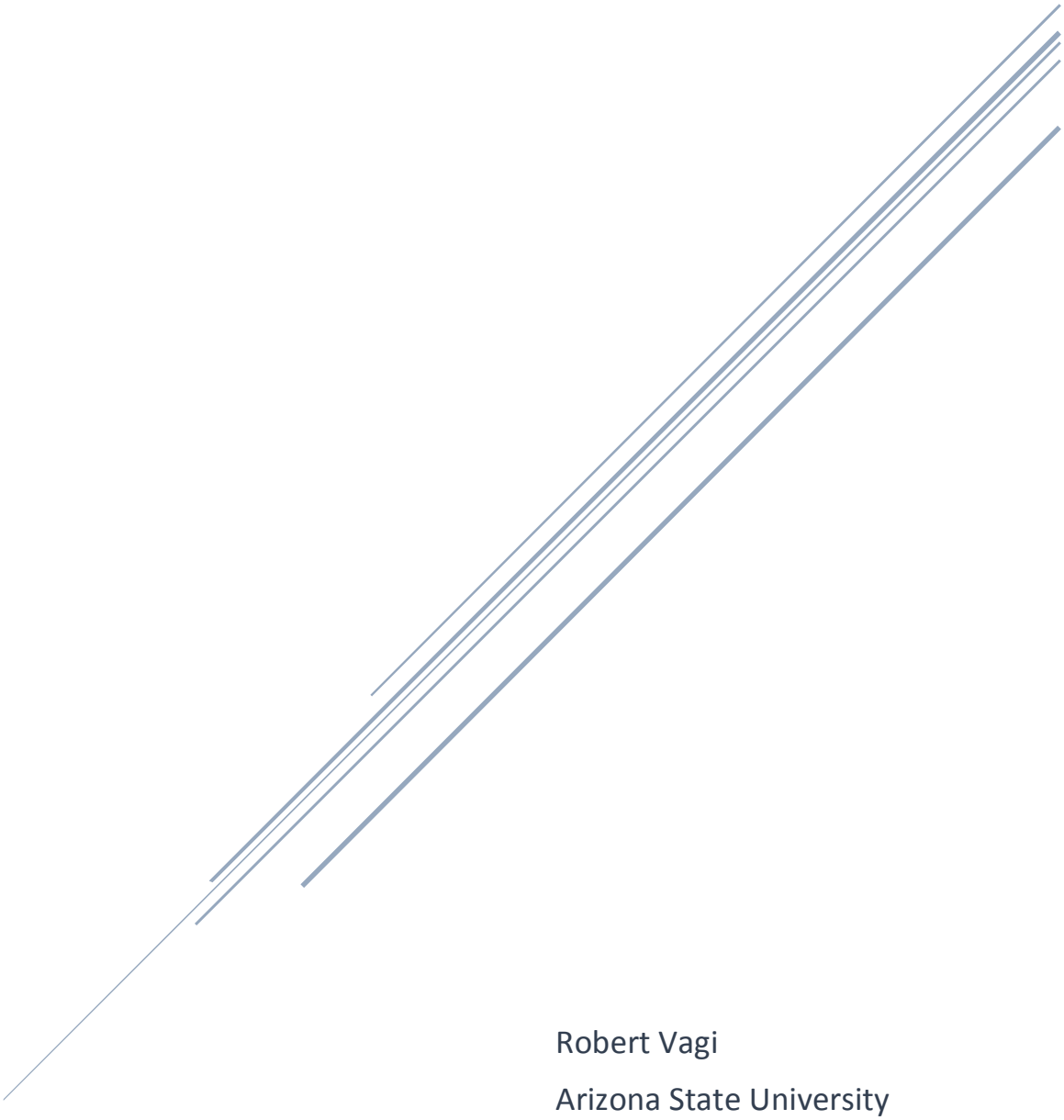
# FACTORS RELATED TO EARLY CHILDHOOD LITERACY

