Secondary student achievement, particularly in science, related to school size, student socio-economic status, per pupil expenditures, student mobility rate, and percentage of non-white students: A modified best-evidence synthesis

Barton, Jerrod
Texas A&M University, USA (jbarton@rma.tx.org)

Irby, Beverly J. ✉
Texas A&M University, USA (irbyb@tamu.edu)

Tong, Fuhui
Texas A&M University, USA (Fuhuitong@tamu.edu)

Torres, Mario
Texas A&M University, USA (mstorres@tamu.edu)

Sandlin, Judy
Texas A&M University, USA (jrsandlin@tamu.edu)

Abstract

The purpose of this study was to conduct a best evidence synthesis of secondary student achievement and science and as related to school size, student socio-economic status, per pupil expenditures, student mobility rate, and percentage of non-White students. Due to the lack of studies inclusive of any, a combination of, or all of those variables, we modified the best evidence process to include student achievement, encompassing of science as available, on standardized tests in U.S. schools. We determined this topic to be important for a number of reasons (a) the focus on science instruction over the years with changes in policies nationally and (b) the emphasis in Texas, in particular, on the biology exam which is a high stakes assessment (one of five exams needed to be passed in order to exit high school). In this paper, we share first the background of the study, the review methods, the results, and a discussion for further study.

Keywords: secondary student achievement; biology; state exams; economic status; mobility rate; per pupil expenditure

Secondary student achievement, particularly in science, related to school size, student socio-economic status, per pupil expenditures, student mobility rate, and percentage of non-white students: A modified best-evidence synthesis

1. Introduction

The importance placed on science instruction has changed drastically to become more thorough throughout the years (107th Congress of the United States of America, 2001a; American Association for the Advancement of Science, 1989, 1993, 2013; Jackson & Ash, 2012). The American Association for the Advancement of Science (AAAS) launched Project 2061 in 1985 (American Association for the Advancement of Science, 2013). Project 2061 is a long-term reform initiative designed to help all Americans become more literate in science, math, and technology. This was the beginning of the impetus in America to ensure that science instruction would change for the better. *Science for all Americans* was published by the AAAS in 1989 (American Association for the Advancement of Science, 1989). That report "*defined science literacy and provided the groundwork for national science-education standards by outlining what students should know and be able to do in science by high school graduation*" (Jackson & Ash, 2012, p. 724). In 1993, *Benchmarks for Science Literacy* was developed by the AAAS and included learning goals that aligned with *Science for all Americans* (American Association for the Advancement of Science, 1993). The purpose of that report was to help educators develop a scope and sequence for science classes. The National Research Council (NRC) published the *National Science Educational Standards* in 1996 (National Research Council, 1996). According to the National Research Council (1996), the purpose of these standards was to "spell out a vision of science education that will make scientific literacy for all a reality in the twenty-first century. They pointed toward a destination and provide a roadmap for how to get there" (p. ix). All of these changes in the field of science led to changes in the importance level that educators placed on the field (Jackson & Ash, 2012).

The next major change in the field of education that took place was the passage of the No Child Left Behind (NCLB) Act of 2001 (107th Congress of the United States of America, 2001). This bill "*requires that educators measure students' yearly progress, encourages high academic standards, and implements greater accountability throughout the nation's school system*" (Jackson & Ash, 2012, p. 724). This bill forced all states to assess students' science understanding on all levels in school yearly. As a result of the passage of NCLB, Texas made Science tested every year in the fifth and eighth grades (Texas Education Agency, 2013a). Also, in the 2011-2012 school year the State of Texas Assessments of Academic Readiness (STAAR) Biology exam became one of the science exams that all high school students in Texas have to pass in order to graduate from high school (Texas Education Agency, 2013d). This placed an added amount of pressure on all high school students in the state of Texas in regard to the science curriculum.

The National Research Council of the National Academies established a committee to develop K-12 science educational standards and in 2012 they published *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (National Research Council, 2012). This project helped to begin to align the science standards for the United States. The committee involved 41 individuals from 26 states. As a result of this, in April of 2013 the *Next Generation Science Standards* were published (National Research Council, 2013). These standards focused on developing students into college and career-ready levels of science preparedness. The standards were ready for states to begin using them for the 2013-2014 school year (National Research Council, 2013).

In the State of Texas, House Bill 5 was passed in 2013. As a result of this, the STAAR Biology exam became the *only required* science exam that *all* high school students are required to pass in order to meet graduation requirements (Texas Education Agency, 2013c). In the 2010-2011 school year 76% of all students in

the state of Texas passed the tenth grade Science TAKS test (Texas Education Agency, 2012). On the same test, 66% of all Hispanic students, 62% of all African American students, 64% of all economically-disadvantaged students, lastly, 87% of all White students passed the test for a difference of 13% when compared to all students (Texas Education Agency, 2012a). In the 2011-2012 and 2012-2103 school years, all 86% of all students in the state of Texas passed the STAAR Biology exam (Texas Education Agency', 2013a, 2013b). 81% of all Hispanic students, 82% of all African American, 82% of all economically-challenged students passed the same test, and 94% of all White students passed the same test (Texas Education Agency, 2013a, 2013b). The fact that the exam must be passed in order to graduate from high school, coupled with the point that students who are Hispanic, African American, and economically challenged consistently perform below their White counterparts, demonstrate a need for school personnel to assist these students to improve their science performance (Research Center, 2011; Robinson et al., 2008).

2. Review Methods

The best evidence synthesis (Slavin, 1986; 1987) approach was selected to review the literature in student achievement and science at the secondary level; however, there were insufficient numbers of studies aligned with specific variables to conduct a best evidence synthesis. Therefore, that approach was slightly modified in order to align with as many of the assumptions or principles of best evidence synthesis and in order to take a systematic approach to our review of literature. The systematic approach was suggested by Slavin (1986, 1987). When discussing the development of the best evidence process, Slavin, (1986) said, "the main idea behind this procedure, which is also called 'best-evidence synthesis,' is to add to the traditional scholarly literature review application of rational, systematic methods of selecting studies to be included and use of effect size (rather than statistical significance alone) as a common metric for treatment effects" (p. 7).

Since there was insufficient information for all the studies to determine the effect sizes, and since there were very few studies per variable, this systematic process was modified by including as many of the principles of best-evidence synthesis as promoted by Slavin. First, the most important principle of inclusion was included, which was that the published research must have germane-ness to the issue at hand (Slavin, 1986). In general, all studies included in the modified best-evidence synthesis contained information that was directly related to student testing. The second principle used was the methodological adequacy of the studies to ensure that bias was minimized (Slavin, 1986); effect size was used when it could be determined. The last principle used was a determination of the internal and external validity for each of the studies included (Slavin, 1986).

