

Supplemental Appendix for Online Publication

Remedial Education:
Evidence from a Sequence of Experiments in Colombia

Horacio Alvarez-Marinelli

Samuel Berlinski

Matias Busso

I Appendix Tables and Figures

I.A Robustness

Table A.1: Robustness: Different specifications
(standardized outcomes)

		Knowledge of letter sounds	Reading of non-words	Fluency of oral reading	Reading compre- hension	Literacy score
		(1)	(2)	(3)	(4)	(5)
Control: Baseline outcome (Y0)	Treatment x End of Grade 3	0.342*** [0.054]	0.073 [0.045]	0.168*** [0.048]	-0.002 [0.043]	0.273*** [0.049]
	Treatment x Beginning of Grade 4	0.324*** [0.062]	0.134** [0.059]	0.207*** [0.078]	0.142** [0.061]	0.265*** [0.071]
	Treatment x End of Grade 4	0.312*** [0.047]	0.082* [0.042]	0.091* [0.052]	0.051 [0.052]	0.148*** [0.045]
Controls: Individual and School	Treatment x End of Grade 3	0.336*** [0.058]	0.078 [0.047]	0.167*** [0.047]	0.006 [0.043]	0.274*** [0.054]
	Treatment x Beginning of Grade 4	0.298*** [0.061]	0.107** [0.052]	0.201** [0.078]	0.154** [0.060]	0.266*** [0.073]
	Treatment x End of Grade 4	0.287*** [0.049]	0.094** [0.040]	0.094* [0.051]	0.059 [0.051]	0.157*** [0.048]
School fixed effects	Treatment x End of Grade 3	0.448*** [0.097]	0.117 [0.086]	0.147* [0.079]	0.044 [0.075]	0.307*** [0.092]
	Treatment x Beginning of Grade 4	0.395*** [0.091]	0.154* [0.089]	0.171* [0.101]	0.184** [0.079]	0.292*** [0.101]
	Treatment x End of Grade 4	0.384*** [0.092]	0.134 [0.081]	0.078 [0.090]	0.094 [0.071]	0.193** [0.096]

Note: Models in panel 1 include baseline outcome as control, cohort, year, and strata fixed effects. Models in panel 2 include as controls: age, age squared, gender, socio-economic and disability status, class size, school number of classes, rural and morning schhol-status, along with cohort, year, and strata fixed effects. Models in panel 3 include year, cohort and school fixed effects instead of strata fixed effects. Each column-panel show the coefficients θ_h of equation (2). That is, the estimated treatment effects at different time horizons for each outcome of interest. Standard errors, shown in squared brackets, are clustered at the school-level (the unit of randomization). * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.2: Robustness: Lee-bounds
(standardized outcomes)

	Outcome	Lower Bound	SE LB	Upper Bound	SE UB
End of Grade 3	Knowledge of letter sounds	0.328	[0.050]***	0.371	[0.049]***
	Reading of non-words	0.042	[0.044]	0.106	[0.050]**
	Fluency of oral reading	0.127	[0.048]***	0.192	[0.053]***
	Reading comprehension	-0.086	[0.043]**	0.034	[0.056]
	Literacy score	0.234	[0.048]***	0.307	[0.054]***
Beg/Med. Grade 4	Knowledge of letter sounds	0.175	[2.547]	0.392	[1.674]
	Reading of non-words	-0.043	[0.061]	0.271	[0.065]***
	Fluency of oral reading	0.060	[0.067]	0.309	[0.057]***
	Reading comprehension	0.112	[0.059]*	0.281	[0.060]***
	Literacy score	0.157	[0.064]**	0.417	[0.059]***
End of Grade 4	Knowledge of letter sounds	0.302	[0.064]***	0.313	[0.069]***
	Reading of non-words	0.019	[0.062]	0.118	[0.047]**
	Fluency of oral reading	0.038	[0.058]	0.143	[0.063]**
	Reading comprehension	0.026	[0.056]	0.093	[0.053]*
	Literacy score	0.152	[0.062]**	0.222	[0.058]***

Note: Each panel shows separate estimates and standard errors of (Lee 2009) bounds for each outcome. We estimate the models separately for each time-horizon and for each outcome of interest tightening at the cohort level (strata tightening is not possible since there is not enough attrition variation at this level). * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.3: Robustness: Multiple Hypothesis Testing
(standardized outcomes)

	Knowledge of letter sounds	Reading of non-words	Fluency of oral reading	Reading com- prehension	Literacy score
	(1)	(2)	(3)	(4)	(5)
Treatment x End of Grade 3	0.349	0.073	0.157	0.001	0.270
p-value	0.000	0.142	0.002	0.977	0.000
Adjusted p-value	0.001	0.522	0.087	0.983	0.008
Treatment x Beginning of Grade 4	0.312	0.105	0.192	0.150	0.264
p-value	0.000	0.063	0.014	0.017	0.001
Adjusted p-value	0.016	0.415	0.259	0.259	0.051
Treatment x End of Grade 4	0.300	0.088	0.083	0.055	0.152
p-value	0.000	0.033	0.117	0.304	0.003
Adjusted p-value	0.002	0.322	0.522	0.598	0.119
Observations	4949	6362	6362	6362	6362

Note: Each column shows the coefficients θ_h of equation (2), that is, the estimated treatment effects at different time horizons for each outcome of interest. Point estimates and p-values are those of Table 2. Adjusted p-values are calculated using Westfall-Young stepdown methodology. All models include cohort, year, and strata fixed effects. Standard errors, shown in squared brackets, are clustered at the school-level (the unit of randomization). * significant at 10%; ** significant at 5%; *** significant at 1%.

