



UNIVERSITY OF ARKANSAS

College of Education and Health Professions
Department of Education Reform
207 Graduate Education Building
Fayetteville, AR 72701
Phone (479) 575-4971
Fax (479) 575-3196

Gary W. Ritter
Endowed Chair in Education Policy
garyr@uark.edu

Memo: ARVA evaluation
To: Scott Sides
From: Marty Lueken and Gary Ritter, University of Arkansas
Date: January 30, 2012

Internal Evaluation of the Arkansas Virtual Academy School

This analysis provides what we believe to be the easiest-to-understand employment of the data we have available for ascertaining the effectiveness of the Arkansas Virtual Academy (ARVA) school. We chose to employ a simple "nearest-neighbor" matching methodology out of an array of methods available. As with all the techniques available to us, this matching technique has strengths and weaknesses. This analysis represents a first take at addressing the question; we intend to employ a variety of other models and methods in order to check the robustness of the results presented in this memo.

Methods

One challenge in evaluating the effectiveness of a particular school is that each school serves different sets of students in different years. However, since ARVA has only been collecting data on its students since 2007-08, we are limited in which cohorts of students we can study. Furthermore, the only standardized data collection that we can rely on occurs for students in grades 3-8. Finally, we believe it is best to evaluate student achievement after students spend more than one year in a new school to temper "transition shocks" that may appear. **As a result, this preliminary analysis of the effectiveness of ARVA is based on the cohort of students who**

were in ARVA in grades 3-6 in the 2008-09 school year. These same students were in grades 5-8 in the 2010-11 school year (most recent available data).

A second, and more fundamental, challenge in any such analysis is choosing an appropriate *counterfactual* or comparison group; that is, we need a best estimate of what would have happened to the students if they had not attended ARVA. To meet this challenge, we employ matching methods to develop a comparison group of students against which we can judge the effectiveness of the Arkansas Virtual Academy Charter School. More specifically, we match each ARVA student in the cohort with two matched peer students from traditional public schools. To ensure that these matched peers represent a fair comparison group for the ARVA students, we choose peers from the same grade and the same resident school district as the ARVA student. Most importantly, the matched peer students are chosen to have nearly identical baseline (2008-09) test scores (within 0.2 standard deviations) in math and literacy. For comparing math outcomes, we place the greatest weight on baseline (spring 2009) math scores. For comparing literacy outcomes, we place the greatest weight on baseline (spring 2009) literacy scores. After ensuring that the students are similar in these categories, we also attempt to match (as best as possible) on socioeconomic status (FRL status), racial group (whether or not the student is a minority), and gender. However, we place the greatest importance on the academic achievement match as prior achievement is the single most important predictor of future academic achievement.

We drop any student that repeated any grade in the sample period for both groups – this consisted of four grade repeaters from the ARVA cohort. We also consider only students who have test scores in 2009 and 2011 (the first and last years in the sample period). Otherwise, we would not be able to ascertain growth in student performance over the two-year study period.

Equivalence of ARVA Students (Treatment Group) and Comparison Group

This strategy can only be effective if the treatment and matched comparison groups are similar at the initial data collection (baseline). Table 1 compares the two groups used in this analysis for each subject (mathematics and literacy). We match 185 ARVA students for the math analysis with 375 students from traditional public schools in Arkansas; for the literacy analysis, we match the 188 ARVA students with 366 comparison students. When matching on math outcomes (columns labeled “Math”), the ARVA group’s average baseline math score is slightly lower than that of the comparison group (about one-tenth of one percentile point difference). When matching on literacy, the ARVA group’s baseline literacy score is less than one percentile point below the comparison group’s score. It is important to note that the two groups are similar with respect to economic disadvantage (47 percent FRL v. 51 percent FRL in the math matched groups and 45 percent v. 49 percent in the literacy matched groups). On the other

hand, the comparison groups have significantly larger proportions of minority students than the treatment groups.

