Volume 9, Issue 1, July 2015

Achieve3000® impacts on student reading and STAAR EOC English I, English II, and Biology exams for the 2014 2015 academic year.

By D. Diego Torres, Ph.D.

Based on the Lexile® Framework, a scientific approach to reading and text measurement backed by more than two decades of ongoing research, Achieve3000®, a web-based differentiated reading program used by HISD, is designed to improve student reading ability and comprehension of increasingly complex texts by initially meeting students where they are academically. The present study focuses on ninth and tenth graders and assesses the impact deriving from the

[redacted]

[redacted] all ninth graders in the district on the STAAR EOC English I exam. The English I mean scale score among all ninth-grade students was significantly greater than the English I mean among Achieve3000 ninth grade students who completed one to five activities ($p < .001$).

Figure 1 also shows mean STAAR EOC Biology scale scores by student group and treatment status for all ninth-grade students in HISD and for ninth-grade students in schools that had licenses for Achieve3000. The mean Biology scale score among all HISD ninth graders was not significantly different than the mean scale score among any of the two bottom Achieve3000 treatment groups. The mean biology scale score among Achieve3000 students who completed more than ten activities was significantly higher than that achieved by Achieve3000 students who completed one to five activities, by 70 points ($p < .05$). There were no other statistically significant mean group differences on the STAAR EOC Biology exam.

The mean STAAR EOC English II scale score, as shown in **Figure 2**, was 3757 among all tenth graders in the district. This mean was not significantly different than that among Achieve3000 students who completed zero (3749) or six to ten activities (3691). Achieve3000 students who completed one to five activities had a mean about 66 points lower than that experienced by all tenth graders in the district (statistically significant at the $p < .001$), while those who completed more than ten activities had a mean of 163 points higher (statistically significant at the $p < .001$) than that experienced by all tenth graders in the district.

Among Achieve3000 students, those who completed more than ten activities performed significantly better than those who completed six to ten activities, who, in turn performed better than those who completed one to five activities.

Similar to the trends just highlighted with respect to the STAAR EOC exams, it was evident that more activities completed was associated with a greater likelihood of having met the passing standard o lET6

cias.(n)6( )] TJETBT/F4 9.96 Tf1 0362.5915.01 332.93 Tm[(F)-2(ig)-4(u2

activities, 7 greater rpr[(ac)orn of nih95(in)7( ) an-(ten)5( )m[(te-10(th

[redacted]

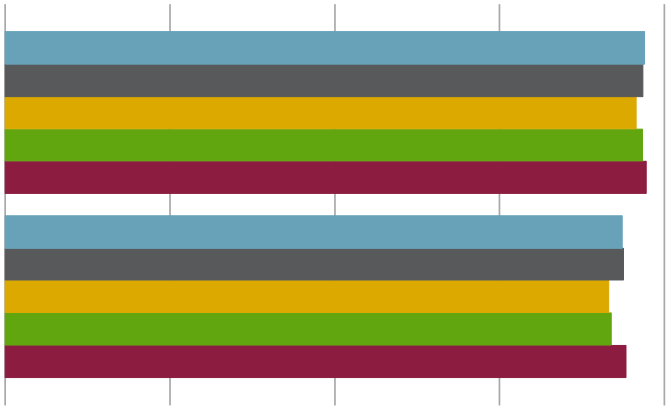

Figure 1. Mean 9th Grade STAAR EOC Scale Score by Treatment Status, Biology and English I.

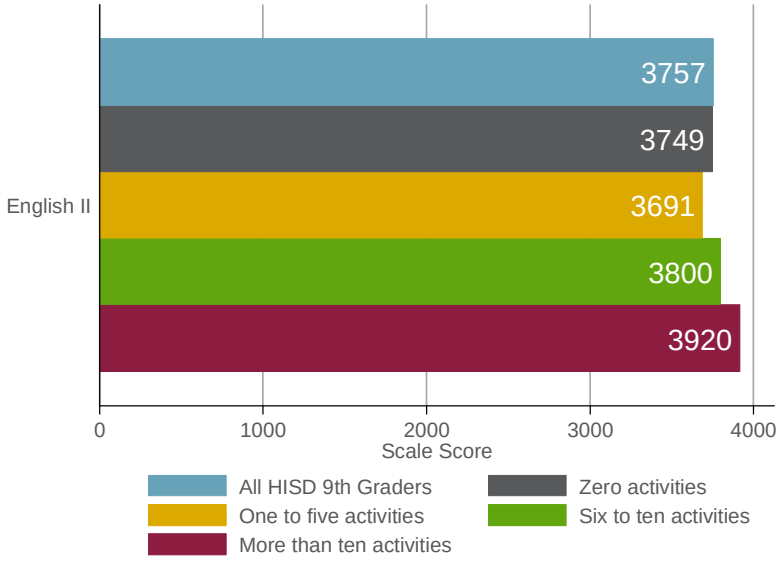

Figure 2. Mean 10th Grade STAAR EOC Scale Score by Treatment Status, English II.

