An Analysis of Achievement Scores of Arthur Academy Schools, 2007 to 2013

Charles Arthur – Early Child Literacy Consultant, Arthur Reading Workshop &
Jean Stockard, Office of Research and Evaluation, National Institute for Direct Instruction
AUGUST 7, 2014

i

An Analysis of Achievement Scores of Arthur Academy Schools, 2007-2013

Arthur Academies are a set of six charter elementary schools in the greater Portland, Oregon metropolitan area**. The mission of the schools is "to accelerate educational achievement and academic competency of all its students" and "to become an effective and innovative model of instruction that can influence teaching practices in other schools" (Arthur, 2009**-10, p. 1). This report examines data on the academic achievement of Arthur students from six consecutive school years: 2007-2008 through 2012-2013 and shows how the schools are accomplishing this mission. The data indicate that, at the start of kindergarten, Arthur students had achievement scores that were similar to or slightly lower than students in the nation as a whole. However, by the end of kindergarten they had achievement scores that were significantly higher than their peers. This high level of achievement persists, and even increases, through later grades. The body of the text below includes selected figures to illustrate this phenomenon. Tables with the supporting data and explanations of the statistical analyses are in the appendix.

Background

The Arthur Academies are managed by the Mastery Learning Institute, a Charter Management Organization. The first of the six schools opened in the fall of 2002. Two more schools opened in 2004, a fourth in 2005, and two more in the fall of 2007. Each of the schools is in a different school district.

All of the schools use Direct Instruction programs for reading and mathematics: *Reading Mastery* and *Connecting Math Concepts*. These Direct Instruction programs use an incremental, mastery-learning approach to teaching basic subjects of reading, math and language. This specialized approach to teaching is based on a comprehensive model of instruction. Providing this model is based on the belief that a powerful way of teaching exists that is not being utilized in most schools, and therefore our charter schools offer it as a choice. **The basis of Mastery Learning is that a child's rate of progress is determined by the** extent to which he or she masters carefully sequenced lessons and activities, culminating in the mastery, or thorough learning, of essential foundational skills and knowledge.

This model, with these programs, is the most thoroughly documented educational reform model in elementary and middle school grades. It emphasizes well-developed and carefully planned lessons, designed around small learning increments and prescribed teaching tasks. Learning is arranged very incrementally so that students find learning easy but challenging and, therefore, can be successful in mastering everything that is taught as they progress

through the programs. The curriculum materials break down all general objectives into very small teaching progressions. All activities are very carefully sequenced so that they can be easily learned, mastered and gradually accumulated towards larger objectives. The activities are presented to children in very exacting, interactive ways so that children are motivated.

The model includes sustained attention to maintaining quality of the implementation of the programs and high fidelity to the guidelines of the DI programs. A great deal of emphasis is placed on having very capable and well-trained teachers, especially in the very early grades. Extensive training is provided yearly during two weeks before fall opening as well as in-class coaching during the year to attain and maintain quality teaching with these specialized programs.

The Direct Instruction model of instruction is academically focused right from the beginning, in kindergarten. It is based on the belief that academic learning can be highly motivating by itself when taught clearly, systematically and enthusiastically. It has been observed that children are excited and gain self-confidence when they are learning how to read and do math in kindergarten. We have learned that, if at-risk children who have the highest likelihood of learning problems can start kindergarten in a strong academic program, many of their learning difficulties can be prevented. The data in the next section illustrate this phenomenon.

Strong Achievement Growth in Kindergarten

All Arthur Academy students are given the Stanford Achievement Test (SAT), a norm-referenced test, in the fall and spring of their kindergarten year. Figure 1 and Figure 2 report the percentage of students who scored at or above the 40th percentile in reading and math at both testing periods from 2008-09 to 2012-13. This benchmark was chosen to indicate a point of "average" achievement. By definition, 60 percent of the students in the nation have scores within this range, a level indicated by the solid horizontal line in the figures.

In most of the comparisons, the percentage of Arthur students meeting this range at the start of kindergarten (the bars labeled "fall") was substantially smaller than the 60 percent national norm. In other words, the Arthur students had lower scores than comparable students in the nation. Yet, by the spring of the year the percentage of students in this range was substantially larger than the 60 percent level. (Compare the bars labeled "spring" with the line labeled "nation.") That is, by spring the Arthur students had much higher scores than comparable students in the nation. (See Table A-1 for statistical details.)
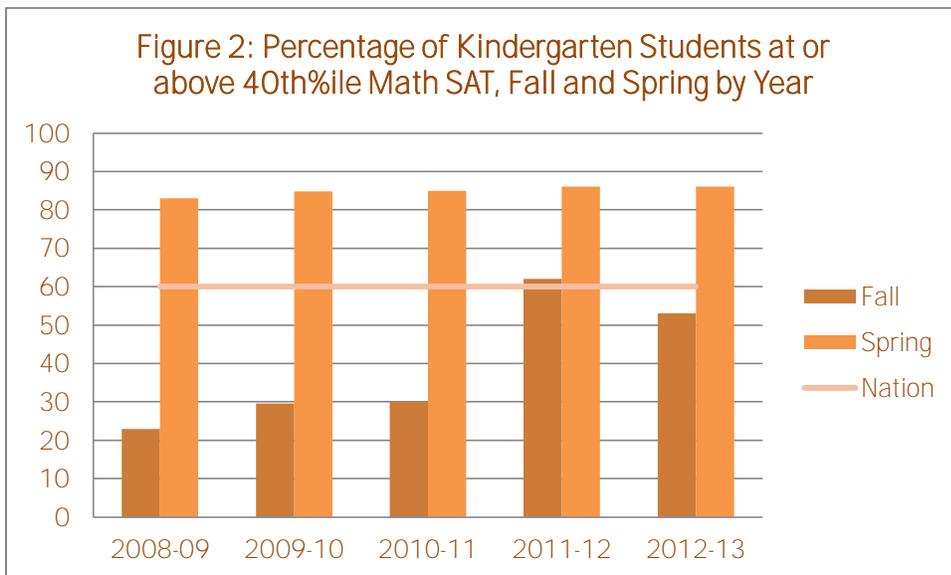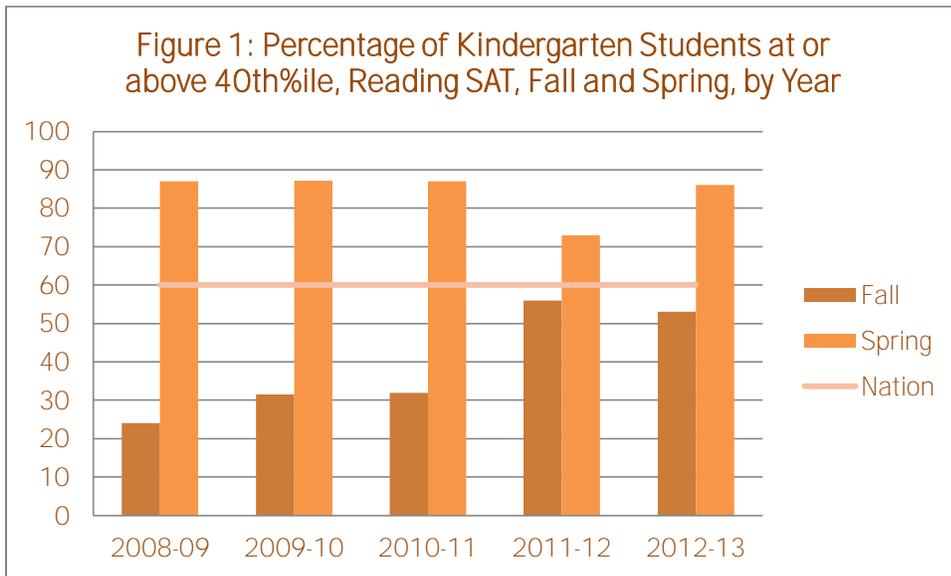
Figure 1: Percentage of Kindergarten Students at or above 40th%ile, Reading SAT, Fall and Spring, by Year

Figure 2: Percentage of Kindergarten Students at or above 40th%ile Math SAT, Fall and Spring by Year

Figure 3 reports results using a much higher standard, the 80th percentile, as the point of comparison, In the nation as a whole, 20 percent of students score at or above that level. At entry to kindergarten ten percent or fewer of the Arthur students (about half of what would be expected) were in this high achieving range. By the end of the year, however, from three to four times as many students were in this range.