Overall, we are confident that our matching procedure generated a comparison group that will allow us to draw reasonable conclusions about the effectiveness of ARVA for the following reasons: the students are nearly perfectly matched on measures of prior achievement in math and literacy (the most important matching indicator), the students are pulled from identical grades and resident districts, and the groups are well-matched on our measure of socioeconomic disadvantage. Nevertheless, as an additional "check" to ensure accuracy, we will conduct all of our analyses using ordinary least squares multiple regression models that will control for important student characteristics including prior student test scores (2009) and student demographic characteristics (FRL, race, and gender). Thus, even slight differences between the groups on these measures will be controlled for in our final analyses.

Table 1: Baseline characteristics of ARVA students and matched kids, 2009 grades 3-6 cohort

	Matched pool (comparison group)		ARVA Students with matches	
	Math	Literacy	Math	Literacy
Sample size	375	366	188	185
Math z-scores percentile	-0.05 48 th	0.06 52 nd	-0.06 47 th	-0.07 47 th
Literacy z-scores percentile	0.03 51 st	0.10 54 th	0.07 53 rd	0.08 53 rd
FRL status (number)	190	178	88	84
(proportion)	50.7%	48.6%	46.8%	45.4%
Minority (number)	89	71	15	15
(proportion)	23.7%	19.4%	8.0%	8.1%
Female (number)	165	162	86	86
(proportion)	44.0%	44.3%	45.7%	46.5%

Outcome Measures Employed

After showing that the groups are similar in the baseline year of the analysis, we then ask whether differences emerged after the ARVA students spent two years in ARVA while their matched peers spent the same two years in traditional public schools. We examine the effect of attending ARVA on math and literacy outcomes. Specifically, we define student achievement as student performance on the Arkansas Benchmark criterion-referenced exams. We normalize these measures so that the distribution of scores has an average score of zero and a standard deviation of one. This normalized value is known as a z-score. This z-score measure allows the effect of ARVA to be expressed in standard deviations (known as an effect size). Perhaps the most understandable way of interpreting these z-scores is by describing the *percentile rank* of the student scores. For example, a positive z-score represents a score above the average score (*above the 50th percentile*) in Arkansas while a negative z-score indicates a score below the average score (*below the 50th percentile*). For illustration, a z-score of 0.5 (which may appear small to the casual observer) represents a student scoring well above the average in the state. This student, who scored at a level that is $\frac{1}{2}$ a standard deviation better than the average student in the state, would be ranked in the 70th percentile (this student's score is better than that of 70% of his peers in Arkansas).