Figure 3. STAAR EOC Phase-in 1 Did-Not-Meet/Did-Meet Standard Rates by Treatment Status by Subject.

Turning to the STAAR EOC English II exam (the third column in Tables B2 and B3), the potential mean probability of meeting the passing standard among tenth graders with access to Achieve3000 would be about 50 percent at the Phase-in 1 level and 44 percent at the Phase-in 2 level. Similar to the average treatment effect shown for the English I exam, had all students completed only between one and five or between six and ten Achieve3000 activities, the mean probability of meeting the passing standard would be the same, regardless of Phase-in level. Had all tenth graders with access to Achieve3000 completed more than ten activities, the mean probability of meeting the passing standard would be more than 13 percentage points higher at the Phase-in 1 level and more than 10 percentage points higher at the Phase-in 2 levels. Though somewhat attenuated, the same trend of treatment effects are achieved when tenth-grade students' English I scores from ninth grade are controlled (the fourth column of Tables B2 and B3). Had all tenth graders with access to Achieve3000 completed more than ten activities, the mean probability of meeting the passing standard would be more than 9 percentage points higher at both the Phase-in 1 and by more than almost 7 percentage points at the Phase-in 2 level.

Had all students failed to complete at least one Achieve3000 activity, their mean probability of meeting the passing standard for the STAAR EOC biology exam would be about 71 percent at the Phase-in 1 level and about 59 percent at the Phase-in 2 level (see the final columns of Tables B2 and B3). The average treatment effect increases significantly with the completion of more exercises. If all students with access to Achieve3000 had completed between one and five of the web-based activities, the mean probability of meeting the passing standard on the STAAR EOC biology exam would be 7 percentage points higher (significant at the $p < .01$ level) at both Phase-in levels than the mean probability had none of them completed a single activity. Had all students completed between six and ten activities, the mean probability of meeting the passing standard would be about 12 percentage points higher (significant at the $p < .001$ level) at both Phase-in levels than the mean probability had none of them completed a single activity. Had all students completed more than ten activities, the mean probability of meeting the passing standard would be about 17 percentage points higher ($p < .001$ level) at the Phase-in 1 level and about 21 percentage points higher ($p < .001$ level) at the Phase-in 2 level than the mean probability had none of them completed a single activity.

Finally, for those students who had pre-treatment Lexile scores on LevelSet and completed at least one Achieve3000 activity, how large were their collective gains on the post-treatment Lexile score?

of the findings for the other outcomes examined here, that Achieve3000 has some positive impact.

Based on these findings, it is recommended that teachers in schools that have Achieve3000 licenses actively encourage their students to complete as many of the exercises as possible during the academic school year. Fidelity to such a recommendation under such a highly decentralized system will no doubt be difficult to achieve. It may therefore be important to allow for the incorporation of Achieve3000 reading solutions into normal classroom time or offer incentives to students who complete a greater number of exercises.

References

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications.* New York: Cambridge University Press.

DeWalt, D. A., Berkman, N. D., Sheridan, S., Lohr, K. N., & Pignone, M. P. (2004). Literacy and health outcomes. *Journal of General Internal Medicine*, *19*, 1228-1239.

DeWalt, D. A., & Hink, A. (2009). Health literacy and child health outcomes: A systematic review of the literature. *Pediatrics*, *124*, S265-S274.

Greenberg, D., Gilbert, A. & Fredrick, L. (2006). Reading interest and behavior in middle school students in inner-city and rural settings. *Reading Horizons*, *47*, 159-173.

National Center for Education Statistics. (2013). *A first look: 2013 mathematics and reading*. Retrieved from http://nces.ed.gov/nationsreportcard/subject/publications/main2013/pdf/2014451.pdf.