## An Exploratory Study



Robert Vagi

Arizona State University

## Introduction

As we enter the 21st Century, new technologies and increased access to information make literacy an indispensable skill. Therefore, it is crucial that students, particularly those in early grades, have access to high-quality reading and writing instruction. In support of this goal, the Arizona Legislature has passed several laws aimed at increasing early childhood literacy across the state. For instance, Arizona Revised Statute 15-704 requires school districts and charters to provide effective reading instruction as well as initial screenings for students, on-going diagnostic tests, and a system to monitor student progress. Similarly, Arizona Revised Statute 15-211 requires all school districts and charters with a K-3 program to submit a comprehensive plan for reading instruction and intervention across these grades. Finally, Arizona Revised Statute 15 – 701 states that if a student scored "falls far below" on the state reading test, the student will not be promoted until they have demonstrated sufficient progress toward proficiency.

In light of the growing emphasis on early childhood literacy, Read On Arizona is leading a statewide effort to improve language and literacy outcomes for all of Arizona's children, from birth to third grade. Read On Arizona is collaborating with public and private agencies, philanthropic and community organizations and citizens, to engage in literacy efforts at the community and state level. As a part of this effort, Read On Arizona has partnered with the Arizona Department of Education, the Arizona Department of Health, First Things First, and Head Start to discuss the factors that affect whether or not a child learns to read and write by the third grade. These partner agencies provided their applicable data to be used in a visual database of factors impacting literacy (e.g., school-, health-, and family-related factors that are correlated with third grade reading achievement). This paper presents the findings from a set of analyses using these data.

Based on prior research, there were relationships between the variables in this analysis and literacy that were expected (e.g., poverty), some relationships that were not expected (e.g., charter status), and other relationships that were expected but were statistically non-significant (e.g., health factors). With regard to this last group, we recommend that readers interpret the results with caution. Statistically non-significant results do not mean that these factors do not impact early childhood literacy. Rather, there may be relationships that these data and this analysis did not allow us to observe (something that is discussed in-depth in the conclusion). Therefore, we recommend that readers focus on those factors that are found to be significantly related to literacy and interpret non-significant results in light of prior research. The following sections describe the data, methods and findings, and implications.

## Data

Data for this study were identified and provided by members of the Read On Arizona data taskforce which includes the Arizona Department of Education, the Arizona Department of Health, First Things First, and Head Start. Because of the large amount of data included in these datasets and the exploratory nature of this analysis, it was important to examine relationships between variables before deciding which variables would be included in the analyses. Therefore,

variables were chosen based on the following criteria: a) their degree of correlation with third grade reading scores, b) their degree of correlation with other variables[1], and c) the completeness of the data for each variable.[2] Below is a list the variables that were included in the analyses and offer a brief description of each.

*School-Level Variables*

Percent Passing AIMS Reading: The percentage of 3rd grade students who scored at either the "Meets" or "Exceeds" achievement levels during the 2013-2014 school year.

Charter or District: An indicator for whether a school was a charter or district school during the 2013-2014 school year.

Enrollment: School enrollment on the last day of October during the 2013-2014 school year.

Percent Free and Reduced Lunch: The percentage of students that qualified for free- and reduced-price lunch during the 2013-2014 school year. This is a commonly-used measure of poverty.

3rd Grade Attendance Rate: The average 3rd grade attendance rate during the 2013-2014 school year. This percentage is calculated by dividing a schools average daily attendance (ADA) by its average daily membership (ADM).

2nd Grade Retention: The percentage of 2nd grade students that were retained during the 2011-2012 school year. 2011-2012 was chosen because students retained during this year would presumably be in the 3rd grade in 2013-2014.

Chronic Absence Rate: The percentage of students schoolwide who were absent for 18 or more days during the 2013-2014 school year.

Kindergarten Type: Indicators for whether a school offered full-day, half-day, or mixed kindergarten during the 2010-2011 school year. 2010-2011 was chosen because this is the year in which 3rd graders in 2013-2014 would have been enrolled in kindergarten.

---

[1] In regression analyses, high correlations among predictors make accurate estimates of the relationships between the variables and the outcome difficult. Therefore, in this analysis, when two predictors were highly correlated and at the same level (school, district, primary care area, etc.), the variable with the highest correlation with the outcome was included in the analysis.