I.B Other results

Table A.4: Treatment Effects on all outcomes and all time horizons
(standardized outcomes)

Experimental cohort	Moment in time	Outcome	Treatment effect	Standard error	
Cohort 1	End of Grade 3	Knowledge of letter sounds	0.307	[0.082]***	
		Reading of non-words	0.024	[0.065]	
		Fluency of oral reading	-0.035	[0.058]	
		Reading comprehension	-0.089	[0.058]	
	Beg/Med. Grade 4	Literacy score	0.138	[0.069]**	
		Knowledge of letter sounds	-	-	
		Reading of non-words	-0.000	[0.084]	
		Fluency of oral reading	0.008	[0.127]	
	End of Grade 4	Reading comprehension	0.296	[0.095]***	
		Literacy score	0.017	[0.118]	
		Knowledge of letter sounds	-	-	
		Reading of non-words	0.108	[0.054]**	
	Cohort 2	End of Grade 3	Fluency of oral reading	-0.012	[0.067]
			Reading comprehension	-0.053	[0.071]
			Literacy score	0.019	[0.064]
			Knowledge of letter sounds	0.144	[0.109]
Beg/Med. Grade 4		Reading of non-words	0.056	[0.085]	
		Fluency of oral reading	0.232	[0.071]***	
		Reading comprehension	0.062	[0.071]	
		Literacy score	0.222	[0.081]***	
End of Grade 4		Knowledge of letter sounds	0.195	[0.087]**	
		Reading of non-words	0.051	[0.083]	
		Fluency of oral reading	0.364	[0.113]***	
		Reading comprehension	0.098	[0.089]	
Cohort 3		End of Grade 3	Literacy score	0.348	[0.098]***
			Knowledge of letter sounds	0.237	[0.071]***
			Reading of non-words	0.026	[0.079]
			Fluency of oral reading	0.068	[0.073]
	Beg/Med. Grade 4	Reading comprehension	0.095	[0.045]**	
		Literacy score	0.158	[0.082]*	
		Knowledge of letter sounds	0.637	[0.108]***	
		Reading of non-words	0.180	[0.074]**	
	End of Grade 4	Fluency of oral reading	0.382	[0.089]***	
		Reading comprehension	0.064	[0.067]	
		Literacy score	0.525	[0.094]***	
		Knowledge of letter sounds	0.439	[0.090]***	
	Beg/Med. Grade 4	Reading of non-words	0.190	[0.080]**	
		Fluency of oral reading	0.196	[0.084]**	
		Reading comprehension	0.175	[0.086]**	
		Literacy score	0.378	[0.100]***	
End of Grade 4	Knowledge of letter sounds	0.384	[0.085]***		
	Reading of non-words	0.131	[0.082]		
	Fluency of oral reading	0.177	[0.096]*		
	Reading comprehension	0.163	[0.121]		
		Literacy score	0.309	[0.094]***	

Note: Each panel shows separate estimates of the treatment effects for each cohort. We estimate the models separately for each time-horizon and for each outcome of interest. All models include strata fixed effects. We did not collect data on the outcomes without an estimated treatment effect (marked as -). Standard errors, shown in squared brackets, are clustered at the school-level (the unit of randomization). * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.5: Treatment Effects Heterogeneity by Gender and Classroom Homogeneity (standardized outcomes)

		Knowledge of letter sounds	Reading of non-words	Fluency of oral reading	Reading comprehen- sion	Literacy score
		(1)	(2)	(3)	(4)	(5)
Gender	Boys	0.303*** [0.064]	0.032 [0.058]	0.149** [0.063]	-0.029 [0.063]	0.231*** [0.062]
	Girls	0.398*** [0.070]	0.103 [0.063]	0.178*** [0.053]	0.053 [0.054]	0.306*** [0.066]
	p-value of equal coeffs.	0.323	0.352	0.652	0.383	0.362
Homogeneity of classroom	s.d. below median	0.340*** [0.086]	0.179** [0.084]	0.202** [0.077]	0.147** [0.070]	0.331*** [0.075]
	s.d. above median	0.318*** [0.110]	-0.063 [0.072]	0.106 [0.081]	-0.117 [0.077]	0.177* [0.090]
	p-value of equal coeffs.	0.932	0.0565	0.442	0.0120	0.228

Note: Each panel shows estimates of θ_1 of equation (2) estimated separately for two different groups of control and treated students. The p-values of equal coefficients correspond to the interaction between treatment and the heterogeneity variable in a separate regression. Homogeneous classroom is an indicator variable that takes value one for classrooms with a s.d of an index of comparable subtasks measured at baseline below the median. All models include cohort, year, and strata fixed effects. The heterogeneity variable is also included as control. Standard errors, shown in squared brackets, are clustered at the school-level (the unit of randomization). * significant at 10%; **significant at 5%; *** significant at 1%.

Table A.6: Costs

		Cost in dollars
Cost per tutor		2158
Training	8 hours. Includes cost of 1 trainer per 15 tutors	40
Wages and transport	Includes wages and transport	2067
Supervision	Supervision of 2 classes per tutor. Includes salaries and transport cost	51
Cost per student	Students per tutor = 26	89
Materials		6
Tutor		83

Note. Parameters: Tutorials per tutor 5; Number of tutors 15; Exchange rate 3,000 Colombian pesos per dollar of 2016; Total number of students in tutorials 385; Total number of tutorials 76; Hourly wage 4.167.

II Data Appendix

This paper relies on two sources of information. First, before the intervention, we administered language and mathematics tests to third grade students at the beginning of each year (2015, 2016 and 2017) in all public schools in the Municipality of Manizales. After that, to measure the impact of the intervention, we collected information in the same schools at the end of third grade, the beginning/middle and the end of fourth grade using the same instrument as well. The former data consist of this test score information administered at each point in time. Details about the test design, content, scoring, and administration are presented in Section [II.C](#).

At the time of baseline of each school year, we also collected administrative school records from the Integrated Enrollment System (Sistema Integrado de Matrícula, SIMAT), the national database for the registration of students in public education in Colombia. The latter provides information on students' age, gender, socio-economic status, and whether or not students change schools or repeat grades over time.

II.A Sample Sizes

Table B.1: Sample Sizes

Cohort	Schools		Classrooms		Students	
	Control	Treated	Control	Treated	Control	Treated
1	46	48	74	75	591	552
2	42	42	57	61	377	376
3	40	40	60	62	353	361

In [Table B.1](#) we show the number of schools, classrooms and students that participated in the experiment. We started off with a sample of 94 public schools in 2015, which we dropped, mostly for cost reasons, to 84 and then 80 in the two subsequent years. The number of third grade classrooms within these schools ranged between 124 and 155. A total of 1,143, 753 and 714 students were eligible in cohorts 1, 2 and 3. It is important to note that when we measure the outcomes during fourth grade we also follow up any retained students in third grade who initially were in our sample either in treatment or control groups but have not been promoted to fourth grade during the next academic year.

