyned	0.195 (0.206)	0.226 (0.420)	0.492** (0.197)	0.033 (0.256)	0.327 (0.239)
Low school	-0.681*** (0.206)	-0.987*** (0.260)	-0.750*** (0.189)	-0.650** (0.280)	-0.953*** (0.214)
t*Low	0.046 (0.280)	0.816** (0.345)	0.485** (0.212)	0.726* (0.412)	0.598** (0.257)
Dyned*Low	-0.289 (0.291)	0.333 (0.357)	-0.283 (0.319)	0.133 (0.364)	-0.087 (0.310)
t*Dyned*Low	0.377 (0.359)	-0.607 (0.464)	0.098 (0.311)	-0.289 (0.485)	-0.068 (0.348)
R <sup>2</sup>	0.265	0.108	0.245	0.160	0.270
<b>Panel A: End of Year Two (n=249)</b>					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	0.334* (0.179)	0.575** (0.236)	0.454*** (0.153)	0.329** (0.130)	0.527*** (0.190)
t	0.781*** (0.131)	0.009 (0.197)	0.662*** (0.073)	0.583*** (0.178)	0.666*** (0.138)
Dyned	0.084 (0.240)	-0.347 (0.322)	-0.078 (0.252)	-0.119 (0.154)	-0.128 (0.248)
t*Dyned	0.203 (0.240)	0.299 (0.347)	0.140 (0.180)	0.238 (0.217)	0.276 (0.213)
Low school	-0.681*** (0.206)	-0.987*** (0.260)	-0.750*** (0.189)	-0.650** (0.280)	-0.953*** (0.214)
t*Low	0.374* (0.212)	0.666** (0.289)	0.361* (0.196)	0.898** (0.376)	0.670** (0.258)
Dyned*Low	-0.289 (0.291)	0.333 (0.357)	-0.283 (0.319)	0.133 (0.364)	-0.087 (0.310)
t*Dyned*Low	0.258 (0.344)	-0.477 (0.441)	0.478 (0.306)	-0.636 (0.435)	-0.020 (0.327)
R <sup>2</sup>	0.337	0.133	0.328	0.276	0.369

Test scores are standardized restricted sample control group baseline test score means and standard deviations. This analysis uses the restricted sample of students without missing test score data at schools with English teachers from the original randomization. Standard errors, adjusted for school-level clustering, are presented in parentheses. \* p<.1; \*\* p<.05; \*\*\* p<.01.

**Table 5b: Effects of Imagine Learning vs. Control for Low-Performing Schools**

<b>Panel A: End of Year One (n=217)</b>					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	0.334* (0.179)	0.575** (0.237)	0.454*** (0.153)	0.329** (0.131)	0.527** (0.190)
t	0.613*** (0.158)	-0.271 (0.304)	0.241*** (0.057)	0.362* (0.190)	0.318** (0.130)
Imagine	-0.209 (0.237)	-0.653** (0.261)	-0.202 (0.173)	0.021 (0.200)	-0.328 (0.198)
t*Imagine	0.027 (0.184)	0.744* (0.396)	0.091 (0.126)	-0.017 (0.264)	0.251 (0.160)
Low school	-0.681*** (0.207)	-0.987*** (0.261)	-0.750*** (0.190)	-0.650** (0.281)	-0.953*** (0.214)
t*Low	0.046 (0.281)	0.816** (0.346)	0.485** (0.213)	0.726* (0.413)	0.598** (0.258)
Imagine*Low	0.155 (0.323)	0.740** (0.336)	0.350 (0.301)	-0.167 (0.444)	0.352 (0.278)
t*Imagine*Low	0.053 (0.319)	-1.116** (0.497)	-0.452 (0.328)	-0.074 (0.542)	-0.480 (0.330)
R <sup>2</sup>	0.179	0.109	0.117	0.201	0.206
<b>Panel A: End of Year Two (n=217)</b>					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	0.334* (0.179)	0.575** (0.237)	0.454*** (0.153)	0.329** (0.131)	0.527** (0.190)
t	0.781*** (0.131)	0.009 (0.198)	0.662*** (0.073)	0.583*** (0.178)	0.666*** (0.138)
Dyned	-0.209 (0.237)	-0.653** (0.261)	-0.202 (0.173)	0.021 (0.200)	-0.328 (0.198)
t*Dyned	0.371** (0.148)	0.663* (0.330)	0.240 (0.172)	0.304 (0.321)	0.492** (0.201)
Low school	-0.681*** (0.207)	-0.987*** (0.261)	-0.750*** (0.190)	-0.650** (0.281)	-0.953*** (0.214)
t*Low	0.374* (0.213)	0.666** (0.290)	0.361* (0.197)	0.898** (0.377)	0.670** (0.259)
Dyned*Low	0.155 (0.323)	0.740** (0.336)	0.350 (0.301)	-0.167 (0.444)	0.352 (0.278)
t*Dyned*Low	-0.394 (0.420)	-0.623 (0.424)	-0.470 (0.336)	-0.123 (0.525)	-0.531 (0.361)
R <sup>2</sup>	0.263	0.135	0.250	0.371	0.332