2.1 Background

Criteria for Study Inclusion - According to Slavin (1987), studies on which a review is based must meet a set of *a priori* criteria with respect to germaneness and methodological adequacy. All studies in this review are directly related to student performance in American public schools grades 9-12 and were quantitative. There were some studies that were mixed-method, but only the quantitative data were included in the modified best-evidence synthesis. The studies were limited based on year of publication (1990-2014). Only studies that were conducted since 1990 were used in order to have more current and relevant research. Some of these studies were performed at all levels, but only the high school portion was used for the modified best-evidence synthesis. We searched dissertations, published research studies, and other technical reports documenting student performance with regard to science.

Literature Search Procedures - Both quantitative and qualitative research was reviewed for the study, though it was decided not to include qualitative studies for this review due to the fact that Slavin's best-evidence synthesis did not include qualitative studies: (a) Education Full Text (H. W. Wilson), (b) Eric EBSCO, (c) Google Scholar, (d) ProQuest Dissertations and Theses Full Text, and (e) JSTOR. There was a total of 15,136 articles found when searching the above databases. When searching in Google Scholar for the term, "Science Achievement," a total 29,500 citations were available from across the years. Afterward, the search was modified

to only include citations from the years 1990-2014. That narrowed the search to only include 16,900 citations. The terms "STAAR Biology," "School Size," "Economic Status," "Expenditure," "Mobility Rate," "Non-White students," "K-12," and "Case Study" were added to produce 32,036 total citations as shown in Figure 1. After these citations, the repeated citations were removed. Citations from outside of the United States were removed, as well as those that were qualitative in nature. Additionally, every attempt was made to acquire an all-inclusive group of the published studies that met the criteria outlined above. Please note that a specific limitation of this study is that the specific number of studies on student achievement related to school size and other specific variables are very limited; therefore, though best-evidence synthesis guidelines were used, more studies would have to be available to be included to be a valid best evidence synthesis. After all studies were filtered, 14 studies met the criterion that was set forth.
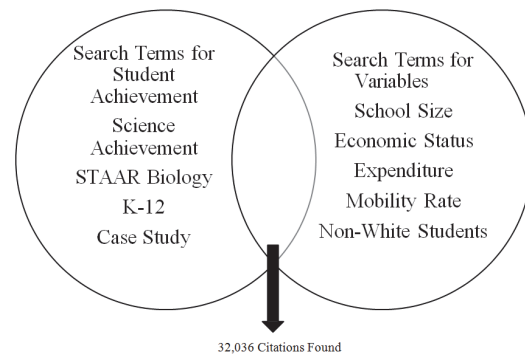

Figure 1. Citations found in the review.

3. Factors Impacting Student Achievement

Information for this modified best-evidence synthesis was found in relation to school size, student socio-economic status, per-pupil expenditures, student mobility rate, and percentage of non-White students. We reviewed the literature for those specific variables related to or impacting the performance of secondary students on standardized tests.

3.1 School Size Related to Student Achievement

From the 26,036 studies found based on the search, there were 635 studies that related to school size and student achievement; however, only three published studies were found that met the requirements to be included in our study. The requirements for inclusion were: (a) all studies are directly related to student performance in American public schools grades 9-12 (b) every study that was used in the research was quantitative in nature. There were some studies that were mixed-method, but only the quantitative data were included in the modified best-evidence synthesis and (c) each study was from the years 1990-2014. The screening process is shown in Figure 2 followed by the three studies that met the criteria.
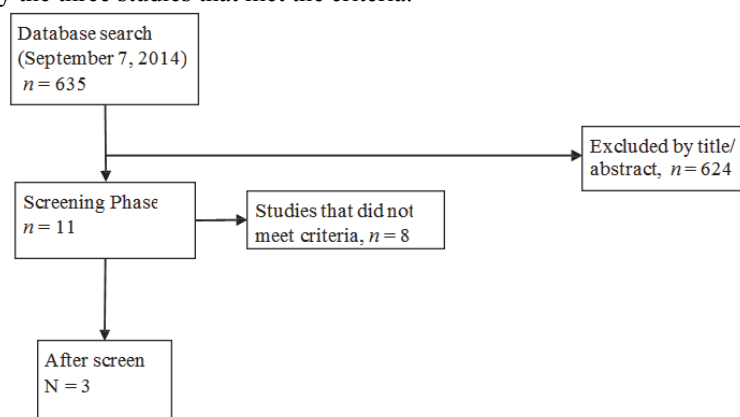

Figure 2. Screening of studies related to school size and achievement

New Jersey Study - Fowler and Walberg (1991) performed a study using data from the New Jersey Department of Education, Bureau of Information Services, and the Bureau of Testing for the 1984-1985 school

year. During the 1984-1985 school year, school-level test results were only available for schools that had ninth grade students enrolled in their campus. There were 332 schools that included ninth graders, and test data were only available for 293 of those schools. The variables included in the Fowler and Walberg study were (a) percentage of students from low-income families/district socioeconomic status, (b)school size and number of schools in each district, (c) teacher characteristics, including salary, degree type, and years of experience, and (d) school outcomes. The researchers investigated the potential influence of 18 social, organizational, and financial variables on 23 learning and related outcomes. The researchers used Statistical Analysis System (SAS) to run a series of regression equations (Fowler & Walberg, 1991). They first entered all possible independent variables and then dropped the variable least closely associated with the outcome, with the other variables held constant. The next least closely associated variable was then dropped, and this procedure continued until all remaining variables were significant at the standard .05 level of significance. This procedure allowed the most consistent variables that are most closely associated with learning to be identified. Socioeconomic status of schools and of districts was allowed to compete with the organizational and financial variables in order to examine their relative contributions to learning, particularly in combination.

The results of this study revealed that seven of the variables were significantly associated with school size: (a) percentage of low-income students in a school, (b) size of school, (c) number of schools in the district, (d) percentage of teachers with a bachelor's degree, (e) pupil-teacher ratio, (f) average teacher salary, and (g) district socio-economic status (SES). The results indicated that the district socioeconomic status was the most consistent of the variables and was significantly and positively associated with 17 of the 18 outcomes. The researchers suggested that smaller school district and smaller schools are more successful in regard to student educational outcomes. One of the limitations that this study has in regard to the best evidence synthesis is that there is no mention of the effect size of the study. The researchers' results revealed that the larger the school size was, the lower the students' achievement test scores were.

North Carolina Study - McMillen, Zhang, Cobb, Williamson, Kaase, Williams, and Fiefs (2000) performed a study that analyzed test scores for Kindergarten through fifth grade schools, sixth to eighth grade schools and ninth to twelfth grade schools in the state of North Carolina. For the purpose of our study, we examined the portion about the ninth to twelfth grade schools. The researchers used North Carolina End of Course (EOC) test scores from the 1998-1999 school year in Algebra I, English I, U. S. History, Biology I, and Economic, Legal, and Political Systems. The researchers grouped the high schools for the study by membership. There were 292 secondary schools, grades 9-12, used for their study. The groups analyzed were: schools with less than 700 students, schools with 700-1000 students, schools with 1001-1500 students, and lastly, schools with more than 1500 students. The variables included in the study were percentage of students enrolled at the school who were non-White, the percentage of students who were eligible for free or reduced price lunch, and the percentage of students whose parents had no formal education beyond high school. The school size groups were compared by using the 1998-1999 standardized EOC test scores in all five subjects.