Despite the small change in the number of schools between cohorts, [Table B.2](#) shows that the composition was not qualitatively altered. We compare the mean of key school-level characteristics between the three cohorts. The last three columns of the Table show that we are not able to reject any of the null hypotheses of equality in means in these variables.

Table B.2: Balance of schools across cohorts

	Cohort 1	Cohort 2	Cohort 3	p-value of differences		
				C1 vs. C2	C2 vs. C3	C1 vs. C3
Grade 3 enrollment	44.64 [34.20]	42.11 [33.42]	47.70 [35.04]	0.28	0.63	0.56
Morning	0.74 [0.44]	0.75 [0.44]	0.76 [0.43]	0.94	0.86	0.79
Rural	0.16 [0.37]	0.14 [0.35]	0.22 [0.42]	0.17	0.73	0.32
Avg. Socio-ec. status (1-6)	1.85 [0.48]	1.86 [0.49]	1.85 [0.45]	0.90	0.96	0.94
Avg. class size	24.07 [9.89]	24.74 [10.81]	23.63 [10.31]	0.48	0.68	0.78
Avg. number of classrooms in Grade 3	1.52 [0.89]	1.40 [0.73]	1.59 [0.77]	0.11	0.34	0.63

Note: Table shows mean and standard deviations of school-level variables for the sample of schools of each cohort along with the p-values of the mean’s differences.

II.B Attrition

We were not able to follow up all the students from our baseline sample in the subsequent times that we visited the schools. We estimate the probability of attrition as a function the baseline standardized scores for knowledge of letter sounds, reading of non-words, fluency of oral reading and reading Comprehension. The following table shows that, in general, the baseline scores of students that attrited from our sample are similar to the scores of those who attrited, suggesting that there is no clear pattern between the characteristics at baseline and attrition across study arms. Most estimates are small. We only reject the null of zero difference in two cases, in different directions and at the 10 percent level.

We also estimate a model for the probability of attrition using a logit specification. We do this separately for each cohort and time horizon. With the exception of age – which positively affects attrition – most variables are not statistically significant and their signs show no clear patterns across cohorts and time horizons. In addition, we are only able to explain a low proportion of the variance in attrition.

Table B.3: Attrition Correlates

		Cohort 1 (1)	Cohort 2 (2)	Cohort 3 (3)
End of Grade 3	Y0	-0.015 [0.012]	-0.017 [0.014]	-0.015 [0.017]
	T	-0.015 [0.020]	-0.025 [0.024]	0.001 [0.021]
	T x Y0	0.011 [0.018]	-0.007 [0.020]	0.004 [0.029]
Beg./Mid. Grade 4	Y0	0.065** [0.026]	-0.002 [0.020]	-0.007 [0.022]
	T	-0.067 [0.082]	-0.050 [0.041]	-0.020 [0.029]
	T x Y0	0.014 [0.038]	-0.005 [0.030]	0.008 [0.027]
End Grade 4	Y0	-0.010 [0.018]	-0.026** [0.013]	-0.027 [0.026]
	T	0.044 [0.033]	0.015 [0.027]	-0.036 [0.034]
	T x Y0	-0.023 [0.026]	0.001 [0.022]	-0.002 [0.035]

Note: Each panel-column shows the coefficients of an OLS regression of an indicator variable equal to 1 if the student is not observed at the corresponding time horizon (i.e. the student attrited from the sample by the end of grade 3, beginning/middle of grade 4, or the end of grade 4), on the baseline literacy composite standardized score (Y0), the treatment variable (T), and the interaction between the two. Standard errors are clustered at the school level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.4: Attrition and baseline characteristics by cohort and time horizon

	Cohort 1		Cohort 2		Cohort 3	
	End G3	End G4	End G3	End G4	Beg./Med. G4	End G4
Age	0.224*** [0.087]	0.145** [0.064]	0.234* [0.133]	0.192** [0.094]	0.216 [0.143]	0.388*** [0.114]
Female	-0.021 [0.241]	-0.080 [0.157]	0.041 [0.256]	-0.220 [0.236]	0.476 [0.293]	-0.126 [0.218]
Fluency of oral reading	-0.089 [0.171]	-0.124 [0.105]	-0.013 [0.155]	0.068 [0.143]	-0.240 [0.170]	0.137 [0.122]
Knowledge of letter sounds	0.074 [0.121]	0.149 [0.106]	-0.268 [0.179]	-0.026 [0.121]	0.001 [0.139]	-0.246** [0.120]
Reading comprehension	0.311 [0.202]	-0.037 [0.124]	0.092 [0.311]	0.160 [0.206]	0.055 [0.159]	-0.034 [0.107]
Additions	-0.217* [0.128]	0.060 [0.078]	0.438*** [0.152]	-0.043 [0.150]	0.226 [0.181]	-0.141 [0.137]
Subtractions	0.137 [0.157]	-0.125 [0.099]	-0.676*** [0.210]	-0.077 [0.160]	-0.048 [0.165]	0.093 [0.152]
Grade 3 enrollment	0.036** [0.018]	-0.017 [0.015]	-0.032 [0.042]	0.003 [0.037]	-0.028*** [0.011]	-0.011 [0.012]
Morning school	-0.201 [0.392]	0.300 [0.276]	-0.485 [0.421]	0.139 [0.338]	-0.630** [0.279]	-0.021 [0.237]
Rural school	-0.665* [0.371]	1.730*** [0.412]	-0.189 [0.391]	0.618 [0.398]	0.363 [0.393]	0.360 [0.364]
School avg. Socio-ec. status (1-6)	-0.175 [0.517]	0.905 [0.586]	1.238* [0.709]	0.848 [0.644]	0.244 [0.532]	0.493 [0.426]
School avg. class size	-0.108*** [0.036]	0.070* [0.040]	0.065 [0.051]	-0.011 [0.053]	0.034 [0.026]	0.016 [0.027]
School avg. number of classrooms in G.3	-1.197* [0.645]	0.117 [0.665]	0.628 [1.508]	-0.683 [1.337]	0.604* [0.340]	0.144 [0.373]
Observations	1,142	1,142	753	753	714	714
Pseudo R2	0.0723	0.131	0.0773	0.0676	0.0459	0.0483
AIC	0.457	1.211	0.550	0.795	0.621	0.875
BIC	-7448	-6586	-4509	-4325	-4184	-4003

Note: Table shows estimates of a logit model of student's attrition on baseline characteristics for each cohort and time horizon. Dependent variable is an indicator for whether the student attrited. Standard errors, shown in squared brackets, are clustered at the school-level (the unit of randomization). * significant at 10%; ** significant at 5%; *** significant at 1%.