Test scores are standardized restricted sample control group baseline test score means and standard deviations. This analysis uses the restricted sample of students without missing test score data at schools with English teachers from the original randomization. Standard errors, adjusted for school-level clustering, are presented in parentheses. \* p<.1; \*\* p<.05; \*\*\* p<.01.

**Table 5c: Effects of Dyned vs. Imagine Learning for Low-Performing Schools**

Panel A: End of Year One (n=331)					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	0.422** (0.179)	0.156 (0.103)	0.512*** (0.137)	0.347*** (0.118)	0.478*** (0.125)
t	0.536*** (0.100)	0.260 (0.221)	0.256*** -0.092	0.489*** (0.141)	0.488*** -0.073
Dyned	0.048 (0.234)	0.268 (0.240)	-0.001 (0.213)	0.017 (0.144)	0.088 (0.184)
t*Dyned	0.391** (0.158)	-0.298 (0.365)	0.486*** (0.174)	-0.091 (0.197)	0.230 (0.182)
Low school	-1.109*** (0.202)	-0.584*** (0.154)	-1.076*** (0.198)	-0.780*** (0.272)	-1.160*** (0.191)
t*Low	0.403** (0.171)	0.058 (0.272)	0.445*** (0.152)	0.442* (0.257)	0.441*** (0.149)
Dyned*Low	0.101 (0.276)	-0.263 (0.279)	-0.023 (0.292)	0.134 (0.321)	-0.012 (0.267)
t*Dyned*Low	0.051 (0.260)	0.390 (0.445)	0.131 (0.287)	-0.008 (0.323)	0.171 (0.287)
R <sup>2</sup>	0.405	0.113	0.328	0.196	0.405
Panel A: End of Year Two (n=331)					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	0.422** (0.179)	0.156 (0.103)	0.512*** (0.137)	0.347*** (0.118)	0.478*** (0.125)
t	1.027*** (0.0968)	0.545*** (0.195)	0.689*** (0.124)	0.917*** (0.160)	1.019*** (0.105)
Dyned	0.0480 (0.234)	0.268 (0.240)	-0.000574 (0.213)	0.0173 (0.144)	0.0876 (0.184)
t*Dyned	0.0715 (0.199)	-0.318 (0.363)	0.0910 (0.182)	-0.146 (0.187)	-0.0548 (0.178)
Low school	-1.109*** (0.202)	-0.584*** (0.154)	-1.076*** (0.198)	-0.780*** (0.272)	-1.160*** (0.191)
t*Low	0.0682 (0.201)	-0.0222 (0.239)	0.201 (0.171)	0.539** (0.253)	0.225 (0.164)
Dyned*Low	0.101 (0.276)	-0.263 (0.279)	-0.0225 (0.292)	0.134 (0.321)	-0.0121 (0.267)
t*Dyned*Low	0.436 (0.297)	0.372 (0.420)	0.645** (0.290)	-0.101 (0.306)	0.484* (0.261)
R <sup>2</sup>	0.459	0.147	0.399	0.353	0.486

Test scores are standardized restricted sample control group baseline test score means and standard deviations. This analysis uses the full sample of students without missing test score data. Standard errors, adjusted for school-level clustering, are presented in parentheses. \* p<.1; \*\* p<.05; \*\*\* p<.01.