Results

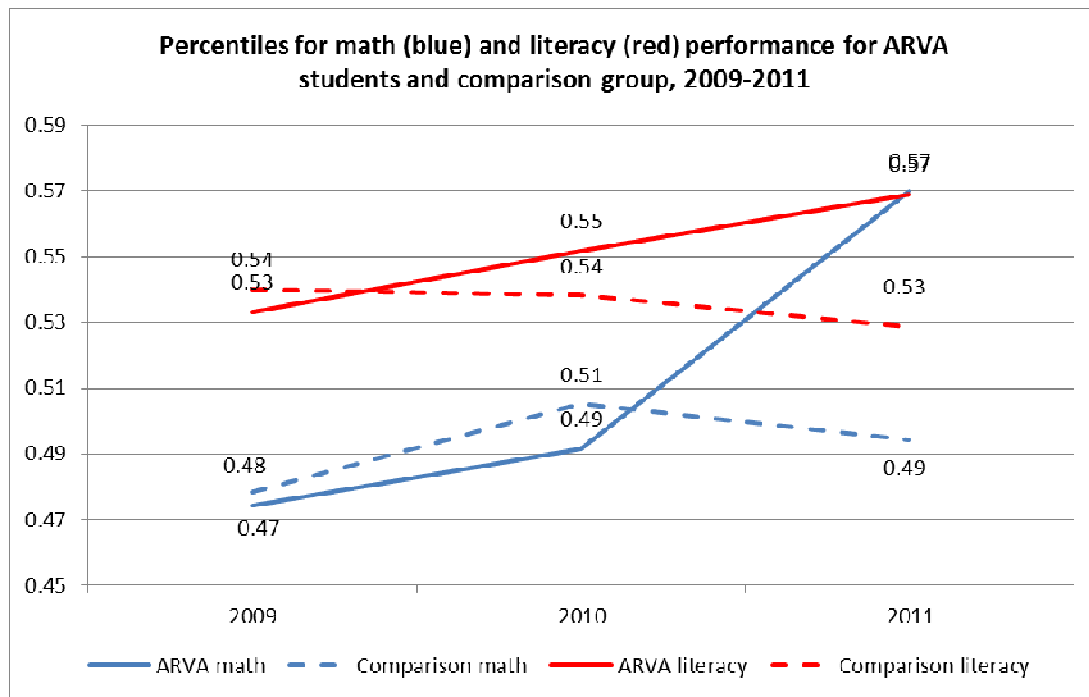
Thus far, we have shown that we will examine the student achievement on Arkansas Benchmark exams in math and literacy in 2011 for ARVA students and comparison students (these groups' benchmark exams in 2009 that were almost perfectly matched). The ARVA cohort averaged math scores about 0.06 standard deviations below the state average while students with comparable backgrounds averaged about 0.05 standard deviations below the state average. In other words, the average ARVA student started out in our study period (in 2009) at the 47th percentile while his/her matched peer scored at the 48th percentile in the baseline year. In literacy, the ARVA cohort scored at the 53rd percentile while its matched comparison group was at the 54th percentile.

Now that we have shown the equality of the two groups at baseline, we can examine whether the next two school years (2009-10 and 2010-11) lead to any improvements for ARVA students. Thus, we ask:

After two years of enrolment at ARVA (in 2011), how does the performance of ARVA students compare with that of similar peers attending schools in their resident districts?

What did we find? The average ARVA student from our cohort performed well in both math and literacy after two years (See Figure 1 below). The ARVA cohort improved significantly as compared to their matched peers. As shown in Table 2, ARVA students improved by .191 z-score points more than their matched peers in math and by .102 z-score points more than their matched peers in literacy. These differences are substantial and might be better understood as percentile ranks -- the average ARVA student jumped from the 47th to 57th percentile in math and from the 53rd to the 57th percentile in literacy. Thus, the average ARVA student increased about 9.6 and 3.6 percentile points over this period in math and literacy, respectively. The average student in the comparison group gained only 1.6 percentile points in math and decreased by 1.2 percentile points in literacy over the same period (Table 2).

Figure 1



The score gains presented above are based on raw differences in student test scores without any adjustment for student demographic differences. Thus, while this first analysis suggests that ARVA students experience much greater growth than do their comparison peers, we will now check the robustness of these findings with more sophisticated statistical models. To do this, we also estimate ARVA effects using regression models that control for background factors in case the overall averages are unduly influenced by any demographic differences between the two groups (Table 2, columns 8-11). Moreover, by estimating a regression model for achievement differences between these two groups that controls for individual student characteristics, we will generate a more precise estimate of the differences in 2011 student achievement that are due to enrollment at ARVA. The adjusted "ARVA effect" for 2011 is 0.126 and 0.085 z-scores in math and literacy, respectively. As we expected, these effects are slightly smaller than those that we found in the unadjusted analysis based in simple differences in means, but they remain in favor of the ARVA group. The math difference is statistically significant.

In sum, using our most cautious model that controls for all student differences, we estimate that, in 2011, the ARVA cohort, on average, scored at the 54th and 56th percentiles in math and literacy, respectively, as compared with the 49th and 53rd percentiles by the comparison group.