II.C Instruments and Test Scores

As described in Section III.A our measures of student learning are captured by EGRA and EGMA tests (Early Grade Reading/Mathematics Assessment).⁴⁵ The test contains four subtasks of literacy and two subtasks of math: knowledge of letter sounds, reading of non-words, fluency of oral reading, reading comprehension, addition and subtraction. In the first two components, children are asked to recognize the letters and invented words. After that, a simple passage is given to students and they are asked to read it aloud and answer several questions about it. The last two subtasks involve students solving math operation of one- and two- digit numbers to measure their math knowledge. We combine all these subtasks scales into an aggregate score that measures literacy and math knowledge: literacy score and math score.

Each subtask was scored separately by counting the number of correct answers. All these raw scores were standardized within grade-subtask-cohort by the mean and standard deviation observed in the control group at the corresponding point of measurement. We also normalized the aggregate literacy and math scores, which are the proportion of correct answers in all subtasks.

II.D Test Subtasks Over Time

Table B.5 presents how the test scales vary from each subtask in terms of item construction. Since it is common for an existing EGRA instrument to be modified into one or more parallel versions, the instrument we administered at each point in time has been modified by re-randomizing the items with grade-level equivalents in the first three subtasks. Even though some minor scaling differences may exist, the outcomes are comparable across the different grades and cohorts because the items have been modified in order to be as similar as possible in terms of length and difficulty.

Table B.5 presents all the items for all the subtasks and tests. Column 1 shows the different composition of items for the Knowledge of Letter Sounds. Letters of the alphabet are distributed randomly and evenly among the upper- and lowercase letters, ten letters to a line. As mentioned before, we have three different scales for this subtask but with equivalent test items in terms of difficulty for each one. We look next at the reading fluency which is a one-paragraph passage that contains same sentence structures and complexity. Note that this subtask, as well as reading comprehension, is not constructed with a constant number of

⁴⁵The Colombian Institute for the Assessment of Education (ICFES) administered a standardized test (Prueba Saber) during 2015-2017 to some students attending the 5th and 3rd grade. The inclusion of these outcomes in the analysis, however, was not feasible for a number of reasons. In the case of the 3rd grade Saber, the 2015 and 2016 waves were designed to measure individual performance. Students received either a math or language booklet but not both and booklets were different among children taking the same subject test and attending the same classroom. In 2017, all children took the test in both subjects. However, the test was administered in early September when our intervention had only started in July. In the case of the 5th grade Saber, only the 2017 version covered children that may have participated from our intervention. These children belong to the 2015 cohort were gains from the program at the end of the third and fourth grade were limited. We requested a match and, unfortunately, only 60 percent of the observations matched to our records. Attrition was statistically different for treatment and control students.

Table B.5: Scales by Grade and Cohort

Grade:	Cohort 1				Cohort 2				Cohort 3			
	3	3	4	4	3	3	4	4	3	3	4	4
Time of School Year:	Beg.	End	Beg.	End	Beg.	End	Beg.	End	Beg.	End	Beg.	End
Scales:												
Knowledge of Letter Sounds	1	1	-	-	1	1	2	2	2	2	2	3
Reading Fluency	1	2	3	4	3	5	5	5	5	5	4	4
Reading Comprehension	1	2	3	4	5	6	7	7	7	7	8	8
Reading of Non-words, Addition, Substraction	Always the same scales in all cohorts and years											

items across instruments. For example, the first scale contains 59 words while the second one has 64 words in the paragraph. However, despite these differences, the scores are comparable since we created a composite score by standardizing them for each grade, year and cohort to place students on the same scale. Lastly, the scales were exactly the same for all cohorts and years for the reading of non-words, addition and subtraction subtasks.

II.E Item Response Theory

Item response theory (IRT) is used in the design, analysis, scoring, and comparison of tests and similar instruments whose purpose is to measure unobservable characteristics of the respondents. IRT models specify a relationship between a single underlying latent achievement variable (ability) and the probability of answering a particular test question (item) correctly. We use the Rasch model to assess items and to score subjects on their abilities or other latent traits. In the Rasch model, the probability of a correct response is given by

$$Pr(Y_{ij} = 1|\theta_j) = \frac{e^{\alpha(\theta_j - b_i)}}{1 + e^{\alpha(\theta_j - b_i)}}$$

where α represents the discrimination common to all items, b_i represents the difficulty of item i , and θ is the latent trait of person j . The probability of a correct response is determined by the item's difficulty and the subject's ability.

Appendix Tables B.8 - B.11 presents a detailed item analysis of four subtasks administered at the endline of third grade to ensure that the items performed well in terms of discrimination and difficulty. We also assess how reliable the scale is for each subtask by computing Cronbach's alpha coefficient. All alpha coefficients of all four different scales are higher than 0.8, suggesting that the items have relatively high internal consistency across study arms.