**Table 6a: Effects of Dyned vs. Control for Low-Performing Students**

<b>Panel A: End of Year One (n=249)</b>					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	0.724*** (0.153)	0.912*** (0.270)	0.804*** (0.105)	0.671*** (0.092)	0.972*** (0.157)
t	0.518*** (0.170)	-0.186 (0.319)	0.224*** (0.070)	0.274* (0.136)	0.300*** (0.107)
Dyned	0.064 (0.201)	-0.236 (0.366)	-0.117 (0.197)	0.044 (0.118)	-0.068 (0.205)
t*Dyned	0.180 (0.207)	-0.146 (0.457)	0.345* (0.186)	-0.157 (0.214)	0.123 (0.224)
Low student	-1.183*** (0.164)	-1.352*** (0.271)	-1.221*** (0.155)	-1.039*** (0.205)	-1.499*** (0.169)
t*Low	0.269 (0.185)	0.658** (0.299)	0.504*** (0.146)	0.841** (0.306)	0.668*** (0.189)
Dyned*Low	-0.031 (0.213)	0.190 (0.366)	0.036 (0.239)	-0.091 (0.248)	0.014 (0.226)
t*Dyned*Low	0.320 (0.279)	0.156 (0.420)	0.280 (0.237)	0.049 (0.364)	0.280 (0.278)
R <sup>2</sup>	0.380	0.185	0.365	0.305	0.455
<b>Panel B: End of Year Two (n=249)</b>					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	0.724*** (0.153)	0.912*** (0.270)	0.804*** (0.105)	0.671*** (0.092)	0.972*** (0.157)
t	0.750*** (0.120)	0.042 (0.194)	0.591*** (0.078)	0.515*** (0.124)	0.644*** (0.103)
Dyned	0.064 (0.201)	-0.236 (0.366)	-0.117 (0.197)	0.044 (0.118)	-0.068 (0.205)
t*Dyned	0.126 (0.218)	0.013 (0.339)	0.097 (0.185)	-0.007 (0.164)	0.091 (0.182)
Low student	-1.183*** (0.164)	-1.352*** (0.271)	-1.221*** (0.155)	-1.039*** (0.205)	-1.499*** (0.169)
t*Low	0.451** (0.176)	0.703** (0.297)	0.480*** (0.123)	0.919*** (0.301)	0.763*** (0.177)
Dyned*Low	-0.031 (0.213)	0.190 (0.366)	0.036 (0.239)	-0.091 (0.248)	0.014 (0.226)
t*Dyned*Low	0.307 (0.210)	0.130 (0.419)	0.396* (0.225)	-0.154 (0.343)	0.262 (0.223)
R <sup>2</sup>	0.448	0.200	0.446	0.403	0.528

Test scores are standardized restricted sample control group baseline test score means and standard deviations. This analysis uses the restricted sample of students without missing test score data at schools with English teachers from the original randomization. Standard errors, adjusted for school-level clustering, are presented in parentheses. \* p<.1; \*\* p<.05; \*\*\* p<.01.

**Table 6b: Effects of Imagine Learning vs. Control for Low-Performing Students**

<b>Panel A: End of Year One (n=217)</b>					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	0.724*** (0.154)	0.912*** (0.271)	0.804*** (0.105)	0.671*** (0.092)	0.972*** (0.158)
t	0.518*** (0.170)	-0.186 (0.320)	0.224*** (0.070)	0.274* (0.136)	0.300*** (0.107)
Imagine	-0.142 (0.207)	-0.634** (0.303)	-0.217 (0.131)	-0.008 (0.148)	-0.294* (0.165)
t*Imagine	-0.079 (0.203)	0.558 (0.444)	0.095 (0.159)	0.050 (0.218)	0.163 (0.169)
Low student	-1.183*** (0.165)	-1.352*** (0.272)	-1.221*** (0.155)	-1.039*** (0.205)	-1.499*** (0.170)
t*Low	0.269 (0.186)	0.658** (0.300)	0.504*** (0.146)	0.841** (0.307)	0.668*** (0.190)
Imagine*Low	0.007 (0.269)	0.595** (0.289)	0.400* (0.205)	-0.080 (0.291)	0.279 (0.196)
t*Imagine*Low	0.306 (0.277)	-0.592 (0.427)	-0.463** (0.219)	-0.237 (0.395)	-0.280 (0.246)
R <sup>2</sup>	0.329	0.181	0.295	0.333	0.422
<b>Panel A: End of Year Two (n=217)</b>					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	0.724*** (0.154)	0.912*** (0.271)	0.804*** (0.105)	0.671*** (0.092)	0.972*** (0.158)
t	0.750*** (0.121)	0.042 (0.194)	0.591*** (0.079)	0.515*** (0.125)	0.644*** (0.103)
Dyned	-0.142 (0.207)	-0.634** (0.303)	-0.217 (0.131)	-0.008 (0.148)	-0.294* (0.165)
t*Dyned	0.284 (0.224)	0.836** (0.329)	0.124 (0.194)	0.192 (0.245)	0.419* (0.211)
Low student	-1.183*** (0.165)	-1.352*** (0.272)	-1.221*** (0.155)	-1.039*** (0.205)	-1.499*** (0.170)
t*Low	0.451** (0.177)	0.703** (0.298)	0.480*** (0.124)	0.919*** (0.302)	0.763*** (0.177)
Dyned*Low	0.007 (0.269)	0.595** (0.289)	0.400* (0.205)	-0.080 (0.291)	0.279 (0.196)
t*Dyned*Low	-0.194 (0.264)	-0.908** (0.405)	-0.204 (0.216)	0.087 (0.397)	-0.362 (0.260)
R <sup>2</sup>	0.406	0.222	0.387	0.470	0.506