Table 2: Standardized math scores for ARVA students and comparison group, 2009-2011[†]

	Comparison group		ARVA		Difference in means in year t		Effect from OLS model		Regression adjusted in ARVA mean score *	
	Math	Literacy	Math	Literacy	Math	Literacy	Math	Literacy	Math	Literacy
2009	-0.054	0.101	-0.065	0.083	-0.011	-0.018	--	--	--	--
	48 th	54 th	47 th	53 rd	-1 pt	-1 pt	--	--	--	--
2010	0.013	0.096	-0.021	0.130	-0.034	0.034	-0.104	0.001	-0.091	0.097
	51 st	54 th	49 th	55 th	-1 pt	+1 pt	--	--	46 th	54 th
2011	-0.014	0.072	0.176	0.174	0.191	0.102	0.126#	0.085	0.111	0.157
	49 th	53 rd	57 th	57 th	+8 pts	+4 pts	--	--	54 th (+5)	56 th (+4)

[†]Scores are given in standard deviations with corresponding percentiles underneath.

*The regression adjusted ARVA mean score is computed by taking the overall outcome mean for the comparison group and adding the treatment group coefficient from the OLS regression model.

This effect is statistically significant at p=.10 level.

Summary of Overall Effects

Up to now, we have shown that ARVA students, overall, started at roughly the same level of their matched comparison peers in 2009 (a bit below the state average in math and a bit above the state average in literacy). By 2011, after 2 years of enrolment in ARVA, we find that the ARVA students outperformed their comparison peers in math by 5 percentile points. This math difference is statistically significant in favor of ARVA. In literacy, ARVA students experienced a growth of 3 percentile points while their comparison peers experienced a decline of one percentile point.

Next, we will examine these changes in more detail by looking at these performance changes for particular subgroups of students (by grade, by FRL status, and by minority status).

Subgroup Effects -- by Grade Cohort

We also break down the comparison by grade for each subject. Figures 2 and 3 plot the standardized scores over the sample period for each grade, and Tables 3 and 4 give the percentiles of the starting and ending scores for each grade. The solid lines in the figures represent grades in the ARVA cohort and the dotted lines correspond to their matches. The columns labeled "Difference" in the tables subtract the comparison group's percentile from the ARVA column. Thus, a negative (or positive) number indicates that students in ARVA, on average, experience negative (or positive) outcomes in the relevant subject relative to their matched comparisons. The columns "ARVA gain" and "Comparison group gain" represent these respective group's average gains over the two year period. The last column ("Diff-diff") provides

the difference in *gains* over the sample period between these two groups. For instance, .08 would indicate that the average ARVA student in the particular grade (indicated by the row) gained 8 percentile points more than the average comparison student over the two-year period. The last row (“Overall”) reproduces the overall results for the two groups.

