Working Paper

Addendum to Effectiveness of Cognitive Tutor Algebra I at Scale

John F. Pane, Beth Ann Griffin, Daniel F. McCaffrey, Rita Karam

RAND Education

WR-1050-DEIES
May 2014
Prepared for Institute of Education Sciences, U.S. Department of Education

Addendum to Effectiveness of Cognitive Tutor Algebra I at Scale

John F. Pane, RAND Corporation

Beth Ann Griffin, RAND Corporation

Daniel F. McCaffrey, Educational Testing Service

Rita Karam, RAND Corporation

Introduction

The authors of this working paper previously reported the results of large-scale effectiveness studies of Cognitive Tutor Algebra I (CTAI) in middle schools and high schools (Pane et al., 2013; Pane et al., 2014). We found no effects of CTAI in the first year of implementation (Cohort 1), but a statistically significant positive effect in the second year of the high school study (Cohort 2).

As part of the analysis previously reported, we used regression calibration (Carroll et al., 2006) to correct for measurement error in the pretests. The motivation for this technique was the apparent purposive selection of students for classes participating in the study.[1] Including error-prone covariates in a regression model to control for systematic group differences without a correction for measurement error will generally result in biased estimates of all model

[1] Deliberate research design decisions, to enable schools to implement CTAI in an authentic fashion, may have enabled schools to shape the population of students participating in the study after becoming aware of treatment assignment (see Pane et al., 2013; 2014 for details).

parameters including the treatment indicator (Greene, 2003; Lockwood and McCaffrey, 2014). We implemented regression calibration by replacing each error-prone test score with a random variable drawn from the conditional distribution of the corresponding error-free test score given the error-prone test score, all other model covariates including classroom indicators, and the outcome. The required conditional distribution was constructed assuming a linear relationship between the error-free scores and these variables and used the conditional standard error of measure of each test score. To support standard error estimation, we imputed twenty values of the error-free test score for each student. An additional benefit of this approach was that we were able to impute missing pretest values for approximately 18% of the high school sample and 7% of the middle school sample – students who joined the study after the administration of the pretest.

The purpose of this working paper is to supplement the analyses previously reported to address concerns raised by some critics regarding the use of imputed pretest scores and the inclusion of study joiners in the analysis. In response to these concerns, we provide alternate analyses that exclude study joiners and use the actual pretest scores as covariate controls in statistical models instead of the imputed scores with measurement error corrections.

This working paper summarizes the results of this alternative analysis. Model 1 estimates group differences without any covariates; Model 2 includes pretest scores; and Model 3 also includes additional student covariates. Model 3 is our preferred specification, and the others are included here to show the sensitivity of estimates to covariate adjustment. These models are specified the same as Models 1 through 3 in the previous articles, except for the substitution of actual pretest

scores for the imputed scores. For convenience, the Appendix reproduces the treatment effect estimates for Models 1 through 3, as reported in the previous articles. The sample is also redefined to exclude joiners. Other details remain as previously reported. We refer the reader to Pane et al. (2014) for more complete details about the intervention, study design, analytic methods, and results not reported here.

Results of Supplemental Analysis

The exclusion of study joiners changes the calculation of the eligible sample and increases attrition rates from those previously reported. However, group differences and p-values are unchanged because the original assessment of balance used the original, unadjusted pretest scores and students lacking scores had no effect on those calculations. Table 1 summarizes information about student participants in the high school sample. The final sample included 11,066 students across the two cohorts. The final sample for Cohort 1 contained 5,328 students after 36.4% attrition (5.8% differential attrition), and the treatment group scored .14 standard deviation units lower than the control group on the pretest (p=.19). The final sample for Cohort 2 contained 5,738 students after 27.7% attrition (5.0% differential attrition), and the treatment group scored 0.11 lower than the control group on the pretest (p=.28).

