

**MEMORANDUM**

November 3, 2014

TO: Board Members

FROM: Terry B. Grier, Ed.D.  
Superintendent of Schools

SUBJECT: **A STUDY OF THE IMPACT OF ABYDOS ON THE WRITING PERFORMANCE OF HISD FOURTH AND SEVENTH GRADE STUDENTS, 2013–2014**

CONTACT: Carla Stevens, 713-556-6700

**Background**

Abydos is a professional development program that includes a three-week writing institute. It provides school districts with teachers who, after a year of implementation and another year of professional development, can train other teachers, thus providing ongoing staff development. The Houston Independent School District (HISD) implemented Abydos during 2013–2014 academic year. The Department Research & Accountability produced an evaluation of the impact of Abydos on fourth and seventh grade students' writing performance on the 2014 STAAR writing test.

This was the first year of the program. A total of 261 teachers were trained impacting 4,374 students. The most notable findings of this evaluation were: a) Abydos students obtained a higher mean scale score than non-Abydos students on the 2014 STAAR writing subtest. However, the mean scale score differences between the groups were not statistically significant with effect size ( $d < 0.15$ ), which was negligible; b) the percentage of Abydos students who met the 2014 STAAR Level II: Satisfactory (Phase-In 1) writing standard was higher than their non-Abydos peers. The percentage differences between the Abydos and non-Abydos groups were not statistically significant with effect size ( $d < 0.15$ ), which was negligible.

**Administrative Response**

Upon review of the findings, further implementation and investigation are warranted. Future research should include additional comparisons such as teacher interviews (as indicated in the report) to gauge fidelity of implementation, results of standardized reading test results for all applicable grades, review of student writing portfolios, and long-term review of writing scores (over time). While these initial results fail to demonstrate significant effect, continued implementation of the Abydos writing pedagogy should provide increased achievement in reading and writing.

Should you have any questions or require any further information, please contact me or Carla Stevens in the Department of Research and Accountability, at 713-556-6700.



TBG

TBG/CS:lp

cc: Daniel Gohl  
Shonda Huery  
Lance Menster  
Cindy Puryear  
Karen Hill



# RESEARCH

Educational Program Report

**A STUDY OF THE IMPACT OF ABYDOS ON THE WRITING  
PERFORMANCE OF HISD FOURTH AND SEVENTH GRADE  
STUDENTS, 2013-2014**



## 2014 BOARD OF EDUCATION

**Juliet Stipeche**

President

**Rhonda Skillern-Jones**

First Vice President

**Manuel Rodriguez, Jr.**

Second Vice President

**Anna Eastman**

Secretary

**Wanda Adams**

Assistant Secretary

**Michael L. Lunceford**

**Paula Harris**

**Greg Meyers**

**Harvin C. Moore**

**Terry B. Grier, Ed.D.**

Superintendent of Schools

**Carla Stevens**

Assistant Superintendent

Department of Research and Accountability

**Lai Pei , Ph.D.**

Research Specialist

**Venita Holmes, Dr.P.H.**

Research Manager

**Houston Independent School District**  
Hattie Mae White Educational Support Center  
4400 West 18th Street Houston, Texas 77092-8501

[www.HoustonISD.org](http://www.HoustonISD.org)

It is the policy of the Houston Independent School District not to discriminate on the basis of age, color, handicap or disability, ancestry, national origin, marital status, race, religion, sex, veteran status, or political affiliation in its educational or employment programs and activities.

# A STUDY OF THE IMPACT OF ABYDOS ON THE WRITING PERFORMANCE OF HISD FOURTH AND SEVENTH GRADE STUDENTS, 2013–2014

## Executive Summary

### Program Description

Abydos is a professional development program that includes a three-week writing institute. It provides school districts with teachers who, after a year of implementation and another year of professional development, can train other teachers, thus providing ongoing staff development (Carroll & Wilson, 2009). Carroll and Wilson (2009) state the vision of Abydos is to train the teachers to achieve the following learning objectives:

- demonstrate the teaching of writing as a process;
- teach language arts (support) skills within the writing process according to students' needs and state curricular guidelines;
- write and share with students;
- create a positive, non-threatening environment that encourages learning, participation, and risk-taking;
- create a student-centered classroom;
- teach students how to address a variety of audiences and write for many purposes in many modes;
- understand the theory that supports the writing process; and
- use reading to teach writing and writing to teach reading.

The Houston Independent School District (HISD) implemented Abydos during 2013–2014 academic year. The purpose of this report was to examine the impact of Abydos on fourth and seventh grade students' writing performance on the 2014 STAAR writing subtest based on the level II, phase-in 1 standard. The following research questions were addressed in this evaluation report:

1. How did students whose teachers completed Abydos training perform on the 2014 STAAR writing subtest compared to their grade-level peers whose teachers did not attend Abydos training?
2. Did the impact of Abydos on students' writing performance vary by student group?

### Highlights

- The matched Abydos and non-Abydos students were similar in terms of demographic characteristics and prior reading and language performance on the 2013 Stanford reading and language subtests.

- Abydos students obtained a higher mean scale score than non-Abydos students on the 2014 STAAR writing subtest. However, the mean scale score differences between the groups were not statistically significant with effect size ( $d < 0.15$ ), which was negligible.
- The percentage of Abydos students who met the 2014 STAAR Level II: Satisfactory (Phase-In 1) writing standard was higher than their non-Abydos peers. The percentage differences between the Abydos and non-Abydos groups were not statistically significant with effect size ( $d < 0.15$ ), which was negligible.

### Recommendations

- Based on students' outcome data, there is no strong evidence showing Abydos training had a significant positive impact on students' performance. The Curriculum, Instruction, and Assessment Department may collect the data regarding teacher implementation of Abydos writing strategies in the classroom to explore the influence of Abydos training on teacher's instructional practices.
- Future evaluations should also include teacher interviews to find out whether Abydos training fosters changes in teachers' attitudes and their implementation of instructional model for improving writing strategies in the classroom.

### Administrative Response

Upon review of the findings, further implementation and investigation are warranted. Future research should include additional comparisons such as teacher interviews (as indicated in the report) to gauge fidelity of implementation, results of standardized reading test results for all applicable grades, review of student writing portfolios, and long-term review of writing scores (over time). While these initial results fail to demonstrate significant effect, continued implementation of the Abydos writing pedagogy should provide increased achievement in reading and writing.