In other words, at the start of their kindergarten year, the average Arthur student had SAT scores that were lower than students in the nation as a whole. However, by the end of their kindergarten year, the situation had reversed, and the average Arthur student scored much higher than peers in the nation. In all cases, the changes over time, relative to the national norms, were statistically significant. The associated effect sizes ranged from .35 to 1.12, with an average of value of .72. Traditionally, effect sizes of .25 or larger are seen as **"educationally important," and those above .60 are seen as large.** Thus, these changes in the achievement of Arthur kindergarten students were both statistically significant and educationally important. Providing this kind of accelerated progress in kindergarten gives all children a huge advantage for success in future grades, a phenomenon that is documented in the next section of this report.

Figure 3: Percentage of Kindergarten Students at or above 80th%ile, SAT, Fall & Spring, by Year and Subject
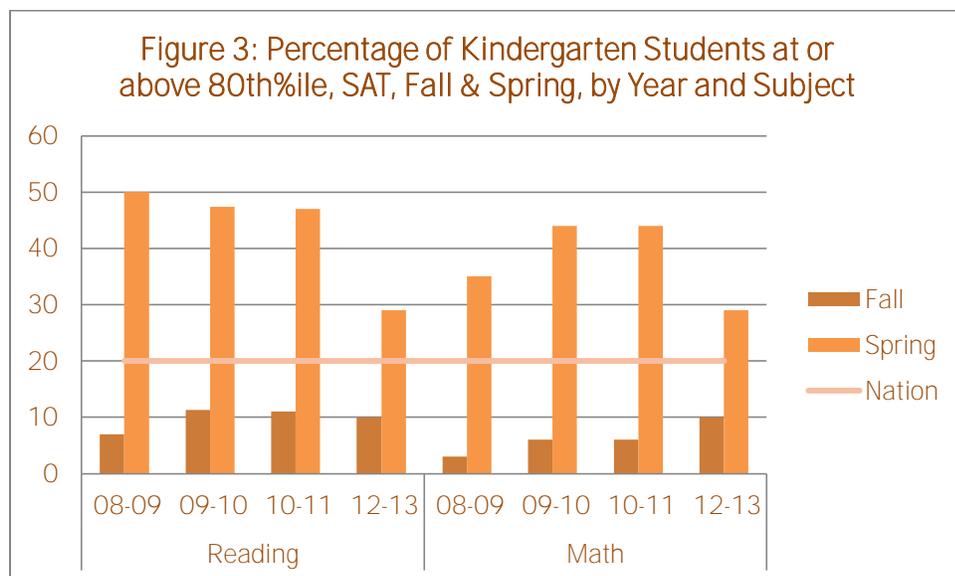
High Achievement Levels Persist Through Elementary School

Two types of achievement data were available for students in grades after Kindergarten: SAT scores like those shown in Figures 1 to 3 were available for grades 1 to 5 in from 2007-08 to 2010-11 and grades 1-2 in 2011-12 and 2012-13. Scores from the Oregon Assessment of Knowledge and Skills (OAKS) were available for grades 3 to 5 for 2009-10 to 2012-13. The remaining figures in this report show how Arthur students maintained their high levels of achievement through their elementary years.

Arthur Students' SAT Scores Surpass National Norms in Grades 1 to 5

Figures 4, 5, 6 and 7 **summarize students'** spring scores on the SAT from 2008 to 2013. Figures 4 and 5 report the percentage of students scoring at or above average, defined as at or above the 40th percentile). Figures 6 and 7 report the percentage of students scoring within the high achieving range, defined as scoring at or above the 80th percentile. The format of the graphs is like that used in Figures 1, 2, and 3 with the bars showing Arthur

Figure 4: Percentage of Arthur Students At or Above the 40th %ile in SAT Reading, Spring, by Grade and Year

Figure 5: Percentage of Arthur Students At or Above 40th Percentile in SAT Math, Spring, by Grade and Year
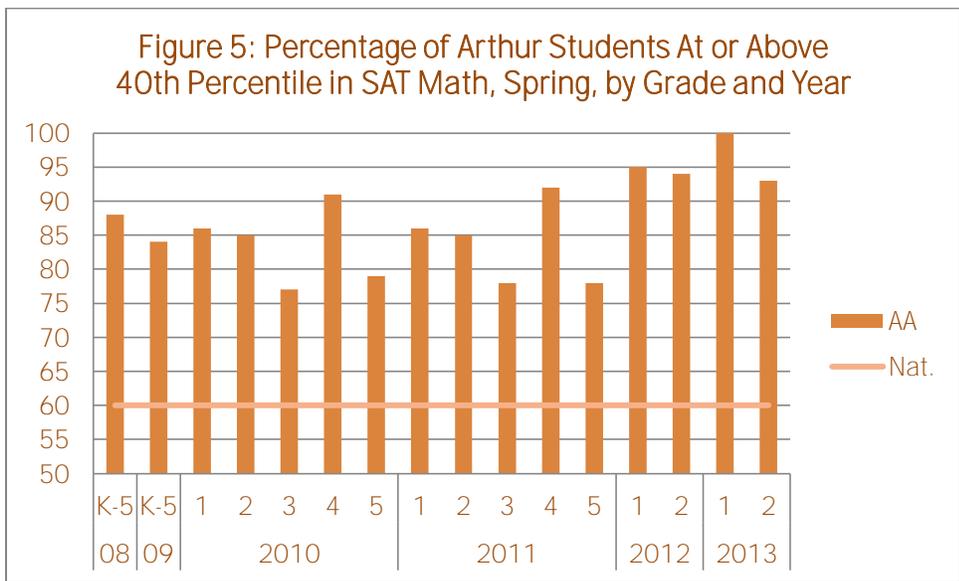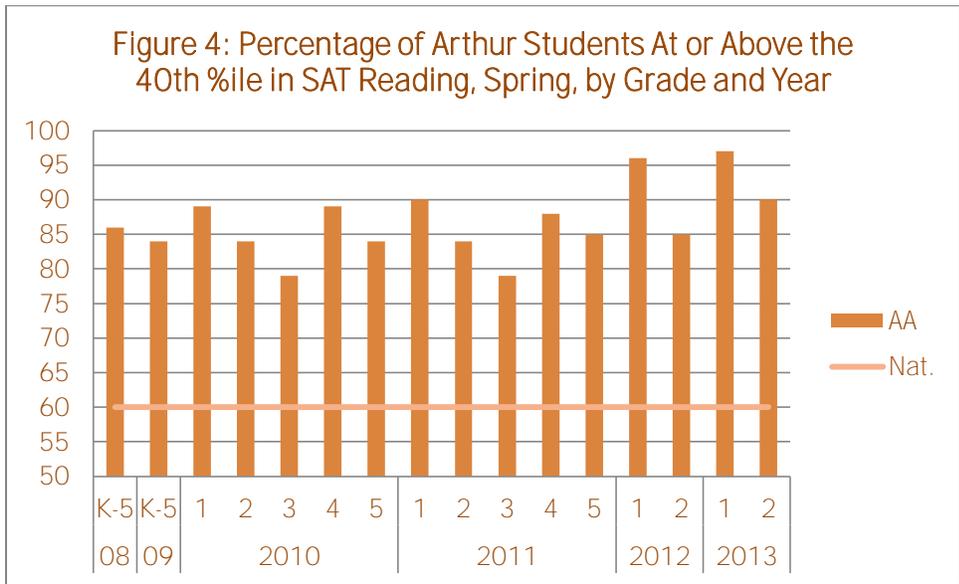
students' scores and the horizontal lines representing the national norm. (Data for 2008 and 2009 were not available separately by grade, so these results are aggregated across all grade levels.)