Test scores are standardized restricted sample control group baseline test score means and standard deviations. This analysis uses the restricted sample of students without missing test score data at schools with English teachers from the original randomization. Standard errors, adjusted for school-level clustering, are presented in parentheses. \* p<.1; \*\* p<.05; \*\*\* p<.01.

**Table 6c: Effects of DynEd vs. Imagine for Low-Scoring Students**

<b>Panel A: End of Year One (n=331)</b>					
Variables	Picture Vocabulary	Verbal Analogies	Understanding Directions	Story Recall	Oral Language
Constant	0.819*** (0.130)	0.792*** (0.093)	0.821*** (0.101)	0.941*** (0.058)	0.761*** (0.100)
t	0.253** (0.125)	-0.350* (0.184)	0.051 (0.085)	-0.013 (0.108)	0.309** (0.122)
DynEd	-0.096 (0.163)	0.358* (0.200)	-0.006 (0.150)	-0.031 (0.084)	0.030 (0.148)
t*DynEd	0.461*** (0.158)	-0.339 (0.333)	0.390** (0.150)	-0.073 (0.179)	0.194 (0.198)
Low	-1.688*** (0.142)	-1.420*** (0.093)	-1.646*** (0.117)	-1.499*** (0.142)	-1.594*** (0.141)
t*Low	0.867*** (0.211)	0.991*** (0.179)	0.836*** (0.129)	1.106*** (0.143)	0.744*** (0.180)
DynEd*Low	0.150 (0.179)	-0.358* (0.200)	0.013 (0.183)	-0.054 (0.178)	0.052 (0.192)
t*DynEd*Low	0.040 (0.279)	0.374 (0.337)	0.307 (0.223)	0.153 (0.224)	0.258 (0.271)
R <sup>2</sup>	0.589	0.349	0.512	0.435	0.549
<b>Panel B: End of Year Two (n=331)</b>					
Variables	Picture Vocabulary	Verbal Analogies	Understanding Directions	Story Recall	Oral Language
Constant	0.819*** (0.130)	0.792*** (0.093)	0.821*** (0.101)	0.941*** (0.058)	0.761*** (0.100)
t	0.770*** (0.153)	0.152 (0.142)	0.436*** (0.127)	0.406*** (0.122)	0.874*** (0.155)
DynEd	-0.096 (0.163)	0.358* (0.200)	-0.006 (0.150)	-0.031 (0.084)	0.030 (0.148)
t*DynEd	0.160 (0.210)	-0.768*** (0.264)	0.116 (0.182)	0.0271 (0.139)	-0.053 (0.198)
Low	-1.688*** (0.142)	-1.420*** (0.093)	-1.646*** (0.117)	-1.499*** (0.142)	-1.594*** (0.141)
t*Low	0.529*** (0.182)	0.597*** (0.152)	0.701*** (0.148)	1.191*** (0.152)	0.482*** (0.165)
DynEd*Low	0.150 (0.179)	-0.358* (0.200)	0.0134 (0.183)	-0.054 (0.178)	0.052 (0.192)
t*DynEd*Low	0.331 (0.236)	1.006*** (0.277)	0.545** (0.233)	-0.200 (0.210)	0.465** (0.226)
R <sup>2</sup>	0.623	0.365	0.577	0.583	0.621

Test scores are standardized restricted sample control group baseline test score means and standard deviations. This analysis uses the full sample of students without missing test score data. Standard errors, adjusted for school-level clustering, are presented in parentheses. \* p<.1; \*\* p<.05; \*\*\* p<.01.