## Introduction

Both researchers and theorists have concurred that a student's ability to communicate ideas through writing has been designated as a top indicator of future academic success (Friedman, 2006; National Commission on Writing, 2003). However, a NAEP writing assessment report (2003) documented that 58% of fourth graders and 54% of eighth graders were writing at the basic level. The basic level was described as lacking attention to audience and elaboration that clarifies and enhances the central idea (National Center for Education Statistics (NCES), 2003). In addition, writers at the basic level or below were not writing well enough to meet the demands faced in higher education and the work environment (NCES, 2003). Hillocks (2005) stated in a meta-analysis that teachers with a negative attitude toward writing chose formulaic strategies to teach instead of implementing authentic writing instruction based on the writing process. However, research has shown that teachers must provide authentic writing instruction in order to increase students' achievement levels (Bloodgood, 2002). Research on teacher development has documented that professional development is a possible venue to enhance the attitude and effectiveness of teachers (Allington, 2005; Darling-Hammond, 1998). Guskey (1985) found that the success of students' learning outcomes started from changes in teachers' attitude toward the subject matter and teachers' implementation of strategies learned in professional development. In order to meet the teachers' needs for effective classroom writing instruction, HISD provided a three-week writing institute, Abydos, to teachers during the summer of 2013.

## Methods

### Data Collection and Analysis

#### Measure

Student writing performance data were collected from two test assessments: STAAR writing test and Stanford Achievement Test (Stanford 10) reading and language subtests.

- The Stanford 10 assesses students' academic achievement in various academic subjects across 9 grade levels (kindergarten through grade 8). In order to compare scores from different administrations and from different instruments, the Normal Curve Equivalents (NCEs) were used for all subtests in this evaluation. Students' total NCE score on the 2013 Stanford reading and language subtests was used to measure their prior reading and language performance in this evaluation.
- STAAR is the state of Texas criterion-referenced assessment, and it replaced the Texas Assessment of Knowledge and Skills (TAKS) program in spring 2012. The Texas Education Agency (TEA), in collaboration with the Texas Higher Education Coordinating Board (THECB) and Texas educators, developed this new assessment system in response to requirements set forth by the 80th, 81st and 83rd Texas legislatures. This new system focuses on increasing postsecondary readiness of graduating high school students, and helps to ensure that Texas students are competitive with other students both nationally and internationally. Students' performance on the STAAR writing test was used as the outcome measure of the Abydos training effect on students' writing performance. The key outcome measure for this evaluation is the 2014 STAAR writing scale scores of fourth and seventh grade students. The 2014 STAAR Level II: Satisfactory (Phase-in I) writing performance standard was also used to measure the proportion of students who met the writing standard.

## Data Analyses

This evaluation combined three analytic approaches to examine the impact of Abydos on student writing performance. First, propensity score matching was used to reduce the selection bias of students by creating a comparable treatment group and control group of students. Second, descriptive statistics (mean scale scores and percentages of students who met STAAR Level II: Satisfactory (Phase-in I) standard) were used to describe the impact of Abydos training on all students and on student subgroups. Third, ANCOVA was used to investigate the association between the treatment effect of Abydos and students' writing performance by controlling students' prior reading and language performance on the Stanford 10 test. The detail of aforementioned analytic procedures is discussed below.

- Propensity Score Matching: Quasi-experimental design was used in this evaluation, which includes a pre-post test design with a treatment group and a control group. The Abydos students (treatment group) was comprised of students whose teachers enrolled and completed the three-week Abydos training, while the non-Abydos students (control group) were students whose teachers have never enrolled in the Abydos training workshop. Propensity score matching was used to select a group of Abydos students that matched the non-Abydos group as much as possible in term of the observable characteristics. The statistical package *MatchIt* in R was used in this evaluation to conduct propensity score matching based on students' grade, gender, ethnicity, economically-disadvantaged status, special education placement, LEP and at-risk status.
- Analysis of Covariance (ANCOVA): To ensure an accurate and fair assessment of the Abydos impact on students' writing knowledge, ANCOVA was used in subsequent analyses to adjust for students' differences in prior reading and language performance on the Stanford 10 test when comparing Abydos and non-Abydos students' writing performance on the 2014 STAAR writing subtest. ANCOVA is a widely accepted statistical procedure that has been used in other quasi-experimental studies (Field, 2013; Wills & Stommel, 2002).
- Effect Size Analysis: Effect size was used to quantify the size of the performance difference between treatment and control group students. Borman and D'Agostino (1996) suggested that the average effect size associated with Title I programs is  $d = 0.15$ . Kulik, Kulik, and Bangert (1984), suggested that the average effect size in achievement test scores is 0.32. Therefore, we used  $d = 0.15$  as small-modest,  $d = 0.3$  as modest-large, and  $d = 0.5$  as large in this evaluation.

## Sample

All HISD elementary and secondary teachers were invited to participate the Abydos training in the summer of 2013. There were 261 teachers enrolled in the Abydos training workshop, and the completion rate was 75.5%. Student demographic data were extracted from district's Chancery database on September 5, 2014. Student performance data were collected from the 2013 Stanford Achievement Test (Stanford 10) and 2014 State of Texas Assessments of Academic Readiness (STAAR) writing test. Only fourth and seventh grade students were included in this evaluation because the STAAR writing test was only administrated in fourth and seventh grade. **Appendix A-Table 1** (p. 13) shows that the demographic information for Abydos and non-Abydos students in the study sample was not similar with respect to grade, ethnicity, economically-disadvantaged status, LEP status and at-risk status. As a result, propensity score matching was used to match Abydos and non-Abydos students in the study sample to create a comparable Abydos and non-Abydos group with respect to their demographic information. Only students who had both 2013 Stanford reading and language scores and 2014 STAAR writing scores were included in this evaluation. Consequently, the sample size of the matched sample was 4,374 students in the

Abydos group, and 4,374 students in the non-Abydos group. The demographic characteristic of students in the matched samples is shown in Appendix A-Table 1 (p. 13).