Figure 6: Percent of Arthur Students At or Above the 80th %ile, SAT Reading, Spring, by Grade and Year

Grade	K-5	K-5	1	2	3	4	5	1	2	3	4	5	1	2
AA	51	41	41	33	35	46	45	41	33	35	46	45	47	48
Year	08	09			2010					2010-11			2012-13	

Figure 7: Percentage of Arthur Students At or Above 80th%ile in SAT Math, Spring, by Grade and Year

Grade	K-5	K-5	1	2	3	4	5	1	2	3	4	5	1	2
AA	54	42	66	40	39	51	39	66	40	39	51	39	70	44
Year	08	09			2010					2010-11			2012-13	
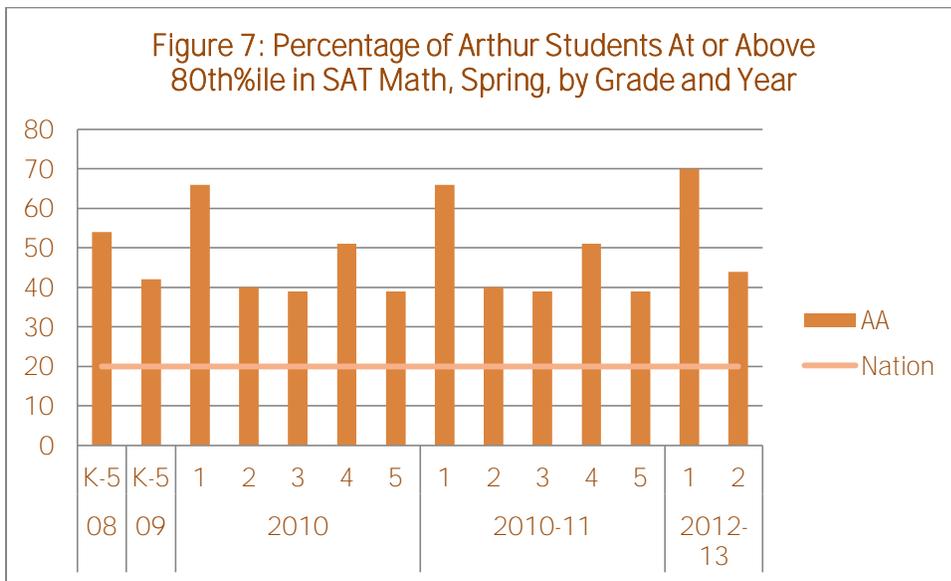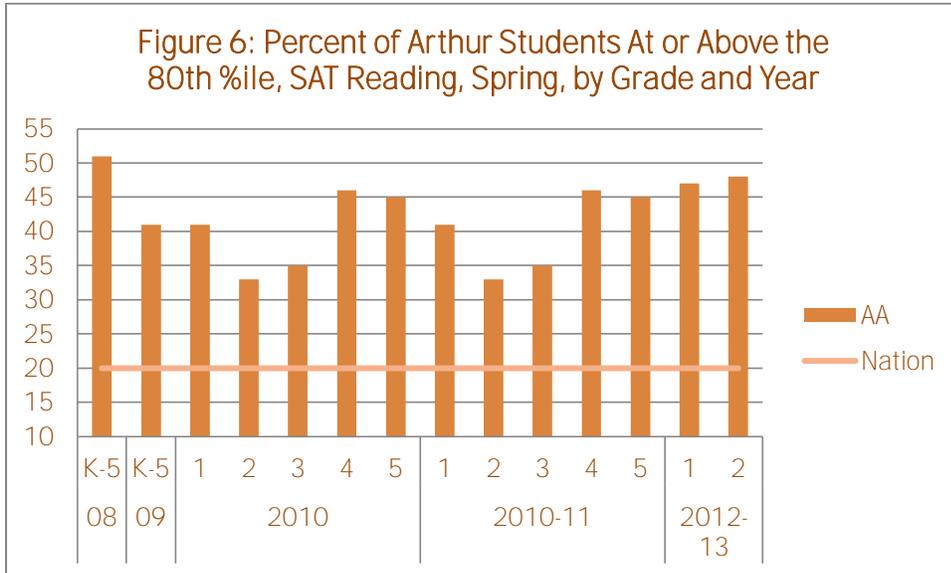
In all cases, the percentage of students in the depicted ranges was substantially higher than would be expected given national norms. All comparisons were statistically significant. The average effect size across all of the results was .59, more than twice the level traditionally

used to denote educationally important results and very similar to that found in other meta-analyses of DI programs. (See Coughlin 2014 for a summary of meta-analyses and systematic reviews of the efficacy of DI programs.)

Tables A-2 and A-3 in the Appendix provide additional details on the results shown in Figures 4 through 7. Tables A-4 through A-7 provide details on changes from fall to spring in SAT scores. In almost all cases, the Arthur students' gains in achievement over a single school year were significantly greater than would be expected given the national norms. The average effect size associated with these changes was .43, again well above the traditional level of educational importance.

Arthur Students' OAKS Scores Surpass District and State Figures

Figures 8, 9 and 10 summarize students' scores on the reading, mathematics, and science sections of the OAKS, given in the spring of each year. The bars represent the percentage of Arthur students who met the state defined benchmark on the assessments, and the lines in the graphs report the average percentage of students in the districts in which the schools are located and in the state as a whole that met the benchmark.

The OAKS benchmarks are, by definition, a much easier criterion to meet than the national mean on the SAT. While only half of the national population is above the mean on the SAT, the data shown in Figures 8 to 10 indicate that over two-thirds of the students in the state met most of the benchmarks. Given such a high percentage of students meeting this mark, the "room" for surpassing this level is thus relatively small.

In almost all cases, however, Arthur students were more likely to score at the proficient level than students in their home districts or in the state as a whole. The average effect size across all comparisons with the district was .48 and the average effect size for comparisons with the state was .32, both above the level traditionally used to denote educationally important effects. Most of the comparisons were also statistically significant. (See Table A-8 for detailed statistical results and discussion in the appendix of the related calculations.)