Table B.6: Intra-Item Correlation Matrix

Subtask	Time of School Year	Baseline	Endline G3	Beg/Mid. G4	Endline G4
Knowledge of letter sounds	Baseline	1.000			
	Endline G3	0.425	1.000		
	Beg/Mid. G4	0.365	0.510	1.000	
	Endline G4	0.425	0.533	0.518	1.000
Reading of non-words	Baseline	1.000			
	Endline G3	0.581	1.000		
	Beg/Mid. G4	0.479	0.560	1.000	
	Endline G4	0.512	0.580	0.585	1.000
Reading Fluency	Baseline	1.000			
	Endline G3	0.693	1.000		
	Beg/Mid. G4	0.305	0.397	1.000	
	Endline G4	0.492	0.585	0.515	1.000
Reading Comprehension	Baseline	1.000			
	Endline G3	0.231	1.000		
	Beg/Mid. G4	0.135	0.163	1.000	
	Endline G4	0.076	0.162	0.317	1.000
Addition	Baseline	1.000			
	Endline G3	0.598	1.000		
	Beg/Mid. G4	0.465	0.562	1.000	
	Endline G4	0.525	0.591	0.570	1.000
Substraction	Baseline	1.000			
	Endline G3	0.467	1.000		
	Beg/Mid. G4	0.319	0.414	1.000	
	Endline G4	0.433	0.493	0.432	1.000

Note: All values are significant at 1%.

Table B.7: Test items

Knowledge of Letter Sounds IDScale	Reading Fluency IDScale	Reading Comprehension IDScale	Reading of Non-words IDScale	Addition IDScale	Substraction IDScale
1 M d r O E C i u p S A n j T b e f r W L m r D E y O a g s Z c v N I k U P x L Q S N O A d T i N a e	1 María y Juan fueron a jugar al parque. En el parque estaba Jaime. Jaime había perdido a su peluche. Estaba muy triste. María y Juan lo ayudaron a buscar. El peluche estaba dormido bajo un árbol. Despierta, peluche, vamos a jugar dijo Jaime. Los niños jugaron hasta que llegó la noche. Todos estaban contentos ese día. El peluche también.	1 ¿A dónde fueron a jugar María y Juan? ¿Por qué estaba triste Jaime? ¿Qué hicieron Juan y María al ver triste a Jaime? ¿Dónde estaba el peluche? ¿Hasta cuándo jugaron los niños en el parque?	1 lete quibe bofa mise garo cafa Celu bede lura mesi lluno Rite duso jata fica luma Altí lufa frate dulte lodo Fosu gesa lemo golpa bosa rale flano trabu bulo pluva arcu cince llusia firta onti zaca queno bana juru foba lise vodo tuzi listu quira cuto ganco rafo duba	1 2+2=(4)3+2=(5)4+2=(6) 1+5=(6)3+4=(7)7+1=(8) 6+2=(8)5+4=(9)4+5=(9) 7+2=(9)6+4=(10)5+5=(10) 8+2=(10)5+6=(11)6+6=(12) 3+9=(12)5+7=(12)8+6=(14) 10+3=(13)2+11=(13) 13+3=(16)6+10=(16) 10+10=(20)15+5=(20) 11+9=(20)	1 4-2=(2)8-1=(7)5-2=(3) 6-2=(4)8-2=(6)6-5=(1) 9-2=(7)9-4=(5)8-3=(5) 9-5=(4)7-4=(3)10-2=(8) 10-3=(7)10-4=(6)20-10=(10) 11-6=(5)11-7=(4)12-9=(3) 12-7=(5)12-14-6=(8)16-3=(13)16-10=(6) 20-5=(15)20-4=(16)20-9=(11)
2 M d r O E F i u p S A n j T b e f r G L m R D E y O a g s Z F V N I b U P R L M S v O A d T i N a e	2 Había un perrito gordo y peludo llamado Toto. La familia con quien vivía lo quería mucho. Toto era un perro obediente, cuidaba muy bien la casa, pero no se comía toda su comida. Un día salió de paseo con su dueño Lucas y se perdió. Lucas se puso triste, pero felizmente Todo apareció en el parque. Lucas lo cargó y lo llevó a casa.	2 ¿Cómo se llamaba el perro? ¿Qué hacía muy bien el perro? ¿El perro es flaco o gordo? ¿Dónde apareció el perro? ¿Quién es Lucas?			
3 M d r O E C i u p S A n j T b e f r W L m r D E y O a g s Z c v N I k U P x L Q S N O A d T i N a e	3 La Gallina y el Cienpiés se pusieron a jugar al fútbol para ver quién era el mejor jugador. Se fueron a la cancha y comenzaron a jugar. La Gallina era rápida, pero el Cienpiés fue más rápido. La Gallina pateó lejos, pero el Cienpiés pateó más lejos. La Gallina comenzó a enojarse. La Gallina anotó un solo gol en todo el juego. El Cienpiés con sus múltiples patas atrapó muchas pelotas. El Cienpiés anotó cinco goles en total. La Gallina estaba furiosa porque perdió. El Cienpiés se echó a reír. Después del partido la Gallina estaba tan enojada que abrió su pequeño pico y se tragó el Cienpiés de un solo bocado. De camino hacia su casa, la Gallina se encontró con la madre del Cienpiés quien le preguntó por su hijo.	3 ¿Qué se pusieron a jugar la gallina y el cienpiés? ¿Por qué el cienpiés pudo atrapar muchas pelotas? ¿Cuántos goles anotó el cienpiés? ¿Por qué la gallina estaba furiosa? ¿Qué hizo la gallina con el cienpiés?			
	4 Pedro y Mateo son amigos. A ambos les gusta jugar fútbol. Les gusta ir a la escuela, entre otras cosas, a jugar pelota a la hora del recreo. Forman equipos, algunas veces por grado, otras por afinidad. El recreo se siente cortico porque jugar pelota es entretenido. Los niños corren, saltan, patean y gritan. Lo mejor es cuando su equipo mete gol. La mayoría de veces Miguel lleva el balón para jugar el partido. El viernes pasado Miguel no llegó a la escuela porque tenía varicela. Pedro pensó que sería un recreo aburrido, pues no habría partido. Mateo tuvo una idea brillante. Buscaron hojas de papel usadas. Preguntaron al maestro si se las regalaba. Él amablemente accedió. Arrugaron una hoja e hicieron una bola. Luego pusieron más hojas hasta hacer una gran pelota. El maestro los vio y los ayudó. Puso cinta adhesiva a su pelota de papel. La idea del maestro era que la pelota durara por más tiempo, hasta que regresara Miguel.	4 ¿Qué les gusta jugar a Pedro y a Mateo? ¿Quién lleva el balón para jugar el partido? ¿Qué día de la semana faltó Miguel a la escuela? ¿Por qué no fue Miguel a la escuela? ¿Por qué pensó Pedro que el recreo sería aburrido?			
	5 El abuelo tomaba café. Era una tarde lluviosa. Recordaba cuando era niño. El abuelo contó, como era la siembra de café. El vivía en un pueblo. El pueblo era grande. El pueblo se llamaba Neira. Al regresar de la escuela, ayudaba a su papá a sembrar café. Le pregunté: ¿cómo se siembra el café? El abuelo dijo: -el café es una planta. Empieza siendo una semilla. Esta crece y se convierte en cafeto. El cafeto da un fruto rojo llamado cereza. Al madurar, se corta. Luego se seca al sol en grandes patios. Después se tuesta y muele. El café se empaqueta y se vende. Esto es lo que saborea mucha gente, en una deliciosa taza de café. El café es conocido en Colombia. El café es famoso en todo el mundo.	5 ¿Qué se pusieron a jugar la gallina y el cienpiés? ¿Quién pateó más lejos? ¿Por qué el cienpiés pudo atrapar muchas pelotas? ¿Cuántos goles anotó el cienpiés? ¿Por qué la gallina estaba furiosa?			
		6 ¿Cómo estaba la tarde? ¿Cómo era el pueblo? ¿Cómo se llamaba el pueblo? ¿A qué ayudaba el abuelo cuando era niño? ¿De qué color es el fruto que da el Cafeto?			
		7 ¿Qué tomaba el abuelo? ¿Cómo se llamaba el pueblo? ¿A qué ayudaba el abuelo cuando era niño? ¿En qué se convierte la semilla cuando crece? ¿De qué color es el fruto que da el Cafeto? ¿Qué se hace primero: secar el café o tostarlo y molerlo? ¿Quién es el personaje principal de la historia? ¿Crees que esta historia podría suceder en la realidad?			
		8 ¿Qué les gusta jugar a Pedro y a Mateo? ¿Cómo se forman los grupos para jugar fútbol? ¿Cómo se siente el recreo cuando juegan pelotas? ¿En qué momento les gusta jugar fútbol a Pedro y a Mateo? ¿Quién lleva el balón para jugar el partido? ¿Por qué no fue Miguel a la escuela? ¿Por qué pensó Pedro que el recreo sería aburrido? ¿Cuál fue la idea brillante de Mateo?			