### Data Limitations

- There are other literacy initiatives being implemented in the district, which may have influenced students' performance on the 2014 STAAR writing test, and were not controlled in this evaluation.
- Student outcome data were used to assess the impact of Abydos, thus, the fidelity of implementation was not considered in the analysis. The results of this evaluation may not be generalized to indicate the overall effectiveness of Abydos due to implementation variation.

## Results

### What were the demographic characteristics of Abydos and non-Abydos teachers and their students?

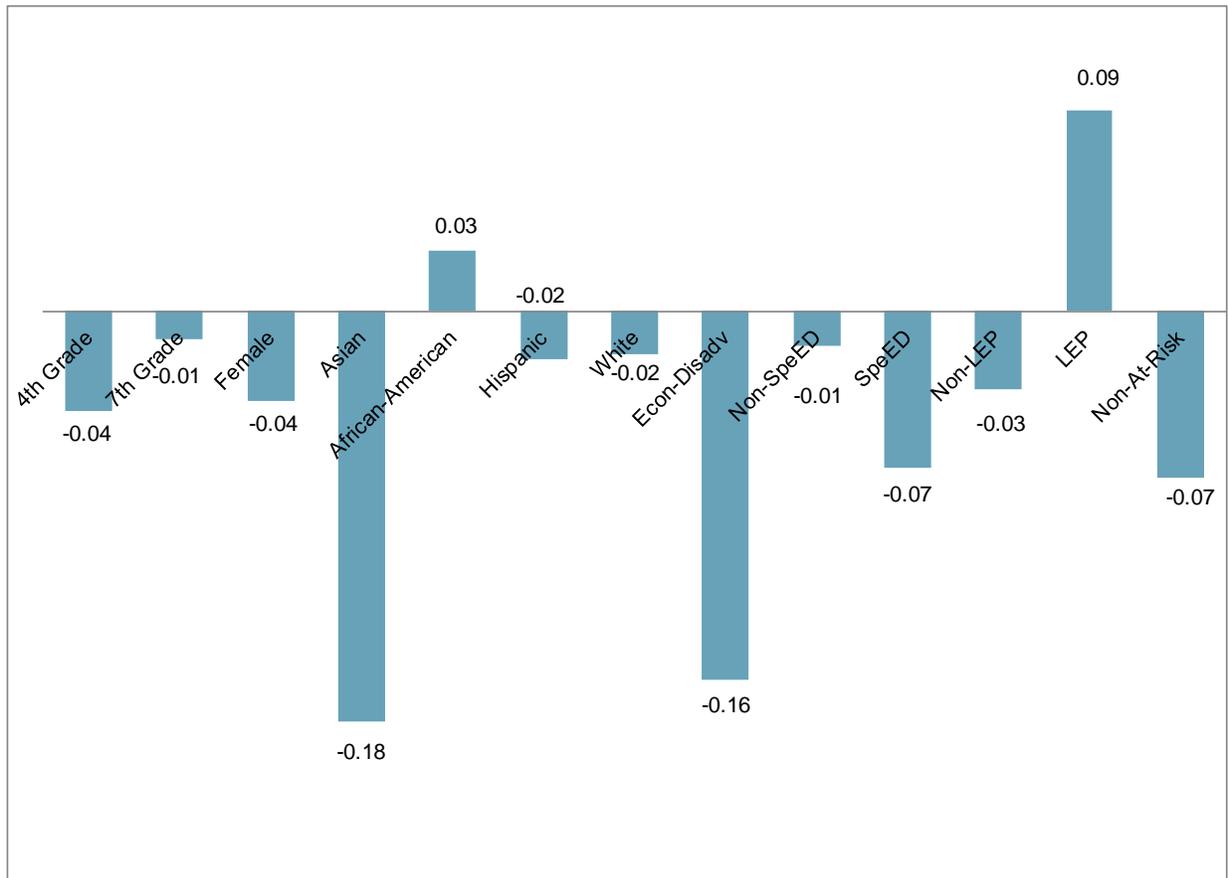
- As Appendix A-Table 1 (p. 13) shows, the demographic characteristics of the Abydos students and non-Abydos students in the analytical sample were comparable with respect to gender, ethnicity, economically-disadvantaged status, special education placement, LEP status, and at-risk status. Notably, in both groups, about 61% of students were Hispanic, 90% were economically-disadvantaged, 33% were LEP, and 58% were at-risk.

### How did Abydos and non-Abydos students perform on the 2013 Stanford reading and language subtests?

A composite score of students' NCE score on the 2013 Stanford reading and language subtests was calculated as an indicator that measured students' reading and language ability before taking 2014 STAAR writing test.

- **Appendix A-Table 2** (p. 14) shows that the mean NCE composite scores of Abydos and non-Abydos students were similar within each student group, except for Asian and non-economically-disadvantaged students. Both Abydos and non-Abydos students had similar reading and language ability before they were exposed to the Abydos writing strategy.
- Asian students in the Abydos group ( $M = 114.3$ ) scored lower than their non-Abydos peers ( $M = 122.1$ ) on the 2013 Stanford reading and language subtests combined (Appendix A-Table 2, p. 14). The corresponding effect size for the mean scale score difference between Abydos and non-Abydos Asian students was  $d = -0.18$ . This effect size indicated that the magnitude of the mean scale score difference was small (Figure 1, p. 6).
- Non-economically-disadvantaged students in the Abydos group ( $M = 117.2$ ) scored lower than their non-Abydos peers ( $M = 123.7$ ) on the 2013 Stanford reading and language subtests combined (Appendix A-Table 2, p. 14). The corresponding effect size for the mean scale score difference between Abydos and non-Abydos Asian students was  $-0.16$ . This effect size indicated that the magnitude of the mean scale score difference was small (Figure 1).