Figure 8: Percentage of AA Students Meeting OAKS Reading Benchmark, by Year and Grade
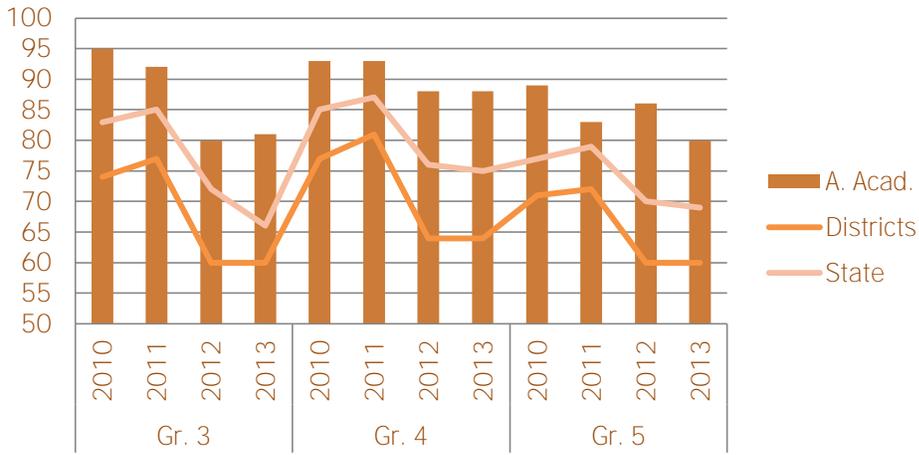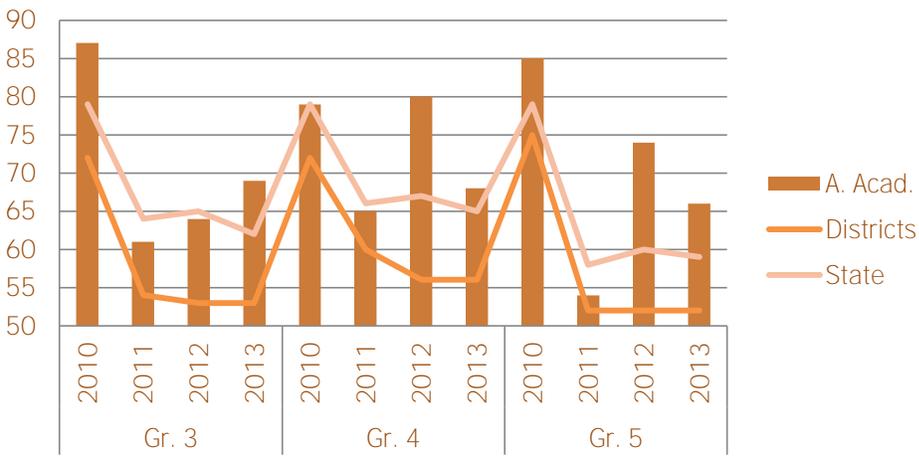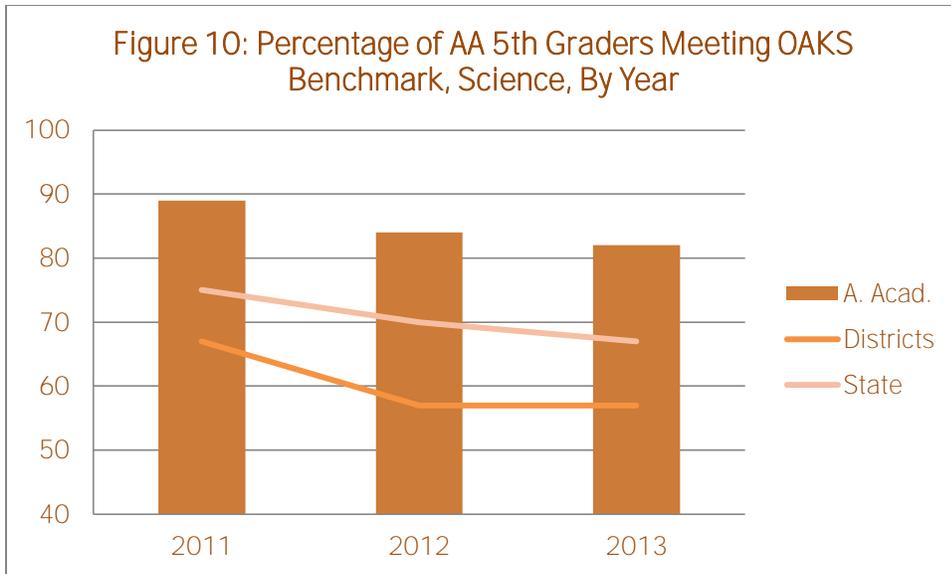

Figure 9: Percentage of AA Students Meeting OAKS Benchmark, Mathematics, by Year and Grade

Figure 10: Percentage of AA 5th Graders Meeting OAKS Benchmark, Science, By Year

Discussion and Summary

This document has analyzed data on the academic achievement of Arthur Academy students over a six year period. The results provide a clear picture of academic success. At the beginning of kindergarten the average Arthur student generally had achievement scores that were lower than national norms. Yet, by the end of kindergarten, the average student scored well above others in the nation. This high achievement continued to appear in the upper elementary grades. In almost all analyses, Arthur students had higher Stanford Achievement Test (SAT) and Oregon State Assessment Scores (OAKS) than counterparts in their districts, state, and nation. Almost all of the comparisons were statistically significant and the associated effect sizes were well beyond the levels used to denote educational importance.

The replication of the results across six academic years is noteworthy. The high achievement patterns do not appear to be a fluke, but instead appear to be a consistent pattern. It should also be noted that the analysis is based on the aggregation of results combined across six schools, allowing more confident use of statistical manipulations. There is no indication that the pattern of strong achievement gains varied from one school to another. Moreover, the fact that the results have been replicated in six different settings is especially noteworthy and should provide additional confidence in the results. Finally, the fact that consistent positive results appeared in different subject matters (reading, mathematics, and, for fifth grade, science) and with two different assessments (SAT and OAKS) provides additional confidence in the results.

Taken together, the data in this report provide substantial evidence that the Arthur Academy schools are meeting their stated mission. They have clearly accelerated the educational achievement and academic competency of their students. Moreover, they can serve as an effective and innovative model of instruction to influence teaching practices in other schools.

Appendix

Details on the analysis summarized above are given in this appendix. They describe the data more completely and explain the procedures used to calculate the effect sizes and tests of significance reported in the text.

Strong Achievement Growth in Kindergarten

Table A-1 presents the statistical results for the data that are shown in Figures 1, 2, and 3. The first two columns of Table A-1 give the subject matter and year of data used in the analysis. The first column of data (third column in the table) reports the percentage of students in a given range for the nation as a whole (60 percent for the 40th percentile and higher and 20 percent for the 80th percentile and higher). The next two columns of data report the percentage of Arthur students within this range at the start and end of the kindergarten year. The following column, labeled ES change (for Effect Size change), gives the effect size associated with the change, relative to the national norms, from fall to spring, The next two columns report the associated t-ratio and probability level; and the final column gives the number of Arthur students in the comparison. The first panel of data reports results for scores at or above the 40th percentile and the second panel reports results for scores at or above the 80th percentile.

The following logic was used in calculating the effect sizes. Results for reading for the 40th percentile for kindergartners in 2010-11 (the third row in Table A-1) are used as an example. In the fall of 2010, 32 percent of Arthur students fell in this range and in the spring 87 percent of the students did so. In the nation, by definition, 60% of the students scored at or above this level at both time periods. This national group is the population to which Arthur students are compared.