Table 1: High School Study Attrition and Group Balance

		Treatment group		Control group		Group difference^	p-value	
		N	*Pretest mean*	*N*	*Pretest mean*			
Cohort 1	Eligible sample	3,986	-0.468	4,395	-0.347	-0.194	0.033	*
	Attrition	1,330	-0.673	1,723	-0.485	-0.155	0.091	
	Final sample	2,656	-0.365	2,672	-0.258	-0.144	0.187	
	Attrition rate		33.4%		39.2%			
Cohort 2	Eligible sample	3,474	-0.390	4,457	-0.359	-0.099	0.284	
	Attrition	1,058	-0.591	1,135	-0.619	-0.067	0.529	
	Final sample	2,416	-0.302	3,322	-0.270	-0.111	0.276	
	Attrition rate		30.5%		25.5%			
Both cohorts	Eligible sample	7,460	-0.432	8,852	-0.353	-0.139	0.082	
	Attrition	2,388	-0.637	2,858	-0.538	-0.106	0.205	
	Final sample	5,072	-0.335	5,994	-0.265	-0.116	0.188	
	Attrition rate		32.0%		32.2%			

Notes: ^ Standardized mean difference in IRT pretest scores between treatment and control groups (negative indicates treatment scored lower than control) as calculated by a model that includes fixed effects for randomization pairs and random effects for classrooms within schools. Eligible sample is defined as students present at pretest. Attrition is defined as the portion of the eligible sample that did not take the posttest.

Table 2 summarizes information about student participants in the middle school sample. The final sample included 5,519 students across the two cohorts. The final sample for Cohort 1 contained 2,765 students after 12.1% attrition (2.3% differential attrition), and the treatment group scored .31 standard deviation units lower than the control group on the pretest (p=.02). The final sample for Cohort 2 contained 2,754 students after 14.4% attrition (5.6% differential attrition), and the treatment group scored 0.35 lower than the control group on the pretest (p=.02).

Table 2: Middle School Study Attrition and Group Balance

		Treatment group		Control group		Group difference^	p-value	
		N	Pretest mean	N	Pretest mean			
Cohort 1	Eligible sample	1,544	0.265	1,602	0.654	-0.296	0.021	*
	Attrition	169	-0.067	212	0.318	-0.295	0.135	
	Final sample	1,375	0.306	1,390	0.706	-0.312	0.016	*
	Attrition rate		10.9%		13.2%			
Cohort 2	Eligible sample	1,488	0.400	1,731	0.738	-0.422	0.003	**
	Attrition	170	0.015	295	0.733	-0.392	0.012	*
	Final sample	1,318	0.449	1,436	0.739	-0.347	0.016	*
	Attrition rate		11.4%		17.0%			
Both cohorts	Eligible sample	3,032	0.331	3,333	0.698	-0.366	0.003	**
	Attrition	339	-0.026	507	0.560	-0.385	0.007	**
	Final sample	2,693	0.376	2,826	0.722	-0.328	0.007	**
	Attrition rate		11.2%		15.2%			

Notes: ^Standardized mean difference in IRT pretest scores between treatment and control groups (negative indicates treatment scored lower than control) as calculated by a model that includes fixed effects for randomization pairs and random effects for classrooms within schools. Eligible sample is defined as students present at pretest. Attrition is defined as the portion of the eligible sample that did not take the posttest.

Table 3 summarizes results from the high school study. Models consistently estimate negative treatment effects for cohort 1, ranging from -.07 to -.17 standard deviation units and not significant. Models for Cohort 2 consistently estimate positive treatment effects, ranging from .12 to .21 standard deviation units; the result for Model 3 with a full set of covariate controls is significant at the .05 level.

Table 3: High School Study Treatment Effect Estimates

Model	Cohort 1				Cohort 2				
	Estimate	Std. Error	t-value	p-value	Estimate	Std. Error	t-value	p-value	
1	-0.17	0.11	-1.53	0.14	0.12	0.13	0.91	0.37	
2	-0.12	0.10	-1.19	0.24	0.17	0.10	1.66	0.11	
3	-0.07	0.10	-0.75	0.46	0.21	0.10	2.14	0.04	*

Notes: Model 1 estimates group differences without any covariates; Model 2 includes pretest scores; and Model 3 also includes additional student covariates.