The researchers analyzed the data utilizing a correlation analysis. This article was developed for the State Board of Education in North Carolina and the researchers do not specifically report their methods for this study. This is a major limitation of the study. The results of the study indicate that there were no differences in test scores on any of the five tests based on school size. The achievement test scores were virtually the same regardless of school size. There was a statistically significant interaction found between larger school size and lower student achievement. This connection was more prevalent in campuses where a larger percentage of students were eligible for free and reduced lunch. The researchers also stated that school size overlaps with many other variables that can impact student performance, so they claimed that it is very difficult to ascertain the true impact that school size has on student performance. The researchers felt that it is problematic to determine which variables are the ones that impact the relationship between school size and student performance (McMillen et al., 2000). They also found that there was a statistically significant relationship between student achievement and larger school size. The data that the researchers studied illustrated that the schools with the larger populations had lower student achievement.

Texas Study - Stewart (2009) conducted a study in which high school size in the state of Texas and student performance on the Texas Assessment of Knowledge and Skills (TAKS) were investigated. Stewart compared high school by sizes that are set by the University Interscholastic League (UIL). There are five categories, and they are as follows: 1A high schools have less than 195 students; 2A high schools have 195 to 414 students; 3A high schools have between 415 and 949 students; 4A high schools have between 950 and 1984, and 5A high schools have 1,985 students or more in their high school. Any high school not labeled a Disciplinary Alternative Education Program (DAEP), Alternative Education Program (AEP), Juvenile Justice Alternative Education Program (JJAEP), or charter school was included in the study.

This data are from the 2005-2006 school year. The researchers investigated the schools' performance on the TAKS test and determined what size of high school had the highest percentage of eleventh graders passing all four sections of the TAKS test. The four sections that the eleventh graders take are reading, writing, math, and science. The researchers conducted quantitative research by conducting a non-experimental, ex post facto design (Stewart, 2009). They used the Statistical Package for the Social Sciences (SPSS) program version 15.0 to perform the one-way ANOVA and the Scheffe analysis. They set the alpha level at .05 level of significance. The researchers used mean test scores of each school within each of the SES quartiles. The researchers conducted the study to determine if there was a relationship between student achievement in Texas (measured by Texas Assessment of Knowledge and Skills (TAKS) scores) and the size of the high school at different socioeconomic levels (Stewart, 2009). The researchers did not manipulate any of the data at any time.

Overall, the data from the study supported that students from the smaller schools are scoring statistically higher than the students at the larger schools. The results demonstrate that there are differences in student performance based on school size and the quartile that the students fall in. In the first quartile that includes less than 25% of economically disadvantaged students the 5A schools have a mean average of 81.71% passing while the 4A schools have a 77.22% mean passing percentage. In the same quartile, 3A schools have a mean passing percent of 78.81, 2A school have a mean passing percentage of 78.70, and the 1A schools have a mean passing percent of 72.5%. In this quartile overall the 5A schools perform statistically better than all of the other classifications with there being an almost 10 percentage point difference between 5A and 1A. In the other three quartiles, the opposite holds true. The 1A and 2A schools perform statistically better than the 5A schools when comparing mean test scores. This exposes that eleventh grade students in the smaller schools performed better on the TAKS test during the 2005-2006 school year. The researchers demonstrated that schools that are smaller have a statistically higher percentage of students passing all four sections of the TAKS test during the 2005-2006 school year.

School Size Related to Student Achievement Studies Summary - It is important to note that in all the three articles that were found that were related to school size and student achievement, and all three noted that students that attend smaller school perform better in school. These articles had different types of methodologies that included quantitative and mixed-methods research. None of the three articles listed effect sizes. Student socio-economic status was a variable that was suggested as one that could have impacted student performance as well as school size in these studies.

3.2 Socio-economic Status Related to Student Achievement

There were 26,036 studies found based on the search, there were 1,714 studies that related to socio-economic status and student achievement. Only one study met the requirements to be included in our study, but was related to math mainly (we did include it as there were no other studies found meeting all the criteria we had specified). The requirements for inclusion were: (a) all studies are directly related to student performance in American public schools grades 9- 12 (b) Every study that was used for the research was quantitative in nature. There were some studies that were mixed-method, but only the quantitative data were included in the modified best-evidence synthesis and (c) each study was from the years 1990-2014. The screening phase is shown in Figure 3. The one study is reported as follows.
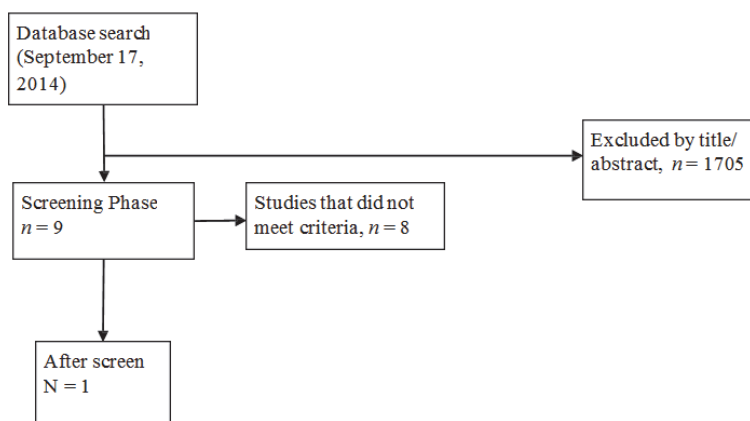
Figure 3. Screening of studies related to socioeconomics and achievement

Louisiana Study - Caldas and Bankston (1997) conducted a study in Louisiana that investigated the relationship between socioeconomic status and individual student achievement. The researchers used the mathematics, English language arts, and written composition components of the Louisiana Graduation Exit Examination (GEE) to measure student achievement. The tests used for this study are the ones that were given to tenth graders in 1990. They used test score data from 42,041 students in the State of Louisiana. The researchers performed a principal-component analysis on the raw mathematics, language arts, and written composition scores and they used the output of the weighted factor as the dependent variable of student achievement, school-level measures of SES, individual- level control variables, and school-level control variable.