Using the logic and terminology associated with analyses of binomial distributions, the value of 60% is equivalent to Pu = .60, the proportion of cases in the population. By definition, the standard deviation of this population,

$\sigma_{pu} = \sqrt{Pu * Qu}$, where Qu = 1-Pu (the square root of the product of Pu and Qu).   (1)

For this population $\sigma_{pu} = \sqrt{(.60 * .40)}$ = .49   (2)

The effect size comparing Arthur students with the nation at any given time point is simply (Ps-Pu)/ σ_{pu} ,   (3)
the difference between the sample and population value divided by the standard deviation of the population. In the fall, the effect size,

$d_{fall,} = (.32-.60)/.49 = .28/.49 = -.57.$ (4)
In the spring, the effect size,
$d_{spring,} = (.87-.60)/.49 = .27/.49 = +.55.$ (5)

In other words, in the fall, the probability that an Arthur student would score in the average or above range was .57 of a standard deviation lower than the probability for the population as a whole. In the spring, the probability that an Arthur student would score in this range was .55 of a standard deviation higher than that for the nation.

The effect size associated with the change from fall to spring is calculated as
$d_{change} = d_{spring} - d_{fall} = .55 - (-.57) = 1.12.$ (6)

The t-ratio and associated probabilities shown in the last two columns of Table A-1 were also calculated using the logic associated with the binomial distribution. This t-ratio tests the null hypothesis that the difference from fall to spring equals zero. The standard error, is defined as
$\sigma_{pu}/\sqrt{N-1}.$ (7)
For the data in the third line of Table A-1, the sample size is 133, and the standard error equals
$\sigma_{pu}/\sqrt{N-1} = .49/\sqrt{(133-1)} = .49/11.49 = .043$ (8)

Then, by definition, the t-ratio,
$t = [(P_{spring,Arthur}-P_{spring,u}) - (P_{fall,Arthur}-P_{fall,u})/[\sigma_{pu}/\sqrt{N-1}]$, and, because $P_{spring,u} = P_{fall,u,}$
$t = [P_{spring,Arthur} - P_{fall,Arthur}/[\sigma_{pu}/\sqrt{N-1}]$ (9)

And, for the data in the third line of Table A-1,
$t = (P_{sp}-P_{fall})/[\sigma_{pu}/\sqrt{N-1}] = (.87-.32)/.043 = .55/.043 = 12.90.$ (10)

This t-value would occur by chance far less than one time out of 1000.

High Achievement Levels Persist Through Elementary School

Figures 4 to 8 in the text compare students' spring scores to those in the nation, state and districts. The sections describe the data in more detail and explain the associated calculations of effect sizes and tests of significance.

Arthur Students' SAT Scores Surpass National Norms in Grades 1 to 5

Figures 4 to 7 and the related discussion compare the spring SAT scores of students to national values, focusing on the percentage at or above the 40th percentile (Figures 4 and 5) and the 80th percentile (Figures 6 and 7). Tables A-2 and A-3 display the data used to construct these figures.

The logic for calculating effect sizes and t-ratios for these data is very similar to that explained directly above. To calculate the effect size associated with the comparison to the national average, the following logic was used, using data in Table A-2 as an example. The first panel of data in Table A-2 compares the percentage of students scoring at or above average (defined as at or above the 40th percentile on the SAT reading test) to the percentage expected given national norms. For example, among Arthur students in 2010-11, 90% of first graders, 84% of second graders, and 79% of third graders scored in this range. The second column of data is 60 for all grades, reflecting the fact that, by definition, 60% of all students would score at the 40th percentile or higher. Using equation (1) above, based on the binomial distribution, the standard deviation for the population is defined as

$$\sigma_{pu} = \sqrt{(Pu * Qu)} = \sqrt{(.60 * .40)} = .49$$

The effect size of the difference between the percentage of Arthur students and that for the nation can then be calculated as

$$d = (Ps - Pu)/\sigma_{pu}, \tag{11}$$

where Ps is the proportion for the sample and Pu is the proportion for the population.
For the data for first graders in 2010-11,
d = (.90 - .60)/.49 = .30/.49 = .61.
In other words, in the spring of 2011, the percentage of Arthur students scoring at or above average was .61 of a standard deviation higher than would be expected given national data.

A simple t-ratio can be used to test the null hypothesis that this effect size (or the difference of Arthur students from the national value) is significantly greater than zero. The calculation of the associated t-ratio is a simple extension of this logic, using the definition of the standard error given in equation (8) above:

$$\sigma_{pu}/\sqrt{N - 1}.$$

The sample size for first grade students in the spring of 2011 was 142. Thus,

$$\sigma_{pu}/\sqrt{N - 1} = .49/\sqrt{(142 - 1)} = .49/11.87 = .041.$$

The t-ratio is then calculated as

$t = (P_s-P_u)/[\ \sigma_{pu}/\sqrt{N-1}]$. (12)

For first grade Arthur students in the spring of 2011,

$t = (P_s-P_u)/[\ \sigma_{pu}/\sqrt{N-1}] = (.90-.60)/.041 = .30/.041 = 7.32$.

A t-ratio of this magnitude would be expected to occur far less than one time out of a 1000 if the null hypothesis were true.

Arthur Students' OAKS Scores Surpass District and State Figures

Table A-4 displays the data represented in Figures 8, 9 and 10, which compare the scores of Arthur Students to the average scores of students in the home districts of the schools and the state as a whole. The format of the table is similar to that of other tables. It shows the relevant percentages for Arthur students, the districts, and the state (the first three columns of data), the effect sizes associated with the comparisons of the Arthur students' scores to those of the district and state, and the associated t-ratios. For instance, the first line of data indicates the percentage of third grade students in Arthur Academies, the districts and the state who met the OAKS reading benchmark in 2010 (95%, 74%, and 83% respectively).

Again, the binomial distribution was used for calculating the effect sizes and t-ratios. However, a correction factor was included to adjust for the "ceiling effect" or limited amount of change possible for Arthur students relative to the larger population. Data in the second line of Table A-4 are used as an example of the calculations.

The correction factor used was based on the maximum effect size that would occur if all students within a population were to increase their achievement levels to meet the benchmark; i.e. if Pu = 1.00. Using equation 10,
In general, the maximum effect size for a given population is calculated as

$d_{max} = (1- Pu)/\sqrt{(Pu * Qu)}$. (13)

For the data in the second row of Table A-4, the maximum effect size for the district, where Pu = .77, would be calculated as

$d_{max, district} = (1-.77)/\sqrt{(.77 * .23)} = .23/.42 = .55$

The maximum effect size for the state, where Pu = .85, would be calculated as

$d_{max, state} = (1-.85)/\sqrt{(.85 * .15)} = .15/.36 = .42$

Using equation 10, the raw, or uncorrected, effect size for the Arthur students relative to the district would be calculated as

$d_{A, district(raw)} = (.92-.77)/\sqrt{(.77 * .23)} = .15/.42 = .36$;

and the raw effect size relative to the state would be calculated as

$$d_{A,\ state(raw)} = (.92 - .85)/\sqrt{(.85 * .15)} = .07/.36 = .19.$$

The corrected effect size is calculated as simply the ratio of the raw effect size to the maximum effect size

$$d_{cor.} = d_{raw}/d_{max}.\tag{14}$$

For the second line of data in Table A-4
$$d_{cor.,\ district} = .36/.55 = .65$$

and
$$d_{cor.,\ state} = .19/.42 = .45.[1]$$

Note that the correction factor only impacts the value of the effect size when Pu approaches unity. In addition, note that the correction factor would not be appropriate in the comparisons involving groups defined by percentiles, as in the previous section.

The t-ratios associated with these comparisons were calculated from the effect sizes using a formula derived from simple manipulations of equations 10 and 11 above:
$$t = d * \sqrt{N - 1}.\tag{15}$$

Arthur Students' Achievement Scores Continue to Improve in Grades 1 to 5
Both SAT and OAKS data were available on changing scores over time for students in grades 1 to 5. While these results are mentioned in the body of this report, associated figures were not given. The sections below describe the findings in more detail.