Table B.8: IRT of Knowledge of Letter Sounds

Knowledge of letter sounds	Item	Cohort 1		Item	Cohort 2		Item	Cohort 3	
		Difficulty Parameter	Std. Error		Difficulty Parameter	Std. Error		Difficulty Parameter	Std. Error
Item1	M	-1.067	0.032	M	-0.980	0.030	M	-1.371	0.037
Item2	d	-0.552	0.027	d	-0.588	0.025	d	-0.734	0.027
Item3	r	-0.678	0.028	r	-0.728	0.027	r	-1.041	0.031
Item4	O	-1.077	0.032	O	-0.939	0.030	O	-1.142	0.033
Item5	E	-1.039	0.032	E	-0.966	0.030	E	-1.163	0.033
Item6	C	-0.729	0.028	C	-0.764	0.027	F	-0.907	0.029
Item7	i	-1.139	0.033	i	-0.981	0.030	i	-1.175	0.033
Item8	u	-1.050	0.032	u	-0.970	0.030	u	-1.152	0.033
Item9	p	-0.137	0.025	p	-0.242	0.023	p	-0.429	0.025
Item10	S	-0.545	0.027	S	-0.659	0.026	S	-1.132	0.032
Item11	A	-1.077	0.032	A	-0.872	0.029	A	-1.081	0.032
Item12	n	-0.305	0.025	n	-0.362	0.023	n	-0.746	0.027
Item13	j	0.099	0.025	j	-0.035	0.022	j	-0.171	0.024
Item14	T	-0.211	0.025	T	-0.206	0.023	T	-0.527	0.025
Item15	b	0.047	0.025	b	-0.057	0.022	b	-0.202	0.024
Item16	e	-0.795	0.029	e	-0.521	0.025	e	-0.727	0.027
Item17	f	-0.071	0.025	f	-0.155	0.022	f	-0.538	0.025
Item18	r	-0.125	0.025	r	-0.167	0.023	r	-0.444	0.025
Item19	W	0.912	0.030	W	0.600	0.025	G	-0.167	0.024
Item20	L	0.082	0.025	L	-0.014	0.022	L	-0.278	0.024
Item21	m	-0.132	0.025	m	-0.081	0.022	m	-0.426	0.025
Item22	R	0.009	0.025	R	0.010	0.022	R	-0.286	0.024
Item23	D	0.124	0.025	D	0.105	0.022	D	-0.142	0.024
Item24	E	-0.321	0.025	E	-0.013	0.022	E	-0.274	0.024
Item25	y	0.280	0.025	y	0.295	0.023	y	0.237	0.024
Item26	O	-0.234	0.025	O	0.100	0.022	O	-0.161	0.024
Item27	a	-0.186	0.025	a	0.135	0.022	a	-0.129	0.024
Item28	g	0.369	0.026	g	0.415	0.024	g	0.131	0.024
Item29	s	0.273	0.025	s	0.367	0.023	s	0.006	0.024
Item30	Z	0.423	0.026	Z	0.486	0.024	Z	0.178	0.024
Item31	c	0.336	0.025	c	0.503	0.024	f	0.196	0.024
Item32	V	0.657	0.027	V	0.681	0.026	V	0.480	0.026
Item33	N	0.558	0.027	N	0.677	0.026	N	0.336	0.025
Item34	I	0.251	0.025	I	0.600	0.025	I	0.319	0.025
Item35	k	0.715	0.028	k	0.855	0.028	b	0.551	0.026
Item36	U	0.349	0.025	U	0.714	0.026	U	0.415	0.025
Item37	P	0.815	0.029	P	0.953	0.029	P	0.639	0.027
Item38	x	1.037	0.031	x	1.079	0.031	R	0.625	0.027
Item39	L	0.931	0.030	L	1.080	0.031	L	0.699	0.028
Item40	Q	1.101	0.032	Q	1.172	0.033	M	0.701	0.028
Item41	S	0.963	0.030	S	1.127	0.032	S	0.738	0.028
Item42	Ñ	1.193	0.033	Ñ	1.310	0.035	v	0.906	0.030
Item43	O	0.780	0.029	O	1.101	0.031	O	0.842	0.029
Item44	A	0.826	0.029	A	1.137	0.032	A	0.890	0.030
Item45	d	1.219	0.034	d	1.349	0.036	d	1.039	0.031
Item46	T	1.317	0.035	T	1.413	0.037	T	1.120	0.032
Item47	i	0.989	0.031	i	1.286	0.035	i	1.086	0.032
Item48	N	1.394	0.036	N	1.476	0.038	N	1.171	0.033
Item49	a	1.064	0.032	a	1.326	0.035	a	1.213	0.034
Item50	e	1.100	0.032	e	1.359	0.036	e	1.297	0.035
Discrimination Parameter		2.521	0.036		3.709	0.058		3.527	0.055
Cronbach's Alpha	Beginning of Grade 3	0.957			0.963			0.969	
	End of Grade 3	0.965			0.972			0.974	
	End of Grade 4	N/A			0.973			0.976	
	Beginning of Grade 4	N/A			0.966			0.983	