[2] Data is masked when the number of students in a given category is small enough that these students might be identified individually. Masked data is most common among smaller schools as they are more likely to have only a handful of students in a given category (i.e. 3rd graders passing AIMS reading, retention in the 2nd grade, etc.). Many of the datasets contained large amounts of masked data. For instance, nearly 60% of schools had masked data for the percentage of at-risk 2nd graders in 2014.

Number of Preschools in Zip Code: The number of preschools in a given school's zip code in 2014.

*District-Level Variables*

Enrollment: Districtwide enrollment on the last day of October during the 2013-2014 school year.

Percent Free and Reduced Lunch: The percentage of students throughout each district that qualified for free- and reduced-price lunch during the 2013-2014 school year.

3rd Grade Attendance Rate: The average 3rd grade districtwide attendance rate during the 2013-2014 school year. This percentage is calculated by dividing each district's average daily attendance (ADA) by its average daily membership (ADM).

2nd Grade Retention: The percentage of 2nd grade students that were retained during the 2011-2012 school year throughout each district. 2011-2012 was chosen because students retained during this year would presumably be in the 3rd grade in 2013-2014.

Chronic Absence Rate: The percentage of students districtwide who were absent for 18 or more days during the 2013-2014 school year.

*Health Data*

Low Birth Weight: The percentage of low birth weight births in each primary care area in 2007-2008.

Preterm Births: The percentage of preterm births in each primary care area in 2007-2008.

Teen Births: The percentage of teen births per 1,000 women in each primary care area in 2007-2008.

Asthma: Rate of ER visits for 6 to 8 year olds with Asthma as the principal diagnosis per 10,000 ER visits in each primary care area in 2011-2012.

*First Things First Data*

This data reports the percentage responses for each of the following questions and response options in each First Things First region:

"During the past week, how many days did you or other family members read stories to your child/children?"

a) 1 to 4 days b) 5 days c) 6 or 7

During the past week, how many days did you or other family members tell stories or sing songs to your child/children?

a) 1 to 4 days  b) 5 days  c) 6 or 7

How many children's books - including library and e-books - do you have right now in your home?

a) 10 or fewer b) 11 to 100 c) 100 or more

## Methods and Findings

Because multiple schools may reside in the same school district, primary care area, or First Things First region, it is likely that schools in these regions have similar test scores due to factors at the higher level. For instance, if a district allots additional time for reading instruction while a neighboring district does not, it is logical to expect schools in the first district to have higher average reading scores. Therefore, a technique known as multilevel modeling was applied to account for this dependency. Specifically, cross-classified multilevel models were used since the schools in the data set lie within multiple regions that often do not overlap (i.e., two schools may be part of the same district but different primary care areas).

The analyses represent several higher level regions. Therefore, they are not well-suited to a single analysis. Instead, we examine relationships between school-, district-, health-, and family-related factors in two separate analyses. The first analysis examines schools in relation to school districts and primary care areas. The second examines schools in relation to school districts and First Things First regions. In each of these analyses, the outcome of interest is the percentage of students passing the AIMS reading test in the 3$^{rd}$ grade in 2014 at the school-level. It also important to note that only schools with all of the relevant data were included in the analysis (i.e. schools with masked data were excluded).

### Analysis 1: Schools, School Districts, and Primary Care Areas

Table 1 presents descriptive statistics for each variable in the analysis for all elementary schools in the state and for those that were included in the analysis. The table shows that the means and standard deviations of the variables in the final sample are nearly identical to those in the original dataset with the exception of the percentage of charter schools, the percentage of schools offering full-day and half-day kindergarten, and school and district enrollment. With regard to the change in the percentage of charter schools, charter schools are not required to report free-and-reduced price lunch data unless they participate in the federal free-and-reduced price lunch program. Therefore, many charter schools did not have this data and were excluded from the analysis. For school and district enrollments, schools were most often excluded because of masked data. Data is masked when the number of students in a given category is