The researchers conducted a "series of OLS regressions in four steps to show the effect of adding school-level variables into the model while reporting standardized regression coefficients to indicate the relative effect of each variable on school achievement" (Caldas & Bankston, 1997, p. 272). Academic achievement was regressed against all of the individual-levels in the first step. On the second step, the school-level poverty status was added to the model. On the third step, the school-level family social status variable was added. On the last step, the researchers included the school-level variable for the racial composition. The results from the study show that the highest correlation between individual-level and school-level variables was between, what the researchers called, minority race and percentage of minority students in the schools ($r = .606$). In the portion of the study in which only individual-level variables were included, the poverty status of individual students had a statistically significant negative effect on academic achievement ($ß = -.069$). Minority race (what we would rather call students of color) had an even higher effect on student achievement ($ß = -.314$). Another variable that had an effect on student achievement was family social status ($ß = .171$). There were other variables that the researchers used in the study, but these are the ones that directly related to this modified best-evidence synthesis, so we did not include other results.

Socio-economic Status Related to Student Achievement Study Summary - There were many articles that related to socio-economic status and student achievement that we found. Only one of them came relatively close to meeting the criteria of our modified best-evidence synthesis. Most of the articles related to K-12 education, not just high schools as is the case in this study.

3.3 Per Pupil Expenditure Related to Student Achievement

Among the 26,036 original studies that were discovered in the initial search, only two published studies were found that related to per pupil expenditure and student achievement. These four studies met the requirements to be included in our study. The requirements for inclusion were: (a) all studies are directly related to student performance in American public schools grades 9-12 (b).There were some studies that were mixed-method, but only the quantitative data were included in the modified best- evidence synthesis and (c) each study was from the years 1990-2014. The screening phase is shown in Figure 4. The two studies are reported as follows.
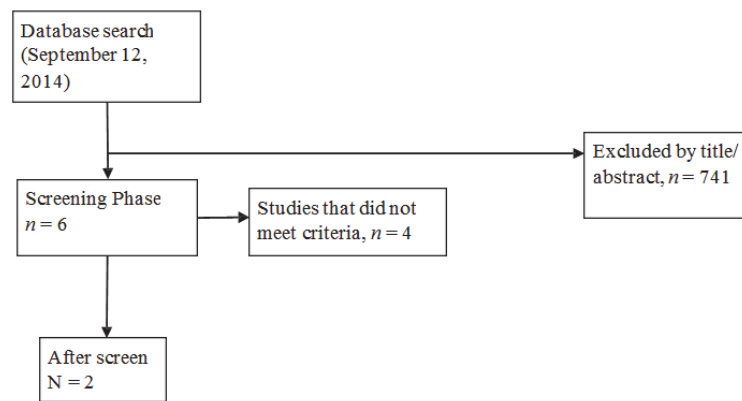
Figure 4. Screening of studies related to per pupil expenditures and achievement

Ohio Study - De Luca and Hinshaw (2013) conducted a study to investigate the role of school district expenditures in predicting student achievement in Ohio. They used data from the 2009-2010 school year and used data from 607 of Ohio's 613 school districts. The variables included in this study are: school district operating expenditures on administration, building operations, instruction, pupil support, and staff support for each of the three academic levels (highest, continuous improvement, and lowest). In the state of Ohio, there are six levels of performance on their standardized tests. They are: advanced, accelerated, proficient, basic, limited, and untested. Students are tested yearly in grades three through eight on reading and math. Fifth and eighth graders are tested in science and tenth graders take the tenth grade graduation assessment. Once the scores are available for the students the school districts are given a score based on their students' performance on the exams. The designations that are given to the school districts are excellent with distinction, excellent, effective, continuous improvement, academic watch, and academic emergency. The independent variables in this study were the percent of total district expenditure for administration, building operations, instruction, pupil support, and staff support. The dependent variable was the performance index score.

The results of this study disclosed that most of the schools spent well below the recommended 65% of their operating budget on instruction. The results of this study indicated that the schools that were the most successful on the performance index were spending their money on pupil support. The results of one of the regression analyses confirmed that only one variable was positive and statistically significant—that of pupil support ($p=$ .027); other variables (staff support, building operation, and administration) were tested. The most successful group from the study spent only 56% of their operating budget on instruction. The researchers also found that the income level of district residents may impact student achievement more that the amount of money spent on instruction.

Texas Study - Jones and Slate (2010) conducted a study that investigated to what extent do Texas public schools compliance with the 65% instructional expenditures ration was related to performance on the TAKS test. The researchers utilized data from TEA for the 2007-2008 school year for this study. There were 943 school districts in the state of Texas that were used for this study. These districts had passing rates on the five TAKS academic measures. 363 school districts provided data for African American students, while 653 school districts provided data for Hispanic students. Texas reports data in a manner that if there are a small number of students the scores are not reported. The researchers used instructional expenditure ratio as the dependent variable for this study. TEA defines instructional expenditure ratio as the money that is spent directly on instructional activities.

The researchers grouped all of the schools into three groups: schools that spent below 60% on instructional expenditures, schools that spent between 60t and 64.99% on instructional expenditures, and lastly schools that spent over 65% on instructional expenditures. They also performed a multivariate analysis of variance (MANOVA) as well as Scheffe post hoc procedures for the subgroups of all students, i.e., African- American students, Hispanic students, and White students. The researchers did the same analysis for all of the students as

well as each subgroup. The results of the MANOVA for all students were Λ =.95, p < .001, $\eta\rho^2$ = .024. The effect sizes for all students are as follows: Math =.04, English =.024, Science =.037, Social Studies = .028, and Writing = 02. The next group that the researchers explored was the African American students. The MANOVA for this group revealed a statistically significant result of Λ = .88, p <.001, $\eta\rho^2$ = .06. The effect sizes for the African American student are as follows: Math=.065, English= .05, Science= .08, Social Studies =.043, and Writing= .038. In regard to Hispanic students in this study, the MANOVA showed that there was a statistically significant result, Λ = .96, p = .001, $\eta\rho^2$ = .02. In regard to the effect sizes for these students, the effect sizes are as follows: Math= .026, English= .014, Science= .017, Social Studies= .01, and Writing= .026. The last group that Jones and Slate (2010) investigated was the White students. The MANOVA presented a result of Λ = .92, p =.001, $\eta\rho^2$ = .04 for the White students. Effect sizes in the form of $\eta\rho^2$ for all students are as follows: Math = .04, English= .03, Science = .06, Social Studies = .05, and Writing=.023.

The overall results from this study expose that there was a strong relationship between 60% instructional expenditures and student test performance. This study demonstrates that there is a clear relationship between instructional expenditure and subgroup student performance (Jones and Slate, 2010). When the researchers conducted the Scheffe post hoc procedures they found that school districts spending below the 60% threshold their students performed significantly lower than did the other two districts that spent more money that this group. The researchers stated that these data are from one year, one state, and should not make broad generalizations because of the results. This study is the first that we found that met the requirements for the best-evidence synthesis.

Per Pupil Expenditure Related to Student Achievement Studies Summary - There were three studies found that included expenditures related to student performance that met the requirements of our search criteria. Two of the researchers focused on the district level spending in schools while one investigated specifically per-pupil expenditure. The two articles included district-level spending and demonstrated that the schools that are more successful spend less than the districts that are not as successful. The article that was about per-pupil expenditure showed that spending more per student improves performance on standardized tests.