Note: IRT estimates using a one-parameter model.

Table B.9: IRT of Reading of Non-words

Reading of non-words	Item	Cohort 1		Cohort 2		Cohort 3	
		Difficulty Parameter	Std. Error	Difficulty Parameter	Std. Error	Difficulty Parameter	Std. Error
Item1	lete	-2.253	0.059	-2.043	0.053	-2.036	0.055
Item2	quibe	-2.293	0.060	-2.049	0.054	-2.097	0.058
Item3	bofa	-1.778	0.044	-1.648	0.040	-1.723	0.045
Item4	mise	-2.594	0.073	-2.347	0.070	-2.269	0.065
Item5	garo	-1.415	0.037	-1.623	0.039	-1.535	0.040
Item6	cafa	-1.884	0.047	-1.695	0.041	-1.671	0.043
Item7	Celu	-2.620	0.075	-2.217	0.062	-2.140	0.059
Item8	bede	-0.960	0.031	-1.238	0.031	-1.047	0.032
Item9	lura	-2.046	0.052	-1.928	0.049	-1.866	0.049
Item10	mesi	-2.263	0.059	-2.192	0.061	-2.000	0.054
Item11	lluno	-2.158	0.055	-1.950	0.050	-2.041	0.055
Item12	Rite	-2.060	0.052	-1.994	0.051	-1.991	0.053
Item13	duso	-2.021	0.051	-2.021	0.052	-1.881	0.050
Item14	jata	-1.956	0.049	-1.822	0.045	-1.897	0.050
Item15	fica	-2.175	0.056	-2.049	0.054	-2.020	0.055
Item16	luma	-1.821	0.046	-1.756	0.043	-1.825	0.048
Item17	Alti	-2.197	0.057	-1.999	0.052	-1.940	0.052
Item18	lufa	-2.258	0.059	-2.010	0.052	-1.944	0.052
Item19	frate	-1.696	0.043	-1.643	0.040	-1.487	0.039
Item20	dulte	-1.859	0.046	-1.741	0.042	-1.679	0.043
Item21	ledo	-1.946	0.049	-1.811	0.044	-1.760	0.046
Item22	Fosu	-1.492	0.038	-1.602	0.039	-1.422	0.038
Item23	gesa	-1.323	0.036	-1.466	0.035	-1.278	0.035
Item24	lemo	-2.082	0.053	-1.695	0.041	-1.690	0.044
Item25	golpa	-1.778	0.044	-1.566	0.038	-1.524	0.040
Item26	bosa	-1.591	0.040	-1.397	0.034	-1.409	0.037
Item27	rale	-1.503	0.039	-1.342	0.033	-1.248	0.034
Item28	flano	-1.505	0.039	-1.295	0.032	-1.280	0.035
Item29	trabu	-1.407	0.037	-1.198	0.031	-1.201	0.034
Item30	bulo	-1.450	0.038	-1.184	0.030	-1.183	0.033
Item31	pluva	-1.160	0.033	-1.008	0.028	-1.005	0.031
Item32	arcu	-1.314	0.035	-1.031	0.028	-1.069	0.032
Item33	cince	-0.965	0.031	-0.837	0.027	-0.845	0.029
Item34	llusia	-1.049	0.032	-0.843	0.027	-0.835	0.029
Item35	firta	-0.979	0.031	-0.752	0.026	-0.705	0.028
Item36	onti	-0.785	0.029	-0.613	0.025	-0.613	0.028
Item37	zaca	-0.842	0.030	-0.587	0.025	-0.619	0.028
Item38	queno	-0.760	0.029	-0.531	0.025	-0.538	0.027
Item39	bana	-0.654	0.028	-0.394	0.024	-0.425	0.027
Item40	juru	-0.327	0.027	-0.187	0.024	-0.125	0.026
Item41	foba	-0.304	0.027	-0.145	0.024	-0.156	0.026
Item42	lise	-0.353	0.027	-0.136	0.024	-0.153	0.026
Item43	vodo	-0.219	0.027	-0.028	0.024	-0.048	0.026
Item44	tuzi	-0.158	0.027	0.046	0.024	0.016	0.026
Item45	listu	-0.034	0.027	0.141	0.025	0.157	0.027
Item46	quira	0.074	0.027	0.255	0.025	0.269	0.027
Item47	cuto	0.214	0.027	0.358	0.026	0.360	0.027
Item48	ganco	0.347	0.028	0.435	0.026	0.472	0.028
Item49	rafo	0.397	0.028	0.504	0.026	0.538	0.028
Item50	duba	0.491	0.028	0.592	0.027	0.696	0.030
Discrimination Parameter		2.174	0.033	3.386	0.056	2.820	0.046
Cronbach's Alpha	Beginning of Grade 3	0.952		0.958		0.957	
	End of Grade 3	0.943		0.954		0.955	
	Beginning of Grade 4	0.950		0.940		0.953	
	End of Grade 4	0.946		0.951		0.939	

Note: IRT estimates using a one-parameter model.