**Table 7a: Effects of Dyned vs. Control by Gender**

<b>Panel A: End of Year One (n=249)</b>					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	0.254 (0.151)	0.324 (0.233)	0.344** (0.154)	0.258 (0.180)	0.371** (0.165)
t	0.551*** (0.162)	0.050 (0.311)	0.357** (0.145)	0.503** (0.216)	0.478*** (0.149)
Dyned	0.002 (0.229)	-0.179 (0.317)	-0.140 (0.263)	-0.008 (0.209)	-0.101 (0.253)
t*Dyned	0.510** (0.222)	-0.008 (0.377)	0.565** (0.229)	-0.020 (0.244)	0.399* (0.224)
Female	-0.050 (0.197)	0.036 (0.192)	-0.097 (0.178)	-0.043 (0.164)	-0.053 (0.171)
t*Female	0.149 (0.166)	0.073 (0.297)	0.145 (0.198)	0.228 (0.229)	0.189 (0.175)
Dyned*Female	-0.017 (0.269)	-0.041 (0.318)	-0.066 (0.267)	-0.099 (0.196)	-0.073 (0.254)
t*Dyned*Female	-0.405 (0.313)	-0.079 (0.425)	-0.128 (0.260)	-0.116 (0.257)	-0.243 (0.276)
R <sup>2</sup>	0.159	0.011	0.139	0.116	0.132
<b>Panel A: End of Year Two (n=249)</b>					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	0.254 (0.151)	0.324 (0.233)	0.344** (0.154)	0.258 (0.180)	0.371** (0.165)
t	0.898*** (0.140)	0.342 (0.257)	0.710*** (0.082)	0.862*** (0.230)	0.910*** (0.146)
Dyned	0.002 (0.229)	-0.179 (0.317)	-0.140 (0.263)	-0.008 (0.209)	-0.101 (0.253)
t*Dyned	0.267 (0.233)	0.089 (0.355)	0.307 (0.199)	-0.039 (0.250)	0.236 (0.216)
Female	-0.050 (0.197)	0.036 (0.192)	-0.097 (0.178)	-0.043 (0.164)	-0.053 (0.171)
t*Female	0.076 (0.147)	-0.009 (0.347)	0.154 (0.145)	0.072 (0.299)	0.101 (0.222)
Dyned*Female	-0.017 (0.269)	-0.041 (0.318)	-0.066 (0.267)	-0.099 (0.196)	-0.073 (0.254)
t*Dyned*Female	0.084 (0.208)	0.025 (0.461)	0.046 (0.226)	0.026 (0.346)	0.060 (0.271)
R <sup>2</sup>	0.257	0.032	0.228	0.227	0.254

Test scores are standardized restricted sample control group baseline test score means and standard deviations. This analysis uses the restricted sample of students without missing test score data at schools with English teachers from the original randomization. Standard errors, adjusted for school-level clustering, are presented in parentheses. \* p<.1; \*\* p<.05; \*\*\* p<.01.

**Table 7b: Effects of Imagine Learning vs. Control by Gender**

Panel A: End of Year One (n=217)					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	0.254 (0.152)	0.324 (0.233)	0.344** (0.155)	0.258 (0.181)	0.371** (0.166)
t	0.551*** (0.162)	0.050 (0.312)	0.357** (0.146)	0.503** (0.217)	0.478*** (0.150)
Imagine	-0.230 (0.229)	-0.403 (0.268)	-0.114 (0.261)	-0.093 (0.257)	-0.253 (0.235)
t*Imagine	0.148 (0.183)	0.536 (0.381)	-0.116 (0.247)	0.056 (0.283)	0.157 (0.194)
Female	-0.050 (0.198)	0.036 (0.192)	-0.097 (0.178)	-0.043 (0.164)	-0.053 (0.171)
t*Female	0.149 (0.166)	0.073 (0.298)	0.145 (0.198)	0.228 (0.230)	0.189 (0.176)
Imagine*Female	0.187 (0.233)	0.046 (0.211)	0.127 (0.274)	0.105 (0.226)	0.156 (0.227)
t*Imagine*Female	-0.187 (0.207)	-0.465 (0.371)	0.055 (0.340)	-0.192 (0.331)	-0.210 (0.251)
R <sup>2</sup>	0.102	0.026	0.053	0.129	0.099
Panel A: End of Year Two (n=217)					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	0.254 (0.152)	0.324 (0.233)	0.344** (0.155)	0.258 (0.181)	0.371** (0.166)
t	0.898*** (0.140)	0.342 (0.258)	0.710*** (0.082)	0.862*** (0.231)	0.910*** (0.146)
Dyned	-0.230 (0.229)	-0.403 (0.268)	-0.114 (0.261)	-0.093 (0.257)	-0.253 (0.235)
t*Dyned	0.313 (0.232)	0.608* (0.334)	0.139 (0.233)	0.308 (0.297)	0.407* (0.233)
Female	-0.050 (0.198)	0.036 (0.192)	-0.097 (0.178)	-0.043 (0.164)	-0.053 (0.171)
t*Female	0.076 (0.148)	-0.009 (0.348)	0.154 (0.146)	0.072 (0.300)	0.101 (0.222)
Dyned*Female	0.187 (0.233)	0.046 (0.211)	0.127 (0.274)	0.105 (0.226)	0.156 (0.227)
t*Dyned*Female	-0.222 (0.241)	-0.321 (0.405)	-0.190 (0.336)	-0.158 (0.325)	-0.279 (0.299)
R <sup>2</sup>	0.209	0.061	0.177	0.301	0.246