**Figure 1. Effect Sizes of Abydos Students vs. Non-Abydos Students on the 2013 Stanford Reading and Language Subtests Before Exposure to Abydos Writing Strategy**



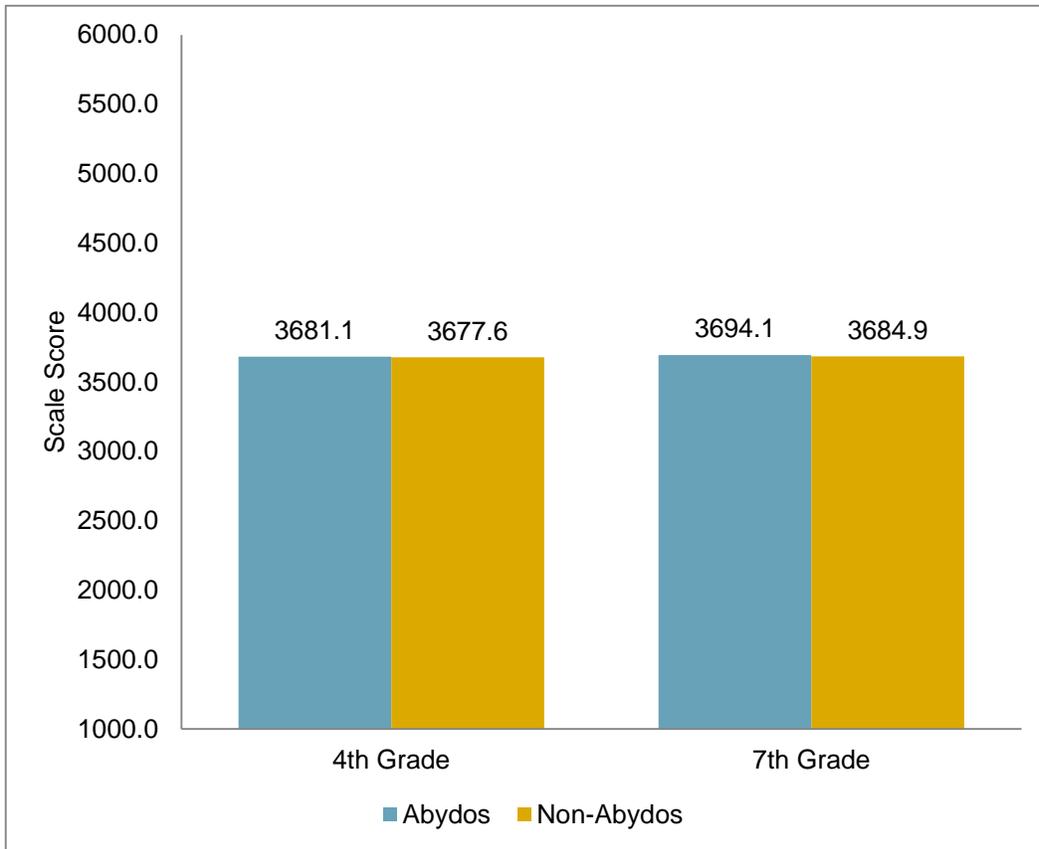
Note. Defined  $d = 0.15$  as small-modest,  $d = 0.3$  as modest-large,  $d = 0.5$  as large. Positive numbers indicate higher performance for the Abydos students.

### How did Abydos and non-Abydos students perform on the 2014 STAAR writing test?

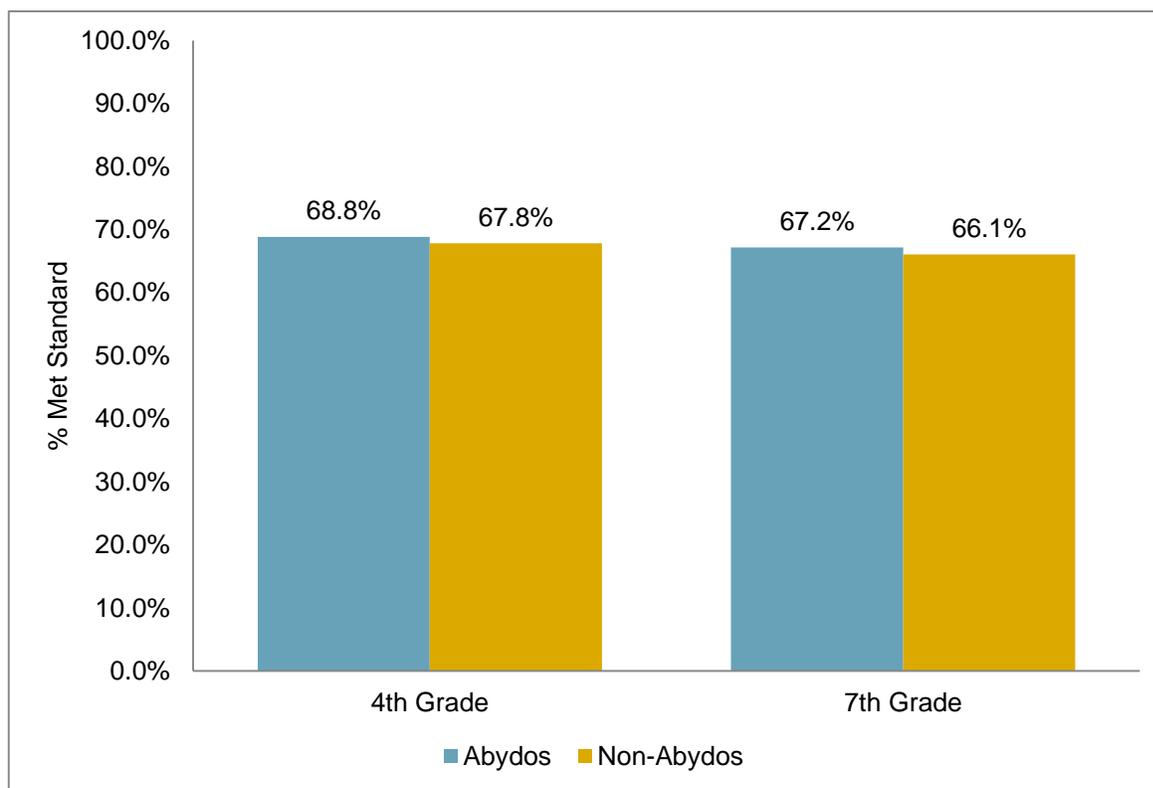
The 2014 STAAR writing performances of Abydos students and non-Abydos students were measured by mean scale score and the percentage of students who met the 2013–2014 STAAR Level II: Satisfactory (Phase-In 1) writing standard.