Table B.10: IRT of Addition

Addition	Item	Cohort 1		Cohort 2		Cohort 3	
		Difficulty Parameter	Std. Error	Difficulty Parameter	Std. Error	Difficulty Parameter	Std. Error
Item1	2+2=(4)	-2.342	0.066	-2.352	0.075	-2.312	0.070
Item2	3+2=(5)	-1.401	0.034	-1.422	0.035	-1.590	0.040
Item3	4+2=(6)	-1.528	0.037	-1.517	0.037	-1.717	0.043
Item4	1+5=(6)	-1.850	0.045	-1.901	0.048	-2.033	0.055
Item5	3+4=(7)	-1.429	0.035	-1.449	0.035	-1.546	0.038
Item6	7+1=(8)	-1.785	0.043	-1.924	0.049	-2.021	0.054
Item7	6+2=(8)	-1.586	0.038	-1.684	0.041	-1.651	0.041
Item8	5+4=(9)	-1.567	0.038	-1.569	0.038	-1.560	0.039
Item9	4+5=(9)	-1.434	0.035	-1.412	0.035	-1.425	0.036
Item10	7+2=(9)	-1.284	0.032	-1.242	0.032	-1.289	0.033
Item11	6+4=(10)	-1.022	0.028	-0.968	0.028	-1.051	0.030
Item12	5+5=(10)	-1.089	0.029	-0.963	0.028	-1.043	0.030
Item13	8+2=(10)	-0.907	0.027	-0.785	0.026	-0.895	0.028
Item14	5+6=(11)	-0.695	0.025	-0.550	0.025	-0.628	0.026
Item15	6+6=(12)	-0.555	0.025	-0.356	0.024	-0.422	0.025
Item16	3+9=(12)	-0.333	0.024	-0.103	0.024	-0.187	0.025
Item17	5+7=(12)	-0.028	0.024	0.186	0.025	0.123	0.025
Item18	8+6=(14)	0.275	0.025	0.601	0.028	0.473	0.027
Item19	10+3=(13)	0.415	0.026	0.724	0.029	0.650	0.029
Item20	2+11=(13)	0.670	0.028	0.932	0.031	0.821	0.030
Item21	13+3=(16)	0.841	0.030	1.147	0.034	0.965	0.032
Item22	6+10=(16)	1.007	0.031	1.330	0.036	1.145	0.034
Item23	10+10=(20)	1.150	0.033	1.444	0.038	1.238	0.035
Item24	15+5=(20)	1.272	0.034	1.598	0.040	1.339	0.037
Item25	11+9=(20)	1.526	0.038	1.758	0.043	1.621	0.041
Discrimination Parameter		3.059	0.049	3.508	0.061	3.566	0.063
Cronbach's Alpha	Beginning of Grade 3	0.914		0.902		0.902	
	End of Grade 3	0.912		0.909		0.913	
	Beginning of Grade 4	0.917		0.886		0.921	
	End of Grade 4	0.908		0.909		0.909	

Note: IRT estimates using a one-parameter model.

Table B.11: IRT of Substraction

Substraction	Item	Cohort 1		Cohort 2		Cohort 3	
		Difficulty Parameter	Std. Error	Difficulty Parameter	Std. Error	Difficulty Parameter	Std. Error
Item1	4-2=(2)	-1.969	0.049	-1.715	0.044	-1.922	0.050
Item2	8-1=(7)	-1.945	0.048	-1.815	0.047	-2.083	0.057
Item3	5-2=(3)	-1.755	0.043	-1.661	0.042	-1.854	0.048
Item4	6-2=(4)	-1.542	0.038	-1.525	0.038	-1.667	0.042
Item5	8-2=(6)	-1.405	0.035	-1.366	0.034	-1.496	0.038
Item6	6-5=(1)	-1.504	0.037	-1.456	0.036	-1.702	0.043
Item7	9-2=(7)	-1.277	0.033	-1.222	0.031	-1.423	0.036
Item8	9-4=(5)	-1.073	0.030	-1.061	0.029	-1.211	0.033
Item9	8-3=(5)	-0.925	0.028	-0.916	0.027	-0.982	0.030
Item10	9-5=(4)	-0.750	0.027	-0.678	0.025	-0.725	0.027
Item11	7-4=(3)	-0.465	0.025	-0.386	0.024	-0.476	0.026
Item12	10-2=(8)	-0.300	0.024	-0.172	0.024	-0.224	0.025
Item13	10-3=(7)	-0.025	0.024	0.102	0.024	0.051	0.026
Item14	10-4=(6)	0.186	0.025	0.333	0.026	0.317	0.027
Item15	20-10=(10)	0.369	0.025	0.539	0.027	0.545	0.028
Item16	11-6=(5)	0.743	0.028	0.904	0.030	0.887	0.031
Item17	11-7=(4)	1.048	0.031	1.213	0.034	1.258	0.035
Item18	12-9=(3)	1.246	0.033	1.556	0.039	1.553	0.040
Item19	12-7=(5)	1.555	0.038	1.838	0.045	1.846	0.046
Item20	12-6=(6)	1.775	0.043	2.066	0.050	2.083	0.053
Item21	13-11=(2)	1.999	0.048	2.309	0.058	2.270	0.059
Item22	14-6=(8)	2.210	0.054	2.516	0.065	2.547	0.071
Item23	16-3=(13)	2.335	0.058	2.675	0.071	2.751	0.081
Item24	16-10=(6)	2.504	0.064	2.977	0.085	2.830	0.086
Item25	20-5=(15)	2.695	0.072	3.179	0.096	2.919	0.091
Item26	20-4=(16)	2.761	0.075	3.250	0.100	2.939	0.092
Item27	20-9=(11)	2.861	0.080	3.447	0.113	3.067	0.100
Discrimination Parameter		2.890	0.045	3.710	0.066	3.255	0.057
Cronbach's Alpha	Beginning of Grade 3	0.900		0.864		0.882	
	End of Grade 3	0.895		0.891		0.884	
	Beginning of Grade 4	0.913		0.881		0.915	
	End of Grade 4	0.902		0.902		0.897	

Note: IRT estimates using a one-parameter model.