Increasing SAT Scores from Fall to Spring
Data from the Stanford Achievement Test are in Tables A-5 through A-8. The interpretation of the tables and the calculation of the effect sizes followed the procedure described above for Table A-1. Results show that students' scores improved significantly from fall to spring in both reading and mathematics in all 5 grade levels.

---

[1] The numbers in these calculations are slightly different than those in Table A-4 because of rounding. The calculations in Table A-4 were done using excel and thus included many more decimal points.

The first two panels of Table A-5 show data for the 2007-08 and 2008-09 school years, in which the available data were aggregated across the 6 grade levels (K to 5). In all cases the percentage of students scoring in either the average or above or high achieving ranges increased strongly and significantly during the school year. Separate data were available for first and second graders were available for students from 2009-10 through 2012-13 (four years, see last panel of Table A-5 first panels in Table A-6, and the entirety of Tables A-7 and A-9). all three years. On average over four-fifths (82% of the students) fell into the average and above range (at the 40th percentile or higher) at the start of the school year. By the spring an additional 10% (92%) were in this range. The increase over time was statistically significant in all but three cases. The exceptions involved subjects and years in which over 85% of the students were in the range at baseline, thus having little range for improvement (2nd grade reading in 2012-13, and 2nd grade math in 2011-12 and 2012-13). The associated effect sizes ranged from zero to .94, with an average of .21, slightly lower than the level traditionally termed educationally important. Similar results appeared with the analysis of first and second grade students scoring at the 80th percentile or higher. In all but two cases there were strong increases from fall to spring and the increases were statistically significant. The two exceptions involve second grade students in the 2012-13 school year, where over twice as many students as would be expected scored in this high range.

Data on SAT scores were available for students in grades three, four and five for the 2009-10 and 2010-11 school years. (See Tables A-5 and A-6). Again, comparisons involved changes in the percentage of students scoring at or above the 40th percentile and at or above the 80th percentile. In all cases the percentage of students scoring in these ranges increased during the school year. The average effect size across all comparisons was .29, and most of the changes were statistically significant. The exceptions usually involved cases in which the students began the school year with relatively high percentages of students already scoring at the designated level.

OAKS Performance Increases Over Time

The OAKS data were **used to compare students' performance over time for two cohorts:** those who were in fourth grade in the spring of 2012 and in fifth grade in the spring of 2013 (called Cohort 1), and students who were in third grade in the spring of 2012 and fourth grade in spring 2013 (Cohort 2). Data for the cohorts and comparison data for each cohort for students in the districts and state are in Table A-9. Effect sizes and t-ratios were calculated following the logic outlined in previous sections. Effect sizes that compare the percentage of Arthur students meeting benchmark to percentages within the district and state were calculated for each time point and cohort using equations 10 and 12 above. The

effect size associated with the change over time was then calculated using equation 6. The t-ratio associated with this change was calculated using equation 13.

The patterns of results differed slightly for the two cohorts and subjects. However, for all comparisons, the percentage of Arthur students meeting benchmark increased or stayed the same from one grade to the next, while the percentage of students in the districts and state stayed the same or decreased. As a result, all of the associated effect sizes were positive and three of the associated effect sizes were statistically significant.

Note, however, that the sample size for each cohort declined from one year to the next. There was no way to determine the extent to which selection bias (lower scoring students dropping out of the Academies from one year to the next) could account for the change over time. Also, because the sample size differed across the two testing periods an average of the sample size for the two periods was used in the calculation of the standard error. Given, however, the very positive and strong results for the SAT data, which involved the same students over time, the possibility of such differential attrition affecting the results seems quite small.

Table A-1

Percentage of Kindergarten Students At or Above 40th Percentile or 80th Percentile, Stanford Achievement Test, Fall and Spring, Nation and Arthur Schools, 2008-09 to 2012-13

			At or Above Average (40th Percentile or Higher)					
		Nation	Start of K	End of K	ES Change	t-ratio	Prob.	N
Reading	2008-09	60	24	87	1.29	15.00	***	137
	2009-10	60	32	87	1.14	13.05	***	133
	2010-11	60	32	87	1.12	12.90	***	133
	2011-12	60	56	73	0.35	4.66	***	181
	2012-13	60	53	86	0.67	8.39	***	156
Math	2008-09	60	23	83	1.22	14.28	***	137
	2009-10	60	30	85	1.13	12.92	***	132
	2010-11	60	30	85	1.12	12.90	***	133
	2011-12	60	62	86	0.49	6.57	***	181
	2012-13	60	53	86	0.67	8.39	***	156
			High Achieving (80th Percentile or Higher)					
		Nation	Start of K	End of K	ES Change	t-ratio	Prob.	N
Reading	2008-09	20	7	50	1	10.24	***	137
	2009-10	20	11	47	1	8.46	***	133
	2010-11	20	11	47	0.90	10.34	***	133
	2012-13	20	10	29	0.48	5.91	***	156
Math	2008-09	20	3	35	0.65	7.62	***	137
	2009-10	20	6	44	0.77	8.85	***	132
	2010-11	20	6	44	0.95	10.91	***	133
	2012-13	20	10	29	0.48	5.91	***	156

*, p< .05; **, p< .01; ***,p<.001; all tests are one-tail reflecting the research hypothesis that students in the Direct Instruction programs will have higher levels of achievement.

Table A-2

Percentage of Students At or Above 40th Percentile or 80th Percentile, Reading, Stanford Achievement Test, Spring, Grades 1-5, Nation and Arthur Schools, by Year and Grade

			At or Above Average (40th Percentile or Higher)				
<u>Year</u>	<u>Grade</u>	<u>Arthur Academy</u>	<u>Nation</u>	<u>Eff. Size</u>	<u>t-ratio</u>		<u>N</u>
2007-08	K - Gr. 5	86	60	0.53	13.54	***	652
2008-09	K - Gr. 5	84	60	0.49	13.99	***	816
2009-10	Gr. 1	89	60	0.60	7.13	***	142
	Gr. 2	84	60	0.49	5.87	***	144
	Gr. 3	79	60	0.40	4.69	***	141
	Gr. 4	89	60	0.58	6.88	***	140
	Gr. 5	84	60	0.50	5.78	***	135
2010-11							
	Gr. 1	90	60	0.61	7.27	***	142
	Gr. 2	84	60	0.49	5.86	***	144
	Gr. 3	79	60	0.39	4.59	***	141
	Gr. 4	88	60	0.57	6.74	***	140
	Gr. 5	85	60	0.51	5.91	***	135
2011-12, All Students							
	Gr. 1	90	60	0.61	7.6	***	155
	Gr. 2	83	60	0.47	5.94	***	161
2011-12, Continuing Students							
	Gr. 1	96	60	0.73	8.54	***	136
	Gr. 2	85	60	0.51	6.1	***	144
2012-13							
	Gr. 1	97	60	0.76	9.43	***	157
	Gr. 2	90	60	0.61	7.5	***	151
			High Achieving (80th Percentile or Higher)				
<u>Year</u>	<u>Grade</u>	<u>Arthur Academy</u>	<u>Nation</u>	<u>Eff. Size</u>	<u>t-ratio</u>		<u>N</u>
2007-08	K - Gr. 5	51	20	0.78	19.77	***	652
2008-09	K - Gr. 5	41	20	0.53	14.99	***	816
2009-10	Gr. 1	41	20	0.52	6.19	***	142
	Gr. 2	33	20	0.33	3.99	***	144
	Gr. 3	35	20	0.39	4.57	***	141