3.4 Mobility Rate Related to Student Achievement

There were 26,036 studies found based on the search, there were 53 studies that related to mobility rate and student achievement. Only one studies met the requirements to be included in our study. The requirements for inclusion were (a) all studies are directly related to student performance in American public schools grades 9-12, (b) there were some studies that were mixed-method, but only the quantitative data were included in the modified best-evidence synthesis and (c) each study was from the years 1990-2014. The screening phase is shown in Figure 5. The two studies are reported as follows.
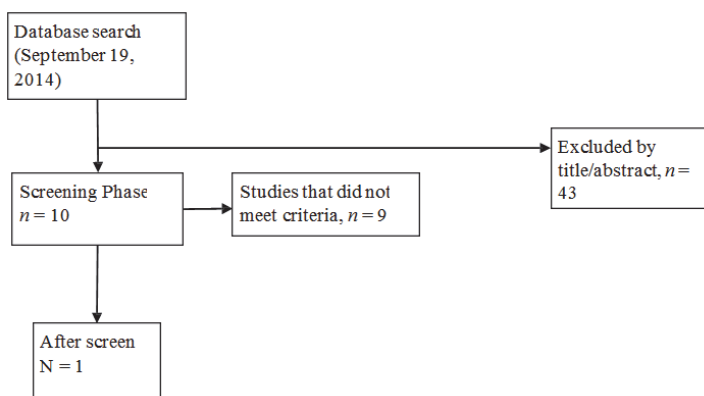

Figure 5. Screening of studies related to mobility rate and achievement

Louisiana Study - Engec (2006) conducted a study in Louisiana that used data from the 1997-1998 and 1998-1999 school years in public schools grades K-12. The researcher used the data from the 1997-1998 school

year to determine mobility rates. The researcher did not use the kindergarten students because there was not test data available. The researcher used the Iowa Test of Basic Skills (ITBS) test data for the 1998-1999 school year to determine student test performance. Another relationship that was investigated was the relationship between mobility rate and suspension rates.

In this study, the grades examined were 3, 5, 6, 7, and 9. The researcher used data from the 1997-1998 school year to investigate suspension rates. The researcher used one-way ANOVAs and ANCOVA to establish the relationships that existed between the variables. The researcher indicated that there is a negative association between ITBS test scores and the number of moves a student makes. In the third grade, students that stay in the same school all year have an average score on the ITBS of 76.05 (*ES* = .45) while the students that move once have a score of 54.46 (*ES* = .11). Students that move two or more times in a school year score on an average of 47.67. In the fifth grade, the results explain that zero moves a student scores an average of 72.07 (*ES* = .43). Students that move one time score an average of 52.68 (*ES* = .12) and the students that move two or more times score 45.14. In the sixth grade, students that stay in the same school all year score an average of 75.76 (*ES* = .50). The students that move once score 51.72 (*ES* =.12) while the students who move two or more times score 44.22. The last group discussed is the seventh grade group. Students in the seventh grade that stay in the same school year round score an average of 74.27 (*ES* = .39). The students that move once during the school year score an average of 50.71 (*ES* = .02). The students that move two or more times during the year score an average of 49.51. There is up to a 20 point difference with some of the grade levels when you look at zero moves to two or more moves during a school year. There were limitations in this study, because the researcher stated that he could not know what caused the frequent moves of the students and that the reasons of the moves could play a role in the negative relationship as well.

Mobility Rate Related to Student Achievement Studies Summary - There was one article that related to mobility rate and student achievement that met our requirements for inclusion in this modified best-evidence synthesis. It was a quantitative study and used data for large groups of students as well. The researcher found that moving more often is linked with lower academic achievement by students.

3.5 Non-White Students Related to Student Achievement

It was found that 26,036 studies found based on our initial search, there were 105 studies that related to the percentage of non-White students and student achievement. Only two published studies found that met the requirements to be included in our study. The requirements for inclusion were (a) all studies are directly related to student performance in American public schools grades 9-12 (b). There were some studies that were mixed-method, but only the quantitative data were included in the modified best- evidence synthesis and (c) each study was from the years 1990-2014. The screening phase is shown in Figure 6. Both of those studies are shared as follows.
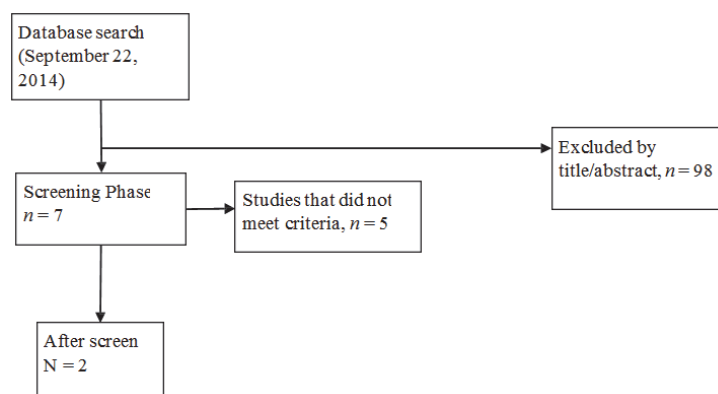

Figure 6. Screening of studies related to Non-White students and achievement

Massachusetts Study - Sanchez, Ehrlich, Midouhas, O'Dwyer, and Regional Educational Laboratory Northeast and Islands (ED) (2009) performed a study in Massachusetts to examine the performance of tenth

grade Hispanic students on the Massachusetts Comprehensive Assessment System (MCAS) tests in English Language Arts and Math. The researchers took data from the 2002-2003 school year to the 2005-2006 school year. They conducted the research because of the achievement gap that existed between the Hispanic students and the other subgroups of students in the state of Massachusetts. The researchers performed multilevel regressions for this research study. There were many variables that were used for the study. Some of these variables include gender, from a low-income household, special education, limited English proficient, former limited English proficient, first languages other than English, and immigrant from other countries. Overall, the results of this study show that students that are Hispanic scored lower on the MCAS ELA test than students who were non-Hispanic. The t-test scores demonstrate this with the Hispanic students having an overall $t(\text{df}) = 20.09$, $p = 13,309$ and the non- Hispanic student having a $t(\text{df}) = 24.43$, $p = (121,863)$. The results of this study show that students that are Hispanic scored lower on the MCAS Math test than students who were non-Hispanic. The t-test scores demonstrate this with the Hispanic students having an overall $t(\text{df}) = 31.72$, $p = (13,309)$ and the non-Hispanic student having a $t(\text{df}) = = 90.75$, $p = (121,863)$. The results of this study explains that the students that were part of the Hispanic population scored significantly lower that non-Hispanic students each year of the study.