- **Figure 2** (p. 7) shows that fourth grade Abydos students ( $M = 3681.1$ ) had a higher mean writing scale score than their non-Abydos peers ( $M = 3677.6$ ) on the 2014 STAAR writing test.
- Seventh grade Abydos students ( $M = 3694.1$ ) obtained a higher mean writing scale score than their non-Abydos peers ( $M = 3684.9$ ) on the 2014 STAAR writing test (Figure 2, p. 7).
- The effect sizes for the mean writing scale score differences on the 2014 STAAR writing test between fourth and seventh grade Abydos and non-Abydos students were negligible ( $d < 0.15$ ) (**Appendix A-Table 3**, p. 15).

**Figure 2. Mean Writing Scale Scores on the 2014 STAAR Writing Test for Abydos and Non-Abydos Students**



**Figure 3. Percentage of Abydos and Non-Abydos Students Who Met the 2014 STAAR Level II: Satisfactory (Phase-In 1) Writing Standard**

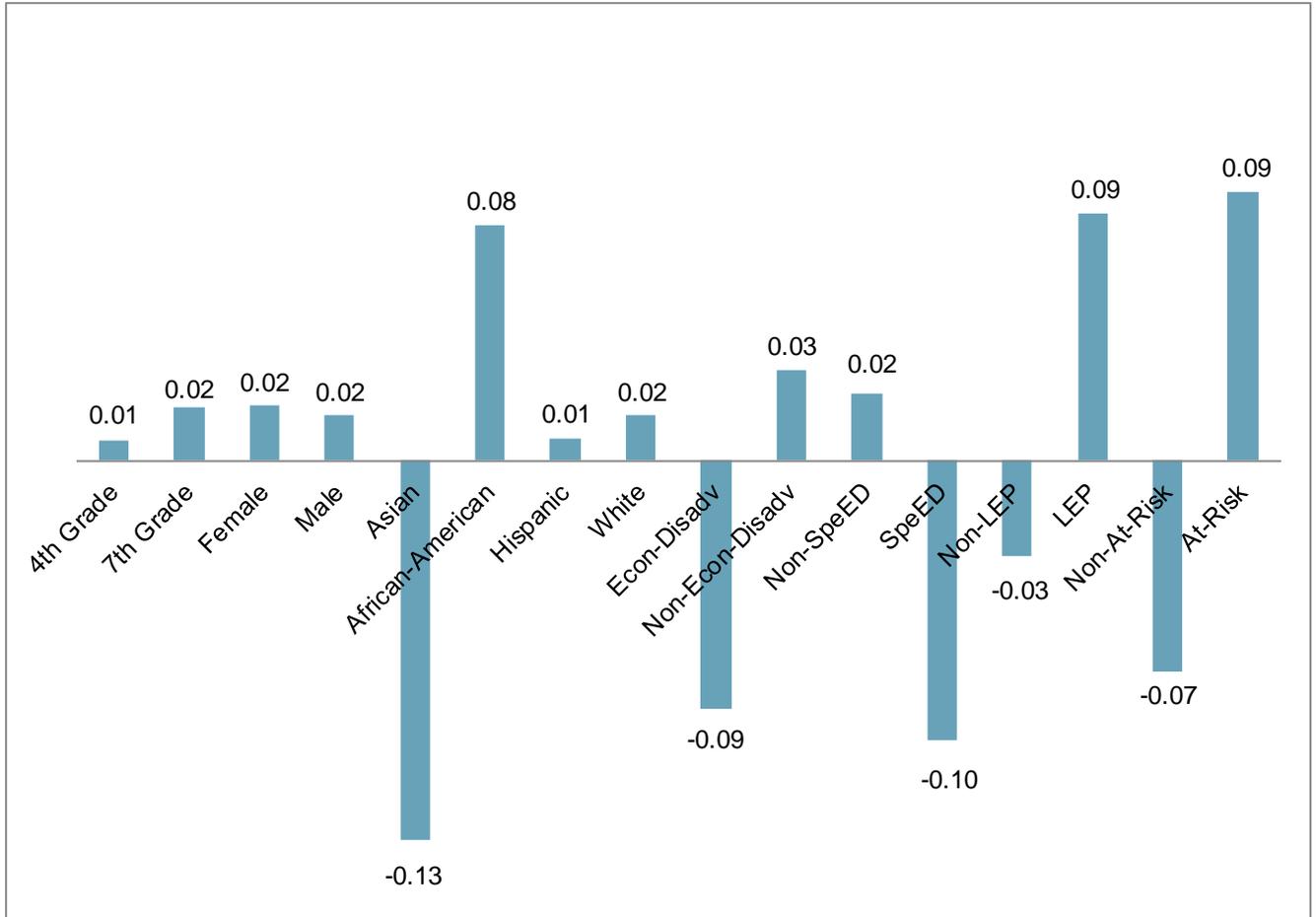


- Overall, 68.8% of the fourth grade Abydos students met the 2014 STAAR Level II: Satisfactory (Phase-In 1) writing standard compared to 67.8% for their non-Abydos peers (**Figure 3**).
- At seventh grade, 67.2% of Abydos students met the 2014 STAAR Level II: Satisfactory (Phase-In 1) writing standard compared to 66.1% for the non-Abydos students (Figure 3).
- The effect sizes for the differences in the percentages of students who met the 2014 STAAR Level II: Satisfactory (Phase-In 1) between fourth and seventh grade Abydos and non-Abydos students were negligible ( $d < 0.15$ ) (**Appendix A-Table 4**, p. 16).

#### Did Abydos impact on students' 2014 STAAR writing performance vary by student groups?

- Appendix A-Table 3 (p. 15) shows that the mean scale scores of Abydos and non-Abydos students on the STAAR writing test were similar within each student group (gender, ethnicity, economically-disadvantaged status, special education placement, LEP status, and at-risk status).
- The effect sizes for mean scale score differences between Abydos and non-Abydos students for all student groups were negligible ( $d < 0.15$ ), which indicated that students in both groups performed comparably on the 2014 STAAR writing test regardless of their demographic information (gender, ethnicity, economically-disadvantaged status, special education placement, LEP status, and at-risk status) (**Figure 4**, p. 9).