| *Table 1: Descriptive Statistics* | | | | |
|---|---|---|---|---|
| | All Elementary Schools | | Elementary Schools in Final Sample | |
| Variable | Mean | Standard Deviation | Mean | Standard Deviation |
| Percent Passing AIMS Reading | 76.82 | 14.05 | 76.32 | 12.48 |
| Charter | .24 | | .10 | |
| Full Day Kindergarten | .77 | | .83 | |
| Half Day Kindergarten | .08 | | .04 | |
| Mixed Kindergarten | .13 | | .13 | |
| School Enrollment | 570.63 | 272.46 | 631.27 | 232.16 |
| School Percent FRL | 63.73 | 26.28 | 63.86 | 26.59 |
| School 3rd Grade Attendance Rate | 95.36 | 1.46 | 95.41 | 1.26 |
| School 2nd Grade Retention | 2.42 | 1.75 | 2.31 | 1.55 |
| School Chronic Absence Rate | 10.00 | 5.78 | 9.48 | 4.47 |
| Number of Preschools in Zip Code | 7.15 | 6.39 | 7.68 | 6.49 |
| District Enrollment | 17,329.35 | 19,093.18 | 20,507.13 | 19,350.34 |
| District Percent Free and Reduced Lunch | 62.74 | 20.82 | 61.57 | 21.30 |
| District 3rd Grade Attendance Rate | 95.32 | 1.19 | 95.37 | .91 |
| District 2nd Grade Retention Rate | 2.29 | 1.58 | 2.13 | 1.23 |
| District Chronic Absence Rate | 11.75 | 5.99 | 11.54 | 4.76 |
| Low Birth Weight | 7.14 | .97 | 7.13 | .82 |
| Preterm Births | 10.24 | 1.16 | 10.26 | 1.07 |
| Teen Births | 56.02 | 7.07 | 55.74 | 6.13 |
| Asthma | 466.30 | 167.94 | 492.89 | 152.53 |
| | N = 1,254 | | n = 758 | |

small enough that these students might be identified individually. Masked data is most common among smaller schools as they are more likely to have only a handful of students in a given category (i.e. 3rd graders passing AIMS reading, retention in the 2nd grade, etc.). Consequently, the average enrollment in our final sample is slightly higher than in the original files. Finally, differences in the percentages of kindergarten types could be related to the previously identified characteristics (i.e. if most of the small schools that were excluded also offered half day kindergarten), some unknown factor, or simply to chance. Regardless of the reason, it is important to note that results from these analyses are only applicable to schools that were included in the final sample.

The results of the analyses are presented in Table 2. This table contains the results of two analyses which indicate relationships between each variable and the percentage of students passing AIMS reading. They are summarized as what are known as partial regression coefficients. Partial regression coefficients represent the average increase in an outcome, in this case the percentage of students passing the AIMS reading test, for a one unit increase in each variable *independent of all other variables*. In other words, each coefficient represents the relationship between the variable and the outcome after statistically controlling for all other

| *Table 2: Partial Regression Coefficients* | | |
|---|---|---|
| | Grand-Mean Centered Analysis | Group-Mean Centered Analysis |
| Variable | | |
| Charter | 6.34***<br>(0.33) | 6.81***<br>(1.51) |
| Half Day Kindergarten | -0.69<br>(2.02) | 0.65<br>(3.20) |
| Mixed Kindergarten | 0.89<br>(1.43) | -0.54<br>(1.98) |
| School Enrollment | -0.00<br>(0.00) | -0.00<br>(0.00) |
| School Percent FRL | -0.30***<br>(0.02) | -0.31***<br>(0.02) |
| School 3rd Grade Attendance Rate | 1.49***<br>(0.33) | 1.36***<br>(0.34) |
| School 2nd Grade Retention | 0.23<br>(0.31) | 0.21<br>(0.31) |
| School Chronic Absence Rate | -0.27*<br>(0.12) | -0.31*<br>(0.14) |
| Number of Preschools in Zip Code | -0.00<br>(0.07) | 0.08<br>(0.08) |
| District Enrollment | 0.00<br>(0.00) | 0.00<br>(0.00) |
| District Percent Free and Reduced Lunch | -0.04<br>(0.04) | -0.35***<br>(0.03) |
| District 3rd Grade Attendance Rate | -0.04<br>(0.72) | 1.35*<br>(0.68) |
| District 2nd Grade Retention Rate | -0.75<br>(0.45) | -0.48<br>(0.33) |
| District Chronic Absence Rate | 0.29<br>(0.17) | 0.07<br>(0.15) |
| Low Birth Weight | -0.34<br>(0.59) | -.40<br>(0.60) |
| Preterm Births | -13.67<br>(47.32) | -15.83<br>(47.86) |
| Teen Births | -4.91<br>(9.62) | -6.13<br>(9.98) |
| Asthma | 0.00<br>(0.00) | 0.00<br>(0.00) |
| | n = 758 | n = 758 |
| Note: * *indicates p < .05, ** indicates p < .01, and *** indicates p < .001* | | |