Gr. 4	46	20	0.66	7.79	***	140
Gr. 5	45	20	0.63	7.29	***	135
2010-11						
Gr. 1	41	20	0.53	6.23	***	142
Gr. 2	33	20	0.33	3.89	***	144
Gr. 3	35	20	0.38	4.44	***	141
Gr. 4	46	20	0.65	7.66	***	140
Gr. 5	45	20	0.63	7.23	***	135
2012-13						
Gr. 1	47	20	0.68	8.43	***	157
Gr. 2	48	20	0.7	8.57	***	151

Table A-3

Percentage of Students At or Above 40th Percentile or 80th Percentile, Mathematics, Stanford Achievement Test, Spring, Grades 1-5, Nation and Arthur Schools, by Year and Grade

		At or Above Average (40th Percentile or Higher)					
Year	Grade	Arthur Academy	Nation	Eff. Size	t-ratio		N
2007-08	K - Gr. 5	88	60	0.57	14.78	***	670
2008-09	K - Gr. 5	84	60	0.49	14.00	***	818
2009-10	Gr. 1	86	60	0.53	6.42	***	145
	Gr. 2	85	60	0.50	6.03	***	144
	Gr. 3	77	60	0.36	4.23	***	142
	Gr. 4	91	60	0.64	7.56	***	140
	Gr. 5	79	60	0.38	4.38	***	135
2010-11							
	Gr. 1	86	60	0.53	6.3	***	142
	Gr. 2	85	60	0.51	6.1	***	144
	Gr. 3	78	60	0.37	4.35	***	141
	Gr. 4	92	60	0.65	7.70	***	140
	Gr. 5	78	60	0.37	4.25	***	135
2011-12							
	Gr. 1	95	60	0.71	8.3	***	136
	Gr. 2	94	60	0.69	8.27	***	143
2012-13							
	Gr. 1	100	60	0.82	10.2	***	157
	Gr. 2	93	60	0.67	8.25	***	151
		High Achieving (80th Percentile or Higher)					
Year	Grade	Arthur Academy	Nation	Eff. Size	t-ratio		N
2007-08	K - Gr. 5	54	20	0.85	21.99	***	670
2008-09	K - Gr. 5	42	20	0.55	15.72	***	818
2009-10							
	Gr. 1	66	20	1.16	13.86	***	145
	Gr. 2	40	20	0.49	5.85	***	144
	Gr. 3	39	20	0.47	5.56	***	142
	Gr. 4	51	20	0.77	9.05	***	140
	Gr. 5	39	20	0.48	5.57	***	135
2010-11							
	Gr. 1	66	20	1.15	13.66	***	142

Gr. 2	40	20	0.5	5.98	***	144
Gr. 3	39	20	0.475	5.62	***	141
Gr. 4	51	20	0.775	9.14	***	140
Gr. 5	39	20	0.48	5.5	***	135
2012-13						
Gr. 1	70	20	1.25	15.61	***	157
Gr. 2	44	20	0.6	7.35	***	151

Table A-4

OAKS Scores Arthur Academy, Districts and State, by Year, Grade, and Subject

Reading									
	A. Acad.-%	Districts -%	State - %	Ef Size-Dist	Ef Size-State	t-district		t-state	
Grade 3									
2010	95	74	83	0.81	0.71	9.79	***	8.56	***
2011	92	77	85	0.65	0.47	8.04	***	5.75	***
2012	80	60	72	0.50	0.29	6.40	***	3.66	***
2013	81	60	66	0.53	0.44	6.66	***	5.60	***
Grade 4									
2010	93	77	85	0.70	0.53	8.23	***	6.31	***
2011	93	81	87	0.63	0.46	7.47	***	5.46	***
2012	88	64	76	0.67	0.50	8.27	***	6.20	***
2013	88	64	75	0.67	0.52	8.03	***	6.26	***
Grade 5									
2010	89	71	77	0.62	0.52	6.18	***	5.19	***
2011	83	72	79	0.39	0.19	4.56	***	2.21	*
2012	86	60	70	0.65	0.53	7.69	***	6.31	***
2013	80	60	69	0.50	0.35	5.63	***	4.00	***

Mathematics									
	A. Acad.-%	Districts -%	State - %	Ef Size-Dist	Ef Size-State	t-district	prob.	t-state	prob.
Grade 3									
2010	87	72	79	0.54	0.38	6.50	***	4.62	***
2011	61	54	64	0.15	-0.08	1.88	*	-1.03	
2012	64	53	65	0.23	-0.03	3.00	**	-0.37	
2013	69	53	62	0.34	0.18	4.32	***	2.34	**
Grade 4									
2010	79	72	79	0.25	0.00	2.96	**	0.00	
2011	65	60	66	0.13	-0.03	1.48		-0.35	
2012	80	56	67	0.55	0.39	6.77	***	4.89	***
2013	68	56	65	0.27	0.09	3.28	**	1.03	
Grade 5									
2010	85	75	79	0.40	0.29	3.98	***	2.84	**
2011	54	52	58	0.04	-0.10	0.48		-1.11	
2012	74	52	60	0.46	0.35	5.42	***	4.14	***
2013	66	52	59	0.29	0.17	3.29	***	1.92	*

Science (Grade 5 Only)

	A. Acad.- %	Districts - %	State - %	Ef Size- Dist	Ef Size- State	t- district	prob.	t- state	
2011	89	67	75	0.67	0.56	7.75	***	6.51	***
2012	84	57	70	0.63	0.47	7.43	***	5.52	***
2013	82	57	67	0.58	0.45	6.55	***	5.12	***

Note: Sample sizes for 2010 were 148 for grade 3, 141 for grade 4, and 100 for grade 5; for 2011 were 153 for grade 3, 141 for grade 4, and 136 for grade 5; for 2012 were 165 for grade 3, 155 for grade 4, and 141 for grade 5; for 2013 were 162 for grade 3, 146 for grade 4 and 128 for grade 5. *, p< .05; **, p < .01, *** p<.001 (two-tail). As with earlier tables, all probabilities are one-tail reflecting the research hypothesis that the DI group would have higher scores.

Table A-5

Changes in SAT Reading and Math Scores, Fall to Spring, 2007-08, 2008-09, 2009-10), Arthur Academy Charter Schools

	Fall %	Spring %	Ef. Size	t-ratio	
Fall 2007 to Spring 2008					
Reading, Grades K-5 (N=652)					
Average or Above (>=40th %ile)	62	86	0.49	12.50	***
High Achieving (>=80th %ile)	20	51	0.78	19.77	***
Mathematics, Grades K-5 (n=670)					
Average or Above (>=40th %ile)	59	88	0.73	18.75	***
High Achieving (>=80th %ile)	20	54	0.85	21.99	***
Fall 2008 to Spring 2009					
Reading, Grades K-5 (N=816)					
Average or Above (>=40th %ile)	53	84	0.63	18.06	***
High Achieving (>=80th %ile)	20	41	0.53	14.99	***
Mathematics, Grades K-5 (n=818)					
Average or Above (>=40th %ile)	51	84	0.83	23.58	***
High Achieving (>=80th %ile)	17	42	0.63	17.86	***
Fall 2009 to Spring 2010					