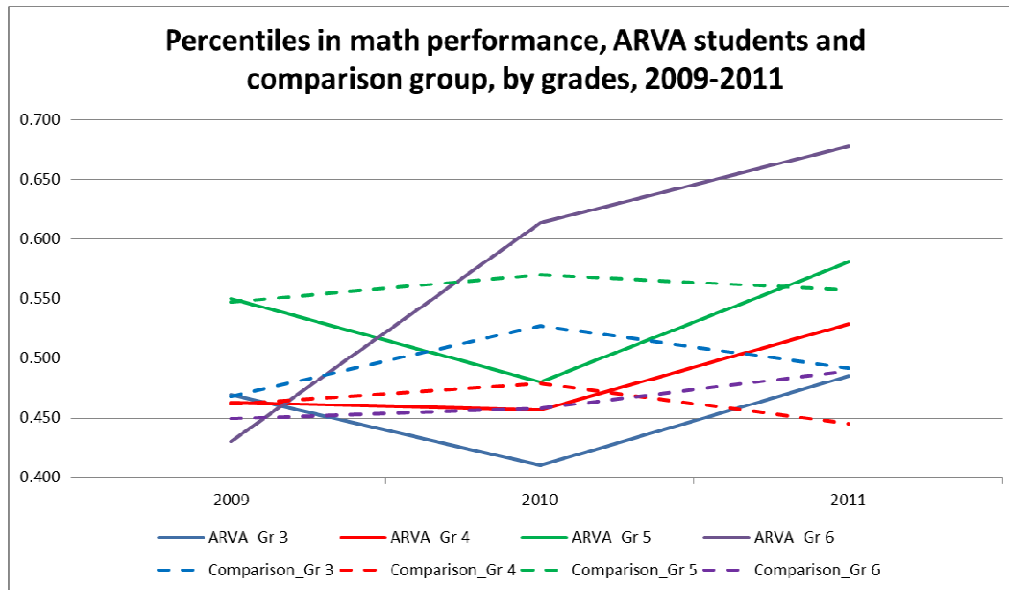
Math

Table 3 provides the starting and ending percentiles ranks for both groups by grade along with their differences in growth overall. The ARVA group’s spring 2011 percentile ranks are adjusted by ordinary least squares using the same regression model that we discussed above. When we disaggregate math results by grade, we find that the average student from both groups started at about the same level. For instance, 3rd graders from both groups started at the 47th percentile. Overall, ARVA students in 3rd grade grew 1 percentile point less than their counterparts. ARVA students in 4th and 5th grade grew 5 and 2 percentile points, respectively, greater than their matched comparisons, but these estimates are statistically insignificant. ARVA 6th graders grew 16 percentile points greater than their matched comparisons, and this result is statistically significant.

Table 3: Description and Mathematics Scores for Grade Level Cohorts of Students

	Gr 3 to 5	Gr 4 to 6	Gr 5 to 7	Gr 6 to 8
ARVA Group - Number of Students	46	46	41	55
Comparison Group - Number of Students	92	92	82	109
ARVA Group - Starting Percentile Rank	47th	46th	55th	43rd
Comparison Group - Starting Percentile Rank	47th	46th	55th	45th
ARVA Group - 2011 Percentile Rank (adjusted)	48th	50th	58th	63rd
Comparison Group - 2011 Percentile Rank	49th	45th	56th	49th
<i>Difference in Growth</i>	-1 pts	+5 pts	+2 pts	+ 16 pts
<i>Statistically Significant (p=.10) Effect of ARVA?</i>	<i>no</i>	<i>no</i>	<i>no</i>	<i>yes</i>

Figure 2



Literacy

Table 4 presents the same information as Table 3, but for literacy. Over the sample period, ARVA 3rd grade students experienced 3 percentile points less than their counterparts in literacy growth, but this result is statistically insignificant. ARVA 4th graders grew 11 percentile points more than their counterparts, and this result is statistically significant. ARVA 5th and 6th graders grew 2 and 7 percentile points more, respectively, than their matched comparisons, but these estimates were statistically insignificant.