variables. Therefore, if two variables are highly correlated with reading achievement *and* with each other, the coefficient for each variable will be lower than it would have been on its own because the shared relationship has been removed or "controlled" for. This makes the choice of centering important. When school-level variables are grand-mean centered, they are allowed to be correlated with district and primary care variables. Consequently, the results from this analysis show relationships that are independent of *all* other variables in the analysis. In the second analysis (group-mean centered), school-level variables are centered within their respective districts making them uncorrelated with district and primary care variables. In this analysis, school-level variables are allowed to be correlated with each other and district- and primary care-level variables are allowed to be correlated, but the relationships between levels has been removed.

Looking at the grand mean centered analysis, four coefficients were statistically significant at $p < .05$. In other words, these relationships are strong enough that they cannot be attributed to chance alone. These four variables are whether or not a school was a charter school, the percentage of students in poverty (FRL), third grade attendance rates, and the schoolwide chronic absence rate. The charter school coefficient indicates that, on average, charter schools had roughly 6% more students pass AIMS reading than district schools after controlling for all other factors (such as poverty, enrollment, etc.). The coefficient for percent FRL shows that for every 1% increase in the percentage of students in poverty, there is an average decrease of .3% in the number of students passing AIMS reading. With regard to the attendance rates, a 1% increase in attendance rate is associated with an average increase of 1.5% of students passing AIMS reading. Finally, the coefficient for chronic absence rate indicates that a 1% increase in chronic absenteeism is associated with a .3% decrease in the number of students passing AIMS reading.

As previously mentioned for the group-mean centered analysis, when variables are centered about their grand means, level 1 (e.g., schools) and level 2 variables (e.g., districts and primary care areas) are correlated. Therefore, level 2 variables will only be statistically significant if their relationships with the outcome are independent of level 1. In other words, the relationships between district and primary care data and school-level reading scores would have to be significant above and beyond other school-level variables. However, in the first analysis, only school-level variables were statistically significant. Therefore, the district- and primary care-level variables included in our analysis are not significantly related to school-level reading scores after we account for school-level variables. Despite these non-significant results, it may be useful to know if any of our district- or primary care-level variables are related to reading achievement independent of school-level variables. To do this, we conduct an analysis where school-level variables are centered at their group means (in this case, their respective districts). As mentioned before, this ensures that school-level variables are uncorrelated with district- and primary care-level variables. However, it is important to note that since districts and primary care areas are both considered level-2 variables, they are correlated with each other making the regression coefficients partial, but still uncorrelated with school-level variables.

As Table 2 shows, the results of the group-mean centered analysis are similar to those of the grand-mean centered analysis with two notable exceptions: district percent FRL and district third grade attendance rates. Much like we saw at the school-level, the coefficient for district

percent FRL indicates that a 1% increase in the district-level poverty rate is associated with .4% decrease in the number of students passing AIMS reading at the school-level. Similarly, a 1% increase in district-level attendance rates is associated with a 1.4% increase in the percentage of students passing AIMS reading at the school-level. These results are not surprising given the strong relationships observed in the grand-mean centered analysis.