Test scores are standardized restricted sample control group baseline test score means and standard deviations. This analysis uses the restricted sample of students without missing test score data at schools with English teachers from the original randomization. Standard errors, adjusted for school-level clustering, are presented in parentheses. \* p<.1; \*\* p<.05; \*\*\* p<.01.

**Table 7c: Effects of DynEd vs. Imagine Learning by Gender**

<b>Panel A: End of Year One (n=331)</b>					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	-0.136 (0.145)	-0.175* -0.096	0.001 (0.184)	0.033 (0.163)	-0.081 (0.162)
t	0.826*** (0.119)	0.505*** (0.156)	0.572*** (0.161)	0.701*** (0.148)	0.825*** (0.123)
DynEd	0.204 (0.209)	0.216 (0.195)	0.124 (0.252)	0.065 (0.203)	0.187 (0.230)
t*DynEd	0.344* (0.172)	-0.332 (0.241)	0.367* (0.216)	-0.153 (0.178)	0.147 (0.185)
Female	0.065 (0.137)	0.104 -0.094	0.003 (0.156)	-0.105 (0.168)	0.022 (0.136)
t*Female	-0.190 (0.144)	-0.414** (0.159)	-0.200 (0.197)	-0.006 (0.184)	-0.243 (0.150)
DynEd*Female	-0.214 (0.207)	-0.115 (0.218)	-0.297 (0.237)	0.029 (0.195)	-0.213 (0.212)
t*DynEd*Female	0.099 (0.259)	0.397 (0.307)	0.368 (0.269)	0.127 (0.204)	0.308 (0.249)
R <sup>2</sup>	0.237	0.024	0.164	0.117	0.199
<b>Panel A: End of Year Two (n=331)</b>					
Variables	Picture Vocabulary	Verbal Analogies	Und. Directions	Story Recall	Oral Language
Constant	-0.136 (0.145)	-0.175* -0.096	0.001 (0.184)	0.033 (0.163)	-0.081 (0.162)
t	1.182*** (0.116)	0.741*** (0.156)	0.821*** (0.142)	1.180*** (0.153)	1.238*** (0.123)
DynEd	0.204 (0.209)	0.216 (0.195)	0.124 (0.252)	0.065 (0.203)	0.187 (0.230)
t*DynEd	0.072 (0.190)	-0.332 (0.254)	0.216 (0.208)	-0.204 (0.186)	-0.020 (0.182)
Female	0.065 (0.137)	0.104 -0.094	0.003 (0.156)	-0.105 (0.168)	0.022 (0.136)
t*Female	-0.233 (0.149)	-0.391** (0.157)	-0.070 (0.173)	-0.017 (0.159)	-0.214 (0.149)
DynEd*Female	-0.214 (0.207)	-0.115 (0.218)	-0.297 (0.237)	0.029 (0.195)	-0.213 (0.212)
t*DynEd*Female	0.424** (0.197)	0.336 (0.277)	0.397* (0.237)	0.002 (0.213)	0.390* (0.203)
R <sup>2</sup>	0.291	0.057	0.230	0.287	0.301

Test scores are standardized restricted sample control group baseline test score means and standard deviations. This analysis uses the full sample of students without missing test score data. Standard errors, adjusted for school-level clustering, are presented in parentheses. \* p<.1; \*\* p<.05; \*\*\* p<.01.