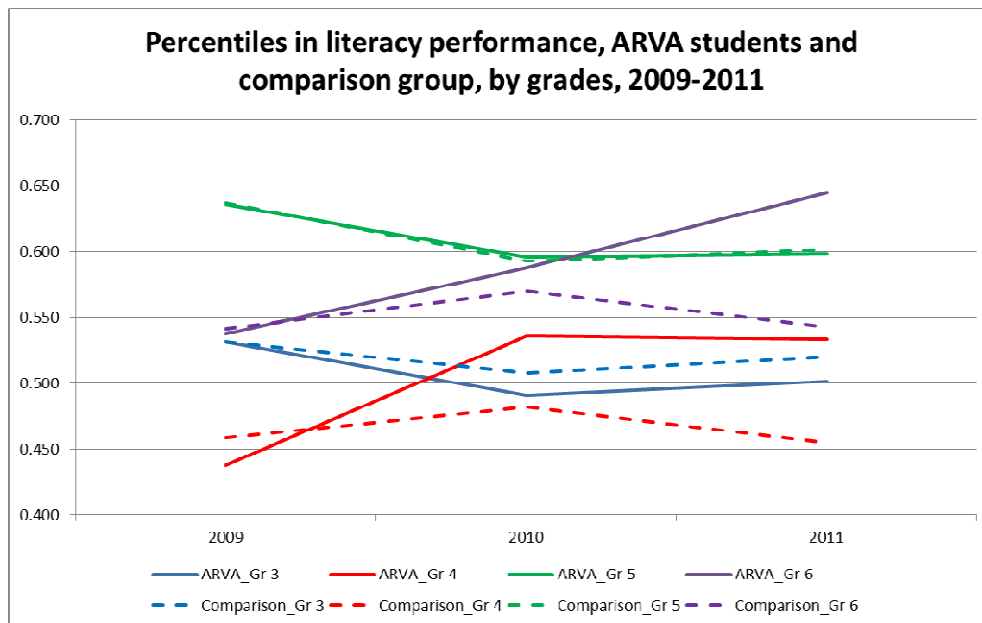
Overall, the average ARVA student made greater gains in literacy than the average comparison group student. Disaggregated by grade, the story becomes somewhat mixed. The average student from both groups started at very similar levels. Over a two-year school period, the average ARVA student in 4th and 6th grades in 2009 outgained students in their matched comparison groups in literacy (Figure 3). ARVA students in both grades gained roughly 10 percentile points while the average comparison group students from both grades made almost zero gains. The average 3rd grader from the ARVA and comparison groups dropped about 3 percentile points and 1.2 percentile points, respectively. The average 5th grader from both groups dropped about 3.5 percentile points.

Over the sample period, there was no difference in gains between the two groups' average 5th graders. ARVA students in 4th and 6th grade, on average, gained 9.9 and 10.6 percentile points, respectively, more than their counterparts. ARVA 3rd graders, on average, dropped about 1.8 percentile points relative to their counterparts over the sample period.

Table 4: Description and Literacy Scores for Grade Level Cohorts of Students

	Gr 3 to 5	Gr 4 to 6	Gr 5 to 7	Gr 6 to 8
ARVA Group - Number of Students	45	46	40	54
Comparison Group - Number of Students	90	90	80	106
ARVA Group - Starting Percentile Rank	53rd	44th	64th	54th
Comparison Group - Starting Percentile Rank	53rd	46th	64th	54th
ARVA Group - 2011 Percentile Rank (adjusted)	49th	55th	62nd	61st
Comparison Group - 2011 Percentile Rank	52nd	46th	60th	54th
<i>Difference in Growth</i>	-3 pts	+11 pts	+2 pts	+7 pts
<i>Statistically Significant (p=.10) Effect of ARVA?</i>	<i>no</i>	<i>yes</i>	<i>no</i>	<i>no</i>

Figure 3



Effects by FRL and minority status

In the last section, we examined the impacts that ARVA has on learning for students who are eligible for free-reduced lunch (FRL) programs and for minority students. The FRL variable represents the best available data to proxy for resource disadvantages. Of the 188 students we examined in the ARVA cohort on math and 185 students on literacy, 88 and 84, respectively, have FRL status (with 175 and 165 matches we compare); similarly, 15 students are minorities with 30 matches we compare (Table 5).

Table 5: Frequency table of sub-analyses, by FRL status and minority status

	FRL		Minority	
	math	lit	math	lit
ARVA	88	84	15	15
Comparison group	175	165	30	30
Frequency with status	131	150	21	20
Percent of comparisons with status	75%	91%	70%	67%

FRL

We now examine how ARVA students with FRL status compare with their matched counterparts. To be clear, we compare ARVA students who have FRL status *with their matches*, who may or may not have FRL status. Indeed, of the 175 and 165 matches for this group of ARVA students, 75% and 91% had FRL status on math and literacy, respectively (Table 5). Table 6 provides the starting and ending percentile ranks for both groups and their overall difference in growth. The ending (spring 2011) percentile rank for ARVA is adjusted with effect estimates from the same regression used in the first section.