Table 4 summarizes the results for the middle school study. Here again, models estimate negative effects for Cohort 1 that are not significant, ranging from -.07 to -.20. For Cohort 2, the

unadjusted estimate is negative and become positive with covariate adjustment. None of the middle school estimates are significant.

Table 4: Middle School Study Treatment Effect Estimates

Model	*Cohort 1*				*Cohort 2*			
	Estimate	*Std. Error*	*t-value*	*p-value*	*Estimate*	*Std. Error*	*t-value*	*p-value*
1	-0.20	0.15	-1.32	0.19	-0.06	0.17	-0.38	0.71
2	-0.07	0.12	-0.62	0.54	0.09	0.14	0.66	0.52
3	-0.09	0.12	-0.76	0.45	0.10	0.14	0.72	0.47

Note: Model 1 estimates group differences without any covariates; Model 2 includes pretest scores; and Model 3 also includes additional student covariates.

Discussion

These alternative analyses produce results that are substantively the same as previously reported. We find a significant positive effect of .21 standard deviation units for high school students in the second year of the study, virtually identical to the results previously reported. In the middle school study, where pretreatment group differences were greater, covariate adjustment does not result in positive estimates as large as previously reported. This suggests that bias may remain in the middle school estimates reported herein due to the lack of correction for measurement error.

Appendix

Table A.1: Previously Reported High School Study Treatment Effect Estimates

Model	*Cohort 1*				*Cohort 2*				
	Estimate	*Std. Error*	*t-value*	*p-value*	*Estimate*	*Std. Error*	*t-value*	*p-value*	
1	-0.19	0.12	-1.68	0.10	0.14	0.12	1.20	0.24	
2	-0.12	0.10	-1.20	0.24	0.19	0.09	2.05	0.05^	*
3	-0.10	0.10	-0.97	0.34	0.22	0.09	2.33	0.03	*

Notes: ^ value is less than 0.05 before rounding. Model 1 estimates group differences without any covariates; Model 2 includes regression-calibrated pretest scores; Model 3 also includes additional student covariates. (Sources: Pane et al., 2013; Pane et al., 2014)

Table A.2: Previously Reported Middle School Study Treatment Effect Estimates

Model	*Cohort 1*				*Cohort 2*			
	Estimate	*Std. Error*	*t-value*	*p-value*	*Estimate*	*Std. Error*	*t-value*	*p-value*
1	-0.20	0.15	-1.34	0.19	-0.07	0.17	-0.41	0.69
2	0.00	0.10	0.01	0.99	0.17	0.13	1.30	0.21
3	-0.03	0.11	-0.24	0.81	0.19	0.13	1.44	0.16

Note: Model 1 estimates group differences without any covariates; Model 2 includes regression-calibrated pretest scores; Model 3 also includes additional student covariates. (Sources: Pane et al., 2013; Pane et al., 2014)

References

Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective* (Second ed.). London: Chapman and Hall.

Greene, W. H. (2003). *Econometric Analysis* (Fifth ed.). Upper Saddle River, NJ: Prentice Hall.

Lockwood, J. R., & McCaffrey, D. F. (2014). Correcting for Test Score Measurement Error in ANCOVA Models for Estimating Treatment Effects. *Journal of Educational and Behavioral Statistics*, *39*(1), 22-52.

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2013). Effectiveness of Cognitive Tutor Algebra I at Scale *Working Paper WR-984-DEIES*. Santa Monica, CA: RAND Corporation.

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of Cognitive Tutor Algebra I at Scale. *Educational Evaluation and Policy Analysis, 36*(2), 127-144. doi: 10.3102/0162373713507480