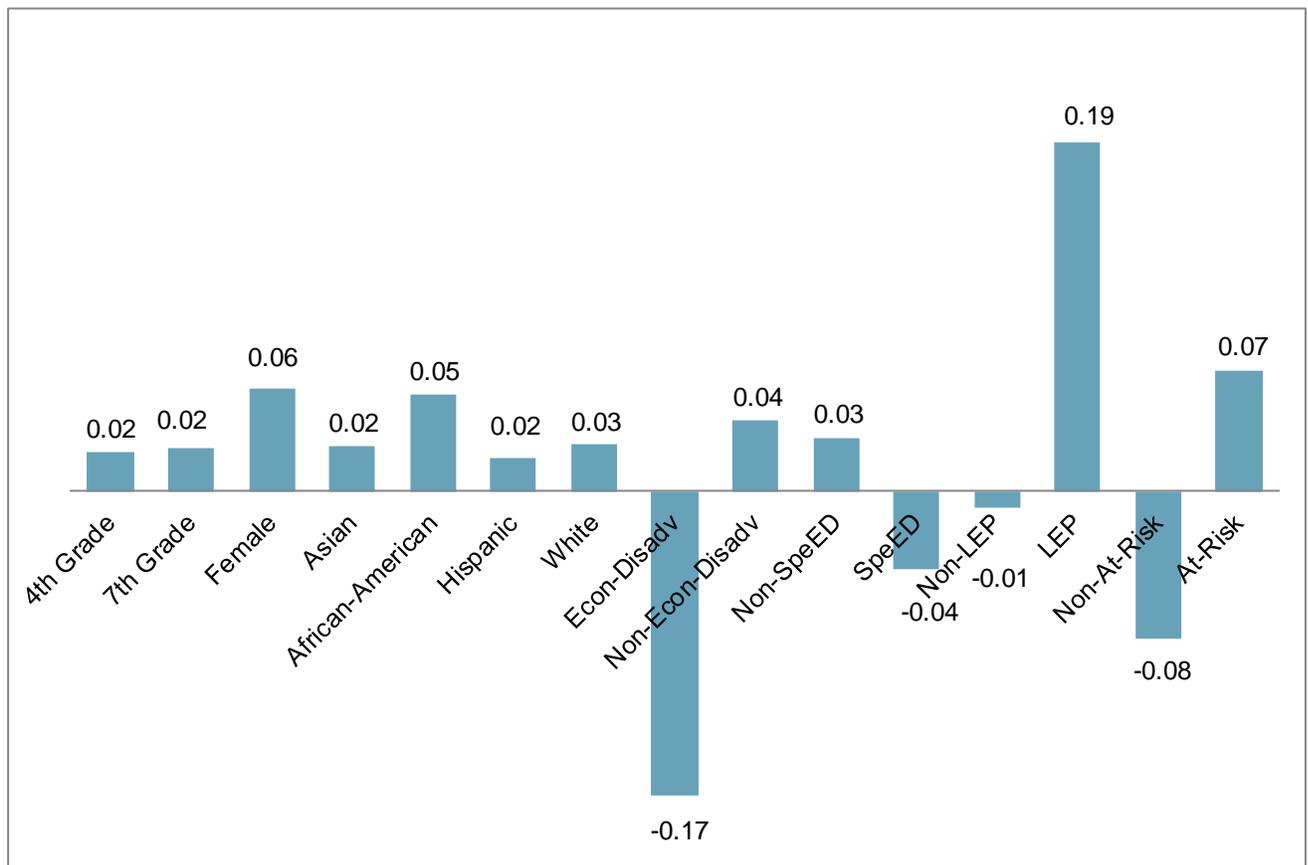
**Figure 4. Effect Sizes for the Mean Writing Scale Score Differences on the 2014 STAAR Writing Test by Student Groups**



Note. Defined  $d = 0.15$  as small-modest,  $d = 0.3$  as modest-large,  $d = 0.5$  as large. Positive numbers indicate higher performance for the Abydos students.

- Appendix A- Table 4 (p. 16) shows that the percentage of Abydos and non-Abydos students who met the 2014 STAAR Level II: Satisfactory (Phase-In 1) writing standard was similar within each student group (gender, ethnicity, economically-disadvantaged status, special education placement, LEP status, and at-risk status).
- The effect sizes for percentage differences between all of the Abydos and non-Abydos student groups were negligible ( $d < 0.15$ ), which indicated that students in the Abydos and non-Abydos groups performed comparably on the 2014 STAAR writing test regardless of their demographic information (gender, ethnicity, economically-disadvantaged status, special education placement, LEP status, and at-risk status) (**Figure 5**).
- Even though the effects were not significant, it is interesting that while only two Abydos student groups (African-American and LEP) performed higher than the non-Abydos students on the 2013 Stanford prior assessment, eleven Abydos student groups performed higher than the non-Abydos students on the 2014 STAAR writing test.

**Figure 5. Effect Sizes for Differences in the Percentage of Students who met 2013–2014 STAAR Level II: Satisfactory (Phase-In 1) Writing Standard**



Note. Defined  $d = 0.15$  as small-modest,  $d = 0.3$  as modest-large,  $d = 0.5$  as large. Positive numbers indicate higher performance for the Abydos students.

### What is the association between the treatment effect of Abydos and students' writing performance by controlling students' prior reading and language knowledge?

Analysis of Covariance (ANCOVA) was used to adjust the treatment effect for the difference in prior reading and language knowledge between the Abydos and non-Abydos groups that existed before the students were exposed to Abydos writing strategies. The dependent variable was student scale scores on the 2014 STAAR writing test, and the independent variable was Abydos treatment effect. Student NCE composite scores on the 2013 Stanford reading and language subtests was used as the covariate. Homogeneity of regression slopes and the linear relationship between dependent variable and covariate assumptions were checked to ensure there was no violation of these assumptions. The ANCOVA results show that teacher's completion of the Abydos training did not significantly affect student performance on the 2014 STAAR writing test with  $p = 0.15$ .