In this analysis, we find that ARVA FRL students did particularly well in math. ARVA FRL students grew 8 percentile points higher than did their counterparts in math, and this difference is statistically significant. ARVA students grew 6 points higher than their counterparts in literacy, but this result is not statistically significant.

Table 6: Description and Mathematics and Literacy Scores for Free and Reduced Lunch Students

	Mathematics Sample	Literacy Sample
ARVA Group - Number of Students	88	84
Percent FRL	100%	100%
Comparison Group - Number of Students	175	165
Percent FRL	75%	91%
ARVA Group - Starting Percentile Rank	43rd	48th
Comparison Group - Starting Percentile Rank	44th	49th
ARVA Group - 2011 Percentile Rank (adj)	52nd	52nd
Comparison Group - 2011 Percentile Rank	45th	48th
<i>Difference in Growth</i>	+8 pts	+6 pts
<i>Statistically Significant (p=.10) Effect of ARVA?</i>	yes	no

Minority students

Finally we examine how minority students in ARVA compare with their matched comparisons. Again, we compare minority students in ARVA *with their matches*, who may or may not be a minority. Of the 30 matched comparisons for this ARVA group in the math and literacy analyses, 21 and 20 students, respectively, are minorities (Table 5). Table 7 presents the same information as Table 6, but it applies to ARVA minority students only. ARVA minority students grew 3 percentile points higher than their counterparts in math while they grew 8 points less than their counterparts in literacy. Both results, however, are statistically insignificant.

The results in this section on minority students should be read with a healthy dose of caution because the treatment sample sizes are quite small. Given the small sample size, we have less confidence that these estimates represent ARVA’s true impact on minority students (that is, statistical power decreases as sample sizes decrease). Before we draw any strong conclusions about the effectiveness of ARVA students, we will need to gather additional data on more cohorts of minority students. As of now, these conclusions are based on the experience of only 15 students.

Table 7: Description and Mathematics and Literacy Scores for Minority Students

	Mathematics Sample	Literacy Sample
ARVA Group - Number of Students	15	15
Percent minority	100%	100%
Comparison Group - Number of Students	30	30
Percent minority	70%	67%
ARVA Group - Starting Percentile Rank	33rd	45th
Comparison Group - Starting Percentile Rank	34th	45th
ARVA Group - 2011 Percentile Rank (adj)	39th	35th
Comparison Group - 2011 Percentile Rank	36th	44th
<i>Difference in Growth</i>	+3 pts	-8 pts
<i>Statistically Significant (p=.10) Effect of ARVA?</i>	<i>no</i>	<i>no</i>

Conclusion

This analysis employed nearest-neighbor matching to examine how students in the ARVA public charter schools perform on math and literacy compared to nearly identical peers educated in traditional public schools in Arkansas. We found very positive trends for ARVA.

- **Overall**, ARVA students outperformed their comparison peers in math; the differences are substantial and statistically significant.
- **Overall**, ARVA students outperformed their comparison peers in literacy; these differences were positive and substantial (+ 4 percentile points).
- These positive trends were apparent in nearly all grade cohorts, but were particularly driven by strong growth in math for the students in grade 6 in 2008-09 and in literacy for the students in grade 4 in 2008-09. These groups experienced large gains that were statistically significant.
- Economically disadvantaged students also did particularly well at ARVA; ARVA free and reduced lunch eligible students outperformed their comparison peers in math; the differences are substantial (+ 8 percentile points) and statistically significant. These differences in literacy were positive and substantial (+ 6 percentile points).
- There were no statistically significant negative effects in any of the numerous analyses conducted.