	Fall %	Spring %	Ef. Size	t-ratio	n
Grade 1 (n-147 Reading, 145 Math)					
Reading, Average or Above (>=40th %ile)	44	89	0.92	11.12	***
Reading, High Achieving (>=80th %ile)	16	41	0.50	6.08	***
Mathematics, Average or Above (>=40th %ile)	53	86	0.68	8.11	***
Mathematics, High Achieving (>=80th %ile)	13	66	1.08	13.01	***
Grade 2 (n=144)					
Reading, Average or Above (>=40th %ile)	53	84	0.64	7.63	***
Reading, High Achieving (>=80th %ile)	17	33	0.34	4.07	***
Mathematics, Average or Above (>=40th %ile)	62	85	0.47	5.59	***
Mathematics, High Achieving (>=80th %ile)	28	40	0.23	2.71	**
Grade 3 (N=141 reading, 142 math)					
Reading, Average or Above (>=40th %ile)	68	79	0.23	2.74	**
Reading, High Achieving (>=80th %ile)	30	35	0.12	1.37	
Mathematics, Average or Above (>=40th %ile)	64	77	0.27	3.24	***
Mathematics, High Achieving (>=80th %ile)	18	39	0.43	5.12	***
Grade 4 (N=140)					
Reading, Average or Above (>=40th %ile)	72	89	0.34	3.95	***
Reading, High Achieving (>=80th %ile)	26	46	0.42	4.99	***
Mathematics, Average or Above (>=40th %ile)	69	91	0.45	5.33	***

Mathematics, High Achieving (>=80th %ile)	25	51	0.52	6.19	***
Grade 5 (N=135)					
Reading, Average or Above (>=40th %ile)	76	84	0.18	2.10	*
Reading, High Achieving (>=80th %ile)	41	45	0.09	1.05	
Mathematics, Average or Above (>=40th %ile)	73	79	0.12	1.40	
Mathematics, High Achieving (>=80th %ile)	30	39	0.18	2.10	*

Table A-6
Changes in SAT Reading and Math Scores, 2010-11, Arthur Academy Charter Schools

	Reading				
	<u>Pre-Test</u>	<u>Post-Test</u>	<u>ES Change</u>	<u>t-ratio</u>	
<u>First Grade (n=142)</u>					
Aver. or Above (>=40th %ile)	44	90	0.94	11.11	* * *
Top 20 (>=80th %ile)	16	41	0.63	7.40	* * *
<u>Second Grade (n=144)</u>					
Aver. or Above (>=40th %ile)	53	84	0.63	7.49	* * *
Top 20 (>=80th %ile)	17	33	0.40	4.78	* * *
<u>Third Grade (n=141)</u>					
Aver. or Above (>=40th %ile)	68	79	0.22	2.66	* * *
Top 20 (>=80th %ile)	30	35	0.13	1.48	
<u>Fourth Grade (n=140)</u>					
Aver. or Above (>=40th %ile)	72	88	0.33	3.86	* * *
Top 20 (>=80th %ile)	26	46	0.50	5.89	* * *
<u>Fifth Grade (n=135)</u>					
Aver. or Above (>=40th %ile)	75	85	0.20	2.42	* *
Top 20 (>=80th %ile)	41	45	0.10	1.16	
	Mathematics				
<u>Kindergarten</u>	<u>Pre-Test</u>	<u>Post-Test</u>	<u>ES Change</u>	<u>t-ratio</u>	
<u>First Grade</u>					
Aver. or Above (>=40th %ile)	53	86	0.67	7.97	* * *
Top 20 (>=80th %ile)	13	66	1.33	15.90	* * *
<u>Second Grade</u>					
Aver. or Above (>=40th %ile)	61	85	0.49	5.80	* * *
Top 20 (>=80th %ile)	28	40	0.30	3.59	* * *
<u>Third Grade</u>					
Aver. or Above (>=40th %ile)	65	78	0.27	3.14	* * *
Top 20 (>=80th %ile)	18	39	0.53	6.23	* * *
<u>Fourth Grade</u>					
Aver. or Above (>=40th %ile)	70	92	0.45	5.31	* * *
Top 20 (>=80th %ile)	25	51	0.65	7.66	* * *
<u>Fifth Grade</u>					
Aver. or Above (>=40th %ile)	72	78	0.12	1.45	
Top 20 (>=80th %ile)	30	39	0.23	2.60	* *

Table A-7

Changes in SAT Reading and Math Scores, Fall, 2011 to Spring, 2012, Arthur Academy Charter Schools

Reading				
	Fall	Spring	Ef. Size	t-ratio
First Grade (continuing students, n = 136)				
Average or Above (>=40th %ile)	87	96	0.18	2.13 *
First Grade (all students (n=155)				
Average or Above (>=40th %ile)	82	90	0.16	2.03 *
Second Grade (continuing students, n= 144)				
Average or Above (>=40th %ile)	74	85	0.22	2.69 **
Second Grade (all students (n=161)				
Average or Above (>=40th %ile)	70	83	0.27	3.36 ***
Mathematics				
	Fall	Spring	Ef. Size	t-ratio
First Grade (n=136)				
Average or Above (>=40th %ile)	81	95	0.29	3.32 ***
Second Grade (n=143)				
Average or Above (>=40th %ile)	87	94	0.14	1.70 *

Table A-8

Changes in SAT Reading and Math Scores, Fall 2012 to Spring 13, by Grade, Arthur Academy Charter Schools

		Reading			
	<u>Fall</u>	<u>Spring</u>	<u>ES Change</u>	<u>t-ratio</u>	
<u>First Grade (n=157)</u>					
Aver. or Above (>=40th %ile)	74	97	0.47	5.86	***
Top 20 (>=80th %ile)	24	47	0.58	7.18	***
<u>Second Grade (n=151)</u>					
Aver. or Above (>=40th %ile)	89	90	0.02	0.25	
Top 20 (>=80th %ile)	44	48	0.10	1.22	
		Mathematics			
<u>First Grade (n=157)</u>	<u>Fall</u>	<u>Spring</u>	<u>ES Change</u>	<u>t-ratio</u>	
Aver. or Above (>=40th %ile)	84	100	0.33	4.08	***
Top 20th %ile (>=20th %ile)	24	70	1.15	14.36	***
<u>Second Grade (n=151)</u>					
Aver. or Above (>=40th %ile)	92	92	0.00	0.00	
Top 20th %ile (>=20th %ile)	56	44	-0.30	-3.67	

Table A-9

Percentage of Students Meeting OAKS Benchmark, Arthur Students, Districts and State, 2012 and 2013, by Cohort and Subject

		Arthur Acad.	District	State	Ef. Size - Dist	Ef. Size- State	t-ratio - District		t-ratio - State	
	Year									
				Reading						
Cohort 1	2012	80	64	76	0.06	0.19	0.65		2.19	*
	2013	80	60	69						
Cohort 2	2012	80	60	72	0.17	0.23	2.07	*	2.91	**
	2013	88	64	75						
				Math						
Cohort 1	2012	64	56	67	0.11	0.26	1.28		3.05	**
	2013	66	52	59						
Cohort 2	2012	64	53	65	0.04	0.11	0.48		1.42	
	2013	68	56	65						

Note: For cohort 1 there were data for 146 students in 2012 and 128 students in 2013. For cohort 2 there were data for 165 students in 2012 and 146 students in 2013. Effect sizes and t-ratios were corrected for possible ceiling effects as explained in the text of the appendix.

*, p< .05; **, p< .01; ***,p<.001

References

Arthur, C. (2009-2010). *Mission and Instructional Model.* Portland, Oregon: Arthur Academy
   Public Charter Schools.

Coughlin, C. (2014). Outcomes of Engelmann's Direct Instruction: Research Syntheses, pp.
   25-54 in J. Stockard (Ed.). *The Science and Success of Engelmann's Direct
   Instruction.* Eugene, OR: NIFDI Press.