## Discussion

This study evaluated impact of Abydos on students' writing performance. The results of this evaluation showed that Abydos students obtained a higher mean scale score and had a higher percentage of students met the 2014 STAAR Level II: Satisfactory (Phase-In 1) standard than their non-Abydos peers on the 2014 STAAR writing test. However, the differences of mean scale score and of the percentage of who met the 2014 STAAR Level II: Satisfactory (Phase-In 1) writing standard between the two groups were not statistically significant. In education, the benchmark for successful implementation of training lies with student performance. However, in this evaluation, only student outcome data were available to assess the impact of Abydos on students' writing performance, and data of teacher attitude and teacher implementation of Abydos writing strategies in the classroom were not available in the analysis. Therefore, the results of this evaluation may not be generalized to overall effectiveness of Abydos.

In future research, an implementation survey could be used to learn more about the changes in teachers' attitude toward writing and their implementation of writing strategies in the classroom after completion of the professional development. Moreover, teacher interviews or focus groups could be used to find out which characteristics of Abydos foster attitude changes and support implementation of writing strategies in the classroom, as well as how those characteristics can be replicated.

## References

- Allington, R. L. (2005). *What really matters for struggling readers*. New York: Allyn & Bacon.
- Bloodgood, J. (2002). Quintilian: A classical educator speaks to the writing process. *Reading Research and Instruction*, 42(1), 30–43.
- Borman, G.D., & D'Agostino, J.V. (1996). Title I and student achievement: A meta-analysis of federal evaluation results. *Educational Evaluation and Policy Analysis*, 18, 309–326.
- Carroll, J. A., & Wilson, E. E. (2009). *The Abydos Learning International Three-Week Writing Course*. Retrieved Jun 1, 2014, from <http://www.abydoslearning.org>
- Darling-Hammond, L. (1998). Teacher learning that supports student learning. *Educational Leadership*, 55(5), 6–11.
- Field, A.P. (2013). *Discovering statistics using IBM SPSS Statistics: And sex and drugs and rock 'n' roll* (4<sup>th</sup> ed). London: Sage.
- Friedman, T. L. (2006). *The world is flat*. New York: Farrar, Straus, and Giroux.
- Fullan, M. (1995). The limits and the potential of professional development. Professional development in education: New paradigms and practices (pp. 253–268). New York: Teachers College Press.
- Gusky, T. R. (1985). Staff development and teacher change. *Educational Leadership*, 42, 57–60.
- Hillocks, G. (2005). The focus on form vs. content in teaching writing. *Research in the Teaching of English*, 40(2), 238–248.
- Kulik, J. A., Kulik, C. C. and Bangert, R. L. (1984) Effects of practice on aptitude and achievement test scores. *American Education*
- National Center for Education Statistics (2003). *NAEP Writing: Achievement levels*. Retrieved January 25, 2008, from <http://nces.ed.gov/nationsreportcard/writing/achieveall.asp>
- Rosenbaum, P.R. & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41–55.
- Wills, C. E., & Stommel, M. (2002). Graduate nursing students' pre- and post course perceptions and preferences concerning full web-based courses. *Journal of Nursing Education*, 40, 193–201.

## Appendix A

**Table 1. Demographic Characteristics of 4<sup>th</sup> and 7<sup>th</sup> Grade Students, 2013–2014**

Demographic Characteristic		Abydos (n = 4,374)		Non-Abydos (Analytical Sample) (n = 4,374)		Non-Abydos (Study Sample) (n = 15,249)	
		n	%	n	%	n	%
Grade	4 <sup>th</sup>	1,104	25.2%	1,093	25.0%	8,184	53.7%
	7 <sup>th</sup>	3,270	74.8%	3,281	75.0%	7,065	46.3%
Gender	Female	2,124	48.6%	2,156	49.3%	7,641	50.1%
	Male	2,250	51.4%	2,218	50.7%	7,608	49.9%
Ethnicity	Asian	100	2.3%	165	3.8%	718	4.7%
	African-American	1,405	32.1%	1,338	30.6%	3,765	24.7%
	Hispanic	2,648	60.5%	2,650	60.6%	9,000	59.0%
	White	188	4.3%	192	4.4%	1,553	10.2%
	Other	33	.8%	29	.7%	213	1.4%
Economically Disadvantaged	No	455	10.4%	475	10.9%	3,448	22.6%
	Yes	3,919	89.6%	3,899	89.1%	11,799	77.4%
Special Education	No	4,171	95.4%	4,178	95.5%	14,659	96.1%
	Yes	203	4.6%	196	4.5%	587	3.9%
Limited English Proficient (LEP)	No	2,935	67.1%	2,946	67.4%	12,432	81.6%
	Yes	1,434	32.8%	1,426	32.6%	2,806	18.4%
At-Risk	No	1,807	41.3%	1,849	42.3%	7,743	50.8%
	Yes	2,567	58.7%	2,525	57.7%	7,504	49.2%

*Note.* The demographic information used in this evaluation was based on student information at the time that the student took the 2014 STAAR writing test.

**Table 2. Mean NCE Composite Scores on the 2013 Stanford Reading and Language Subtests by Student Groups**

Student Group	Abydos (n = 4,374)			Non-Abydos (n = 4,374)			Mean Difference	Effect Size (d)	
	Mean	SD	n	Mean	SD	n			
Overall Sample	87.5	37.6	4,374	88.3	38.2	4,374	-0.8	-0.02	
Grade	(2013–2014) 4 <sup>th</sup>	90.3	35.9	1,104	91.9	38.4	1,104	-1.6	-0.04
	(2013–2014) 7 <sup>th</sup>	86.6	38.1	3,270	87.0	38.0	3,270	-0.4	-0.01
Gender	Female	91.4	36.1	2,124	92.8	36.7	2,146	-1.4	-0.04
	Male	83.9	38.5	2,250	83.9	39.1	2,228	0.0	0.00
Ethnicity	Asian	114.3	45.7	100	122.1	43.7	144	-7.8	-0.18
	African-American	83.6	36.3	1,405	82.6	35.6	1,326	1.0	0.03
	Hispanic	85.4	34.9	2,648	86.1	36.0	2,685	-0.7	-0.02
	White	127.5	44.6	188	128.3	42.4	191	-0.8	-0.02
	Other	119.7	45.5	33	114.3	46.1	28	--	--
Economically disadvantaged	No	117.2	42.7	455	123.7	39.4	452	-6.5	-0.16
	Yes	84.1	35.3	3,919	84.2	35.9	3,922	-0.1	0.00
Special Education	No	89.3	37.0	4,171	89.8	37.7	4,185	-0.5	-0.01
	Yes	51.3	28.6	203	53.3	31.9	189	-2.0	-0.07
Limited English Proficient (LEP)	No	92.8	36.7	3,643	94.1	36.8	3,662	-1.3	-0.03
	Yes	61.0	29.6	726	58.4	30.4	711	2.6	0.09
At-Risk	No	116.1	29.6	1,807	118.2	28.8	1,802	-2.1	-0.07
	Yes	67.4	28.4	2,567	67.3	28.9	2,572	0.1	0.00