With regard to the MCAS English Language Arts test, students that were from a low-income household scored lower than students that were not. Another result of the study is that the scores of the Hispanic students increased over the course of the study by a significant amount. Female Hispanic students outscored their male counterparts by a significant amount on the English Language Arts portion of the assessment while the Male Hispanic students scored significantly higher on the Math portion than the female students. Hispanic students in the study who were from low-income homes, special education, or limited English proficient or formerly limited English proficient in the previous two years performed significantly lower on the Math and English language arts portions of the assessment. The students that went to school on campuses with higher attendance rates scored significantly higher on the English language arts and Math parts of the test than student that went to school on campuses with lower attendance rates.

The study had several limitations including the fact that the study described statistical associations, not causal relationships between the variables of the study. Another limitation of the study is that there was a portion of the data that was excluded from the study appeared to include students performed lower on the test. This could have impacted the study by giving bias to the study. There were many variables that could possibly explain the differences in the performance on the exam that were not analyzed in this study. Another limitation that was discussed is the fact that the federal testing policy was changed in February of 2004 and this was not accounted for in the study. Effect sizes were not discussed in this article.

St. Louis Study - Trent (1997) conducted a study that used data from the St. Louis Public School district to determine if Black students are performing lower than white students. The researcher used data from the 1994-1995 data file for the first part of the study. The second portion of the study was based on a national survey that focused on the effects of race on education, employment, and attitudes. The researcher used the Stanford Achievement Test (SAT) as the measure of achievement for the purpose of this study. Math and reading are the two portions of the test that were used for this study. The researcher used the data to determine if the Black and White students' scores were different because of differences in students' backgrounds. Variables used in this study were age, sex, socioeconomic status (SES), prior test scores, and school characteristics. Some of the school characteristics used include school size and poverty concentration.

The researcher used regression analysis to determine the differences in achievement scores between Black and White students. He did this at the state and national level. The overall findings of this study illustrate that poverty is a significant predictor on the SAT reading exam with β being -.832 for the entire sample of students in regard to Reading performance. In regard to Math performance, β is -1.007, and this shows that poverty concentration is a significant predictor on student performance. The findings from this study include the fact that students that live in neighborhoods with higher poverty concentration score lower than students that live in

neighborhoods with a lower poverty concentration. The findings also confirmed that Black students are more likely to attend schools with a higher concentration of poverty. Another thing that the researcher found is that students that go to schools with a higher population of poverty the worse they perform.

Percentage of Non-White Students Related to Student Achievement Studies Summary - There were only two articles that were found that related to the percentage of non-White students and student achievement. Both of the articles establish that students that were non-White performed lower than expected in school. Also, they revealed that schools that had more students that were non-White performed had lower performance ratings.

3.6 Multiple Variables Related to Student Achievement

There were 26,036 total studies found based on the search. However, there were 26 studies that related to multiple variables and student achievement and only four published studies found that met the requirements to be included in our study. The requirements for inclusion were: (a) all studies are directly related to student performance in American public schools grades 9-12 (b) every study was quantitative in nature that we used for our research. There were some studies that were mixed-method, but only the quantitative data were included in the modified best-evidence synthesis and (c) each study was from the years 1990-2014. The screening phase is shown in Figure 7. All four of those studies are shared as follows.
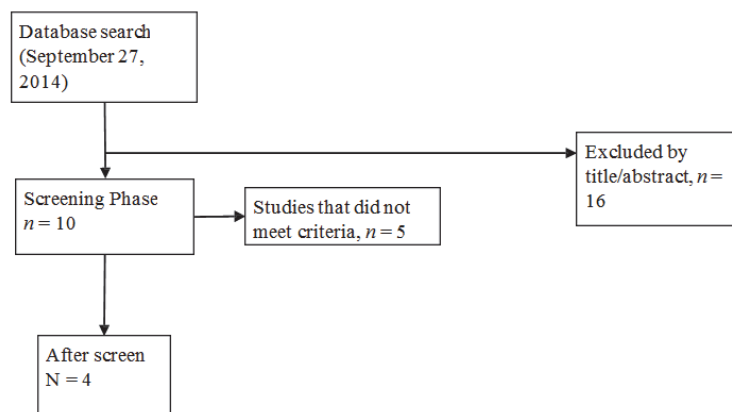

Figure 7. Screening of studies related to multiple variables and achievement

Louisiana Study - Bankston and Caldas (1998) investigated the influence of schoolmate family structure, racial concentration, and socioeconomic status on the academic achievement of individual African American and White students in Louisiana. The researchers pulled their data from the test results of 18,000 tenth graders who took the Louisiana Graduation Exit Examination. They also only used students who were either White or African American. They excluded students who were in special education. There were a total of 42,041 students that could be used in this study and the researchers used random sampling to limit the sample to 18,310 students. The researchers wanted to determine academic achievement for this study so they used raw scores on the math, language arts, and written composition to do this. The researchers had the following variables as part of the study: Female-headed family structure, race, parents' educational level, family poverty status, and characteristics of schoolmates. Female-headed family structure was determined based on a family that was a single-parent, female-headed household and by coding them as 1 while coding all other families as 0. Race was determined by coding African American students as "1" and White students were coded a "0." Students were coded as 0 if they were not participants in the free or reduced lunch program and as 1 if they were participants.

The researchers used SAS Proc Mixed to multilevel unconditional means and random coefficient models. They had multilevel models that they referred to as "random coefficient regression models," "random-effect models," or "multilevel linear models" (Bankston & Caldas, 718, 1998). The results of the study show that 71% of the African American students performed below the median on the study's measure while only 34% of White students performed below the same level. Also, findings reveal that there are strong negative correlations between academic achievement and the African American race with a score of ($r = -.363$), percentage of African

American race in school with a score of (r =-.303), female-headed family with a score of (r =-.373), and percentage of female-headed families in schools with a score of (r =-.352). The largest correlation in the study is between the percentage of African Americans in school and percentage of students from female-headed families in school with a score of (r =.797). The study results revealed that students of color had the strongest negative correlation with test performance. The level of the parents' education was positively associated with test performance. The order of correlation was students of color, followed by parental education, poverty status, and then single-parent family structure.

Missouri Study - Vorthmann (2011) performed a study that examined the relationship between school size and student achievement in Missouri. The researcher investigated elementary, middle, and high schools to determine the optimal school size to maximize student achievement on the Missouri Assessment Program (MAP) grade level assessments as well as EOC assessments for the 2009-2010 school year. The researcher used a causal-comparative quantitative research design. School size was the independent variable was established by using the data from the Missouri Department of Elementary and Secondary Education (DESE). The researcher divided the school sizes into groups and used the following labels for the high schools: Very Small (35-153 students), Small (154-265 students), Medium (266-558 students), Large (559-1040 students), and Very Large (1041-2421 students). The researcher used the following independent variables location, ethnicity, poverty, and special education classifications. Included in the study were all 2,334 public schools in Missouri excluding charter, alternative, special education, career, vocational, and technical schools. From this, 492 of the schools were high schools. The EOCs that were included in the study are Algebra I, English II, Biology, and Government. There are four score ranges and achievement levels for these EOC assessments. They are as follows Below Basic (100-176), Basic (177-199), Proficient (200-224), and Advanced (225-250).