**Analysis 2: Schools, Districts, and First Things First Regions**

In order to examine possible relationships between parental involvement and student achievement, we used the same models as in the first analysis but with First Things First variables replacing primary care area variables. Because all of the First Things First variables were highly correlated with each other ($r$ =.2-.8), we ran a series of analyses that included only one First Things First variable in each analysis. The results of these analyses were nearly identical to the results from the primary care area analysis. When school-level variables were grand-mean centered, whether or not a school was a charter school, school-level poverty, school-level attendance rates, and school-level chronic absenteeism were all significantly related to the percentage of third grade students passing AIMS reading at the school-level. As before, the group-mean centered analysis showed that in addition to the previously-identified school-level variables, district-level poverty and district-level attendance rates were significantly related to school-level third grade reading achievement. None of the analyses showed a significant relationship between parent involvement and student achievement independent of the other variables in the analysis.

## Implications

First and foremost, it is important to point out that the relationships described in this report are correlational, not causal. In other words, one cannot conclude with any degree of certainty that these variables caused changes in school-level reading scores. Further, this analysis is exploratory in nature, meaning that the theoretical underpinnings of these relationships can only be speculated. That said, four school-level variables consistently showed a statistically significant relationship across all of the analyses: whether or not a school is a charter school, percent poverty, third grade attendance rates, and chronic absenteeism.

With regard to charter schools, this relationship should be interpreted in light of two caveats. First, a significant number of charter schools were excluded from the final sample. This was due mostly to missing free-and-reduced price lunch data. In speculation, it is likely that the charter schools in the sample, on average, serve higher percentages of poor students since only schools that participate in the federal free and reduced price lunch program are required to report this data. Although this does not affect the results of the analysis, it is important to remember that the findings only apply to the schools in the final sample. If all charter schools were included in the analysis, a different relationship might be observed. Second, charter schools are subject to a rigorous review process where low-performing schools are routinely closed. It should come as no surprise, then, that charter schools in the sample perform higher, on average, than district schools simply for the fact that consistently low-performing charter schools

are not allowed to enroll students. Still, the difference between charter and district schools in the sample is quite large and is likely due to factors other than those mentioned above.

It should come as no surprise to those who study education that poverty is negatively correlated with student achievement. A long-standing body of research indicates that children from poor families perform below their affluent peers. Although the exact mechanisms that link poverty to lower academic achievement are the topic of much debate, the relationship is strong and well-documented.

Finally, analysis shows that third grade attendance rates and chronic absenteeism are strongly related to third grade reading scores even after controlling for poverty. Because the correlation between poverty and student achievement is so high in this data set ($r = -.71$) and because both attendance and absenteeism were correlated with poverty ($r = -.35$ and $.49$, respectively), the fact that these remained statistically significant is noteworthy. Again, it is important to state that these relationships may not be causal in nature. For instance, a third unobserved variable might be driving both higher attendance rates and higher student achievement. However, it seems logical that students who spend more time in school are more likely to learn.

Finally, some may interpret the fact that health and parental involvement variables were not statistically significant to mean that these factors do not influence student achievement. This, however, would be misguided. The relationships observed in these analyses are partial relationships; they reflect relationships independent of all other variables. The fact that these variables were not statistically significant could reflect the fact that they are correlated with other variables included in the model and, therefore, are redundant. For instance, students who are afflicted with chronic asthma may be forced to miss more days of school than their peers which results in lower levels of learning. Since our analyses show such a strong relationship between attendance rates and student achievement, it could be that the relationship between asthma and student achievement is already captured by attendance. With regard to the First Things First variables, it should be noted that First Things First regions are fairly large and, therefore, fewer in number than districts or primary care areas. Having fewer regions results in a loss of statistical power. In other words, a relationship must be large for it be statistically significant. Further, the correlations between the First Things First variables and school-level reading scores were roughly the same as the correlations between First Things First variables and district-level FRL. As with asthma and attendance, it could be that the relationship between First Things First variables and student achievement was already captured by district-level FRL. Finally, these statistically nonsignificant results should be understood in light of the fact that they are average relationships. In other words, some students may be affected by these factors while others are not. Whatever the case, these analyses are only a first step towards understanding the complex relationships that shape student learning and further research is needed before definitive statements about these relationships can be made.