Note. 1.) Effect size and mean difference were not reported when n < 30, and were denoted by "--"; 2.) Defined d = 0.15 as small-modest, d = 0.3 as modest-large, d = 0.5 as large; 3.) The composite score is the sum of NCE scores on the 2013 Stanford reading and language subtests, therefore scores can range from 1-200.

**Table 3. Mean Scale Scores on the 2014 STAAR Writing Test by Student Groups**

Student Group		Abydos (n = 4,374)			Non-Abydos (n = 4,374)			Mean Difference	Effect Size (d)
		Mean	SD	n	Mean	SD	n		
Overall Sample		3690.8	485.1	4,374	3683.0	497.8	4,374	7.8	0.02
Grade	4 <sup>th</sup>	3681.1	452.9	1,104	3677.6	497.2	1,104	3.5	0.01
	7 <sup>th</sup>	3694.1	495.5	3,270	3684.9	498.0	3,270	9.2	0.02
Gender	Female	3768.1	489.1	2,124	3758.6	502.1	2,146	9.5	0.02
	Male	3617.8	469.8	2,250	3610.3	482.7	2,228	7.5	0.02
Ethnicity	Asian	4108.8	641.5	100	4195.1	665.0	144	-86.3	-0.13
	African- American	3645.9	473.6	1,405	3607.6	456.3	1,326	38.3	0.08
	Hispanic	3660.5	442.6	2,648	3657.0	465.0	2,685	3.5	0.01
	White	4145.8	615.4	188	4136.3	590.2	191	9.5	0.02
	Other	4177.6	674.2	33	4026.2	548.7	28	--	--
Economically disadvantaged	No	4071.8	628.7	455	4123.9	579.0	452	-52.1	-0.09
	Yes	3646.6	445.0	3,919	3632.2	461.3	3,922	14.4	0.03
Special Education	No	3714.0	479.3	4,171	3702.6	493.5	4,185	11.4	0.02
	Yes	3214.2	334.0	203	3248.9	382.0	189	-34.7	-0.10
Limited English Proficient (LEP)	No	3741.0	487.5	3,643	3747.0	490.1	3,662	-6.0	-0.03
	Yes	3440.3	386.6	726	3353.2	396.7	711	87.1	0.09
At-Risk	No	3997.3	461.9	1,807	4030.2	437.2	1,802	-32.9	-0.07
	Yes	3475.1	371.6	2,567	3439.8	379.4	2,572	35.3	0.09

Note. 1.) Effect size and mean difference were not reported when n < 30, and were denoted by "--"; 2.) Defined d = 0.15 as small-modest, d = 0.3 as modest-large, d = 0.5 as large.

**Table 4. Percentage of Students Who Met the 2014 STAAR Level II: Satisfactory (Phase-In 1) Writing Standard by Student Groups**

Student Group		Abydos (n = 4,374)		Non-Abydos (n = 4,374)		Difference	Effect Size (d)
		%	n	%	n		
Overall Sample		67.6%	4,374	66.5%	4,374	1.1	0.01
Grade	4 <sup>th</sup>	68.8%	1,104	67.8%	1,104	1.0	0.01
	7 <sup>th</sup>	67.2%	3,270	66.1%	3,270	1.1	0.02
Gender	Female	74.3%	2,124	71.8%	2,146	2.5	0.04
	Male	61.3%	2,250	61.4%	2,228	-0.1	0.00
Ethnicity	Asian	91.0%	100	90.3%	144	0.7	0.02
	African-American	64.0%	1,405	61.4%	1,326	2.6	0.03
	Hispanic	67.0%	2,648	66.1%	2,685	0.9	0.01
	White	87.8%	188	86.9%	191	0.9	0.02
	Other	84.8%	33	82.1%	28	--	--
Economically disadvantaged	No	84.2%	455	89.8%	452	-5.6	-0.12
	Yes	65.7%	3,919	63.8%	3,922	1.9	0.02
Special Education	No	69.6%	4,171	68.3%	4,185	1.3	0.02
	Yes	25.6%	203	27.5%	189	-1.9	-0.02
Limited English Proficient (LEP)	No	71.7%	3,643	72.1%	3,662	-0.4	-0.01
	Yes	47.2%	726	37.7%	711	9.5	0.11
At-Risk	No	90.6%	1,807	92.8%	1,802	-2.2	-0.06
	Yes	51.4%	2,567	48.1%	2,572	3.3	0.04

Note. 1.) Effect size and mean difference were not reported when n < 30, and were denoted by "--"; 2.) Defined d = 0.15 as small-modest, d = 0.3 as modest-large, d = 0.5 as large.

## Appendix B

### Propensity Score Matching

Propensity score matching can be used to address the concern of quasi-experimental studies, selection bias, due to the inherently non-experimental nature of the design. A quasi-experimental design assigns members to the treatment group and control group by a method other than random assignment. A random assignment is an ideal method for observational studies because randomization can produce comparable treatment and control groups prior to the treatment. In this evaluation, the teachers in the treatment group and the control group may not be comparable due to their demographic characteristics and experience. In order to recreate a situation that resembles a randomized experiment, propensity score matching was used to select a group of students in treatment group that matched the control group students as much as possible in term of the observable characteristics. Propensity score analysis can yield unbiased causal effect estimates because the balance between groups' propensity score produce on average balance on observed covariates, even though matched individuals will typically differ on many observed covariates (Rosenbaum and Rubin, 1983).