The researcher conducted the research by "using a one factor analysis of variance (ANOVA) to determine if statistically significant differences in student achievement existed as measured by the MAP Grade-Level Assessments in communication arts and mathematics and MAP EOC Assessments in English II, Algebra I, biology, and government among schools of different sizes. Statistical significance for the ANOVA was set at α = .05. A follow-up post hoc analysis, the Tukey Honestly Significant Difference (HSD), was used to determine which interaction effect means were statistically significantly different with α = .05. A single hypothesis was tested to address each research question for RQ 1 through RQ 8" (Vorthmann, 2011). The researcher also conducted an ANOVA for each of the other variables as well.

In regard to the findings, there were mixed findings, but there was higher student achievement in English II, biology, and government in very large high schools. Algebra I was the exception where the higher achievement was seen in small high schools. The interaction effect of school location, ethnicity, poverty, and special education classification on elementary, middle, and high school size was also examined by the researcher. The researcher found that large city high schools had the lowest levels of student achievement in English II and biology. High minority schools produced a higher level of student achievement in English II and biology. In high-ethnicity high schools, medium and very large high minority high schools produced higher levels of student achievement in English II and biology when compared to large high minority high schools. The study revealed that lower levels of student achievement in high poverty schools rang true in all four areas tested on the EOCs. Among the high poverty schools, the very small, small, and medium high poverty high schools had higher scores on Algebra I and Biology compared to the big school schools. In regard to English II, the small and medium poverty schools produced higher levels of student achievement than the large and very large schools. Biology compared to large and very large high poverty high schools, while small and medium high poverty schools produced higher levels of student achievement in English II compared to large and very large high poverty schools. The very small and medium poverty high schools produced higher levels of student achievement on the government EOC. The last variable discussed was special education and its relationship to student achievement. Schools that were high special education serving schools produced the lowest levels of student achievement in all areas tested. All of these interactions show that different variables impact student performance and the combination of some account for student performance changes. This research study covered elementary schools, middle schools, as well as

high schools, and the only portion that we included was the portion about high schools.

South Carolina Study - Durbin (2001) carried out a study to determine the relationship between the size of high schools, per pupil expenditure, and student achievement while controlling for the effects of socioeconomic status (SES). The study used all eleventh graders in 192 South Carolina public high schools that took the Metropolitan Achievement Test-Seventh Edition (MAT-7) in the spring of 1998. The subjects used for this study were reading, written language, and mathematics. The researcher used the mean total scores for all three subjects as the measure of student achievement.

School size was defined by the total number of students tested in the spring 1998 administration of the MAT-7. Socioeconomic status was measured by the percent of eleventh grade students identified in the "Total Number Tested" that received free lunch in the school-lunch program during the 1997-1998 school year. In another portion of the study, school size was defined as the daily average membership in the school. The last variable that was used in the study is per pupil expenditure. It was defined as the 1997- 98 fiscal year operational cost per pupil for each of South Carolina's public high schools with an eleventh grade (Durbin, 2001).

The schools were grouped into quartiles and the 192 schools were broken down into the following groups: very small schools 17-99 eleventh graders tested, small schools 101-174 eleventh graders tested, medium schools 175-251 eleventh graders tested, and large schools 254-629 eleventh graders tested. The researcher broke the schools into groups based on SES and established the following groups: low SES group 53%-85% free lunch eleventh graders tested, medium SES group 28%-52% free lunch eleventh graders tested, and the high SES group 3%-27% free lunch eleventh graders tested.

The researcher used multiple regression to determine the relationships between all of the variables included in the study. The researcher also set α =.05. The data sets were collected from the South Carolina State Department of Education. The relationship between school size and achievement was examined while controlling for socioeconomic status using a multiple regression process, a review of descriptive statistics, and a post hoc multiple comparisons technique. These data were used to identify school sizes that yielded academic success for students while maintaining cost effectiveness.

The results of the study start with the relationship between school size and mean total reading scores while controlling for SES. Both school size and SES were predictors on the mean total reading standard scores. The partial correlation between school size and mean total reading score was p = .201 with a significance of .005. The next portion of the study involved mean total mathematics scores while controlling for SES. The relationship between school size and mean total mathematics standard scores when controlling for SES, was statistically significant, with p =.012 and a partial correlation of .182. The third portion of the study examined the relationship between school size and mean total written language standard scores while controlling for SES. Both school size and SES were found to have a statistically significant relationship to the mean total written language scores. The relationship between school size and the mean total written language standard scores, when controlling for SES, was statistically significant with p =.010 and a partial correlation of .185. Overall, as school size increased, scores in reading, mathematics, and written language increased significantly, even after controlling across size groupings for SES. There were other portions of the results that did not directly pertain to the modified best- evidence synthesis so they were not included. This study was part of a dissertation.

West Virginia Study - Howley (1996) conducted a study in West Virginia that investigated the relationship between school size and student performance. This study was based on the researcher's dissertation. Both school size and district size were used as units of analysis in the study. West Virginia schools that include third grade, sixth grade, ninth grade, or eleventh grade were included as well as all West Virginia school districts. The variable of size was defined as the number of students in each grade level grouped by school or school district. SES was defined by the rate of students receiving free or reduced lunches. Student achievement was measured by the Comprehensive Test of Basic Skills (CTBS). This test is administered to all students in West Virginia that are not in special education classes. The data for this study is from the 1990 school year. The researcher used

regression analysis to determine the relationship between school size and achievement.

The researcher found that the regression pointed out that there was not a statistically significant correlation between seven of eight analyses. The only correlation that was statistically significant existed in the third grade analysis (r = .11). This article was basic in the explanation of the results and was missing some tables that demonstrated the results. The overall findings of this study showed that the relationship between school size and student performance changed based on different variables. SES impacted student performance in some cases, and school size impacted it other cases. In some situations, both of these variables impacted student performance.

Multiple Variables Related to Student Achievement Studies Summary - The four articles that were found that discussed multiple variables that included school size, district size, school location, ethnicity, poverty, schoolmate family structure, racial concentration, and socioeconomic status, special education classifications, per-pupil expenditure, and student performance. The results from our research synthesis exposed that such variables were measured against different levels of students across all four studies. Bankston and Caldas (1998) used tenth graders as the subject of their study, Vorthmann (2011) used elementary, middle, and high school students as the subjects of his study, Durbin (2001) used eleventh graders as the subject of his study, and lastly, Howley (1996) used third, sixth, ninth, and eleventh graders as the subjects for his research. The studies varied greatly in regard to organization as well as outcome.