Pitcher, S. M., Albright, L. K., DeLaney, C. J., Walker, N. T., Seunarinesingh, K., Mogge, S., …Dunston, P. J. (2007). Assessing adol si g a

Appendix A

This appendix provides information on the analytic strategy used in this study.

Analytic Strategy

Treatment of STAAR EOC Scale Score Outcomes

Because neither the implementation of Achieve3000 at a specific school nor its use by particular students within those schools are random processes, statistical analyses examining its relationship to specific outcomes must address the non-representative nature deriving from these non-random processes. While simple regression techniques may reveal statistically significant associations between the use of Achieve3000 and students' achievement, such associations may not be viewed as causal since there may be unobserved differences among students that drives them to both do better academically, generally, and to also complete more Achieve3000 activities. To be able to say that Achieve3000 produced specific results and was not merely associated with them, the counterfactual model of causal inference requires the use of statistical methods that remove bias. Failure to account for potential omitted variables or the bias in selecting greater use of Achieve3000 among students can lead to erroneous conclusions about the causal link between students' use of Achieve3000 and their

Appendix B

Table B1. Treatment level predicted scale score outcomes deriving from the inverse-probability-weighted regression-adjusted estimator.

			Controlling for English I Scores	
ATE				
Treatment				
(1 to 5 vs. 0)	5.69	-25.66	-3.54	33.06
	(19.54)	(15.08)	(11.89)	(20.49)
(6 to 10 vs. 0)	11.17	24.96	20.48	68.43**
	(24.19)	(17.61)	(13.59)	(24.64)
(More than 10 vs. 0)	93.35***	89.36***	44.40**	95.81***
	(25.33)	(18.62)	(13.69)	(24.79)
POMean				
Treatment				
0	3661.64***	3758.33***	3758.98***	3801.08***
	(18.03)	(13.82)	(11.51)	(17.48)

Note: All model coefficients are net of all controls listed in the Data and Method section of this research brief. Robustv(f)29(.78T1(-)18(p)-20(e)-4(t-26(st)30(e)-4(d)-4( )29(i)-2.e)-4(d)-4)30(l)35(i)3

Table B2. Treatment level predicted probabilities of meeting the Phase-in 1 passing standard deriving from the inverse-probability-weighted regression-adjusted estimator.

	English I	English II		Biology
			Controlling for English I Met Standard	
ATE				
Treatment				
(1 to 5 vs. 0)	.020	.012	.008	.068**
	(.021)	(.019)	(.017)	(.022)
(6 to 10 vs. 0)	.030	.018	.005	.117***
	(.030)	(.022)	(.020)	(.029)
(More than 10 vs. 0)	.109***	.126***	.093***	.167***
	(.032)	(.023)	(.021)	(.026)
POMean				
Treatment				
0	.405***	.495***	.507***	.708***
	(.019)	(.016)	(.017)	(.020)

Note: All model coefficients are net of all controls listed in the Data and Method section of this research brief. Robust standard errors are in parenthesis.
*p < .05, **p < .01, ***p < .001; two-tailed tests.

			Controlling for English I Met Standard	
ATE				
Treatment				
(1 to 5 vs. 0)	.013	-.005	.005	.072**
	(.021)	(.018)	(.017)	(.023)
(6 to 10 vs. 0)	.031	.001		

Table B4. Random coefficient models of Lexile score change due to Achieve3000 usage.

Variables	Restricted Model	Full Model
Fixed Effects		
Intercept	842.63***	892.48***
	(33.06)	(34.13)
# Activities Completed	19.16	15.78***
	(16.95)	(3.08)
Random Effects		
SD of the Intercept	161.58***	19.59***
	(24.26)	(3.79)
SD of # Activities Completed	52.21***	7.05**
	(21.92)	(3.48)
Correlation between Intercept and # Activities Completed	.46	.77
SD of the Residuals	240.50***	63.23***
	(4.29)	(1.12)
Log-likelihood	-11303.35	-9101.62

Note: The restricted model coefficients are net of only the number of activities covariate, whose effect was also allowed to vary by school. The full model coefficients are net of all controls listed in the Data and Method section of this research brief, including the pre-treatment Lexile score. Again, the effect of the number of activities completed was allowed to vary by school. Maximum likelihood estimation was used to obtain the estimates. Standard errors are in parenthesis.
*$p < .05$, **$p < .01$, ***$p < .001$; two-tailed tests.