4. Conclusion

All of the researchers' studies included in the modified best-evidence synthesis promulgated that there are numerous variables that impact student performance. The variables that were researched were school size, student socio-economic status, per-pupil expenditures, student mobility rate, and percentage of non-White students as those specific variables relate to or impact the performance of students on standardized tests. From the research that exists, a proposed model was developed that demonstrates the impact that each variable may have on student secondary science achievement. This model is shown in Figure 8. The model suggests, that based on the published literature, there is a link between non-White students and mobility rate, as well as SES and per pupil expenditures. It is possible that there is a predicted path between SES and non-White students and another predicted path between SES and school size.
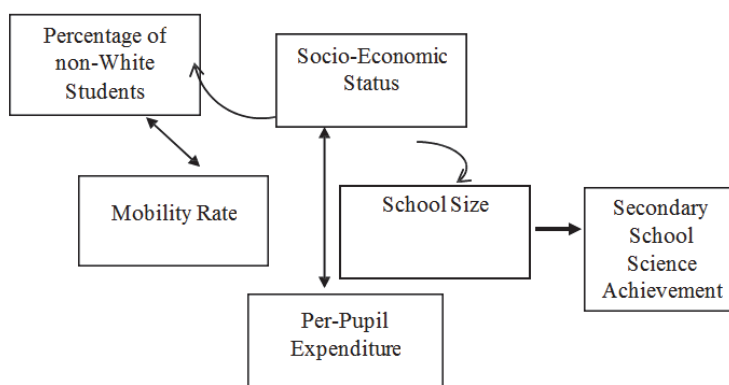

Figure 8. Proposed model of the relationship of secondary school science achievement and school size considering the variables of percentage of non-White students, mobility rates, per pupil expenditure, and socio-economic status based on published research.

5. References

107th congress of the United States of America. No child left behind. (2001). Retrieved from
        http://www2.ed.gov/policy/elsec/leg/esea02/index.html
American Association for the Advancement of Science. (1989). Science for all Americans. Retrieved from

Barton, J., Irby, B. J., Tong, F., Torres, M., & Sandlin, J.

http://www.project2061.org/publications/sfaa/online/sfaatoc.htm

American Association for the Advancement of Science. (1993). Benchmarks for Science Literacy. Retrieved from http://www.project2061.org/publications/bsl/

American Association for the Advancement of Science. (2013). Project 2061. Retrieved from http://www.project2061.org/about/default.htm

Bankston, C. L., & Caldas, S. J. (1998). Family structure, schoolmates, and racial inequalities in school achievement. *Journal of Marriage and the Family*, *60*(3), 715–723. https://doi.org/10.2307/353540

Caldas, S. J., & Bankston, C. L. (1997). Effect of school population socioeconomic status on individual academic achievement. *Journal of Educational Research, 90*(5), 269-277. https://doi.org/10.1080/00220671.1997.10544583

De Luca, B. M., & Hinshaw, S. A. (2013). Predicting student achievement in Ohio: The role of expenditure distribution. *Educational Considerations*, *40*(3), 9–16.

Durbin, M. K. (2001). *The relationship of high school size, student achievement, and per-pupil expenditures in South Carolina* (Ph.D. dissertation). University of South Carolina.

Engec, N. (2006). Relationship between mobility and student performance and behavior. *Journal of Educational Research*, *99*(3), 167–178. https://doi.org/10.3200/JOER.99.3.167-178

Fowler, W. J., & Walberg, H. J. (1991). School size, characteristics, and outcomes. *Educational Evaluation and Policy Analysis*, *13*(2), 189–202. https://doi.org/10.3102/01623737013002189

Howley, C. (1996). Compounding disadvantage: The effects of school and district size on student achievement in West Virginia. *Journal of Research in Rural Education*, *12*(1), 25–32.

Jackson, J. K., & Ash, G. (2012). Science achievement for all: Improving science performance and closing achievement gaps. *Journal of Science Teacher Education*, *23*(7), 723–744. https://doi.org/10.1007/s10972-011-9238-z

Jones, T. B., & Slate, J. R. (2010). The 65% instructional expenditure ratio and student achievement: Does money matter? *Current Issues in Education*, *13*(4). Retrieved from http://cie.asu.edu/ojs/index.php/cieatasu/article/view/487/100

McMillen, B., Zhang, G., Cobb, C., Williamson, G., Kaase, K., Williams, J., & Feifs, H. (2000). School size and its relationship to achievement and behavior. Retrieved from http://www.ncpublicschools.org/docs/data/reports/size.pdf

National Research Council. (1996). National science education standards. Retrieved from https://www.nap.edu/read/4962/chapter/1

National Research Council. (2012). A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. Retrieved from https://www.nap.edu/read/13165/chapter/1

National Research Council. (2013). The next generation science standards. Retrieved from http://www.nextgenscience.org/next-generation-science-standards

Research Center. (2011). Research center: Achievement gap. Retrieved from http://edweek.org/ew/issues/achievement-gap/

Robinson, V. M. J., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly*, *44*(5), 635–674. https://doi.org/10.1177/0013161X08321509

Sanchez, M. T., Ehrlich, S., Midouhas, E., O'Dwyer, L., & Regional Educational Laboratory Northeast & Islands (ED). (2009). *Analyzing performance by grade 10 Hispanic high school students on the Massachusetts state assessment. Issues and answers*. Regional Educational Laboratory Northeast & Islands. ERIC document ED505420

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher*, *15*(9), 5–11. https://doi.org/10.3102/0013189X015009005

Slavin, R. E. (1987). Best-evidence synthesis: Why less is more. *Educational Researcher*, *16*(4), 15–16. Retrieved from http://doi.org/10.3102/0013189X016004015

Stewart, L. (2009). Achievement differences between large and small schools in Texas. *Rural Educator*, *30*(2), 20–28.

Texas Education Agency. (2012). TAKS statewide summary reports 2010–2011. Retrieved from
        http://www.tea.state.tx.us/student.assessment/taks/rpt/sum/yr11/

Texas Education Agency. (2013a). 2011-2012 State AEIS report. Retrieved from
        http://ritter.tea.state.tx.us/perfreport/aeis/2012/state.html

Texas Education Agency. (2013b). 2011-2012 State AEIS report. Retrieved from
        http://ritter.tea.state.tx.us/perfreport/aeis/2012/state.html

Texas Education Agency. (2013c). House bill 5. Retrieved from
        http://www.tea.state.tx.us/index2.aspx?id=25769806149

Texas Education Agency. (2013d). STAAR statewide summary reports 2012-2013. Retrieved from
        http://www.tea.state.tx.us/student.assessment/staar/rpt/sum/yr13/

Trent, W. T. (1997). Why the gap between black and white performance in school? A report on the effects of race
        on student achievement in the St. Louis public schools. *Journal of Negro Education*, *66*(3), 320–329.
        https://doi.org/10.2307/2967169

Vorthmann, C. (2011). *The relationship between school size and student achievement in Missouri* (Ed.D.
        dissertation). Northwest Missouri State University.