Asuncion Los Banos

DS-670 Capstone: Big Data & Business Analytics

Assignment 8

**Contribution**

Do students who score high on standardized tests naturally good test takers or are there other factors that take place? Popham (1999) states that “standardized test is any examination that's administered and scored in a predetermined, standard manner”. Students in United States will likely take two major kinds: aptitude and achievements tests. For college bound high school students they will partake in an aptitude test that will predict how well they are likely to perform in educational settings. Those aptitude tests are the SAT and ACT, which forecast how well a high school student will perform in college (Popham, 1999). Students while in elementary, middle, and high school will be exposed to the other type of standardized test, the achievement test. These score results are what citizens and school board members rely on to base on how well the school is performing (Popham, 1999). Understanding what factors that can influence a student’s test score in an achievement test will help school districts become more efficient.

The State of New Jersey’s achievement test score from 1983 to 2014 was the High School Proficiency Assessment (HSPA). It tested students who were in their junior year in Reading and Mathematics. Then in 2015 to present the standardized test score was changed to the Partnership for Assessment of Readiness for College and Careers (PARCC) which believes that assessments should work as tools for enhancing teaching and learning. The PARCC assessments are aligned with the Common Core State Standards.

Achievement test results are used to evaluate a school's effectiveness, but since a school is made up of different individual’s does their socioeconomic background come into play? My contribution is to find what factors can influence a student’s test performance. The factors that I will be utilizing are average household income, being of white ethnicity, married households, and education by county in New Jersey. I am using white as the only ethnicity because there have been many conversations regarding education gap of African American and Hispanic. Therefore, I would like to know if being Caucasian helps raise your test score. I hypothesize that counties in New Jersey with high incomes will have higher average scores.

I will be using the 2014 and 2007 HSPA reading and math results. The PARCC results are not viable as there are only two years of data and the first year the test was implemented numerous students throughout that state opted-out, therefore it does not paint a correct picture of each district.

The data for 2014 and 2007 HSPA results were retrieved from the State of New Jersey’s Department of Education website and these results will be my dependent variables. My independent variables of household income, ethnicity, married households, and education were gathered from the U.S. Census Bureau’s American FactFinder. To model my data, I will be using linear regression.

**State of the Art**

[1] In Dahl and Lochner’s paper titled “The impact of family income on child achievement: Evidence from the earned income tax credit” a strategy of instrumental variables was used. They estimated the causal effect of income on children’s math and reading achievement. Their identification derived from large, nonlinear changes in the Earned Income Tax Credit. From their paper, they found that “the largest of these changes increased family income by as much as 20 percent, or approximately $2,100, between 1993 and 1997. Our baseline estimates imply that a $1,000 increase in income raises combined math and reading test scores by 6 percent of a standard deviation in the short run. Test gains are larger for children from disadvantaged families and robust to a variety of alternative specifications.” Their estimation strategy was based on the

fact that low- to middle-income families benefitted substantially from expansions of the EITC in the late-1980s and mid-1990s while higher-income families did.

[2] Anger and Heineck’s paper examined the intergenerational transmission of cognitive abilities. Their finding was not compatible to a pure genetic model, but rather point to the importance of parental investments for children’s cognitive outcomes. They found that “individuals’ cognitive skills are positively related to their parents’ abilities, despite controlling for educational attainment and family background”.

[3] In the paper titled “The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood” by Chetty, Friedman, and Rockoff measured if teacher’s impacts on students test scores a good measure of their quality. They analyzed school district data from grades 3 to 8 and linked the student’s test scores to tax records on parent characteristics and adult outcomes. Their paper found that “students assigned to high-value-added teachers are more likely to attend college, attend higher- ranked colleges, earn higher salaries, live in higher SES neighborhoods, and save more for retirement.”

[4] The paper by Matthew M. Chingos titled “Strength in Numbers: State Spending on K-12 Assessment Systems” provides the most current, comprehensive evidence on state-level costs of assessment systems. I felt this paper was important because achievement tests are used to evaluate how well a school is performing my the public and law makers.

[5] The paper titled “What high-achieving low-income students know about college” is interesting because there are studies on the education gap and that a student from a low-income family does not perform well in school. This paper provides “provides individualized information about colleges' net prices, resources, curricula, students, and outcomes. Our prior study shows that the intervention raises students' applications to, admissions at, enrollment, and progress at selective colleges.” A survey data was used to show changes in students’ knowledge and decision-making.

[6] Mayer focuses on the effect of parent’s income and isolated the effect of parental income on children’s outcomes, in particular the effect of low parental income on poverty. The author concluded that “evidence suggested that a 10 percent increase in parental income was associated with .024 to .104 additional years of schooling.”

[7] In the paper “Why Standardized Tests Don't Measure Educational Quality” there is an argument of why standardized test should not be used as a measure for a student’s ability. The paper states, “The task for those developing standardized achievement tests is to create an assessment instrument that, with a handful of items, yields valid norm-referenced interpretations of a student's status regarding a substantial chunk of content. Items that do the best job of discriminating among students are those answered correctly.”

[8] The paper “Income Level, Gender, Ethnicity, and Household Composition as Predictors of Children's School‐based Competence” they state that “In the United States, being black, male, or growing up in a low-income and/or single-parent household have all been identified as risk factors for maladjustment during childhood.” The authors compared predictions of 3 different forms of children's competence from each of these 4 variables. They used a sample of 868 black and white elementary school children from 2-parent and mother-headed 1-parent homes, we studied 3 aspects of school-based competence: conduct, peer relations, and academic achievement. The result was “Overall, income level and gender were thus the strongest predictors of children's competence.”

[9] The paper “The influence of parent education and family income on child achievement: the indirect role of parental expectations and the home environment” examined the process of how socioeconomic status, specifically parents' education and income, indirectly relates to children's academic achievement through parents' beliefs and behaviors. The data came from a a national, cross-sectional study of children were used for this study. The subjects were 868 8-12-year-olds, divided approximately equally across gender (436 females, 433 males). This sample was 49% non-Hispanic European American and 47% African American. A structural equation modeling techniques, the author found that the socioeconomic factors were related indirectly to children's academic achievement through parents' beliefs and behaviors but that the process of these relations was different by racial group.

[10] The authors of “Ethnic differences in children's intelligence test scores: Role of economic deprivation, home environment, and maternal” examined differences in intelligence test scores of black and white 5-year-olds. The data from Infant Health and Development Program included 483 low birthweight premature children who were assessed with the Wechsler Preschool and Primary Scale of Intelligence. The children had been followed from birth, with data on neighborhood and family poverty, family structure, family resources, maternal characteristics, and home environment collected over the first 5 years of life. The paper found that “Black children's IQ scores were 1 SD lower than those of white children. Adjustments for ethnic differences in poverty reduced the ethnic differential by 52%”.

[11] Attanasio and Lechene used the Progresa data from Mexico to investigate intrahousehold decision making using a variety of outcomes. They tested “for global pooling of resources within households, which would correspond to the unitary model of household decision making. We also exploit a set of questions about power and the decision-making process in the household to investigate aspects of strategic interactions between household members.” The confirmed previous rejections of income pooling.

[12] Duckworth, Quinn, and Tsukayama proposed that standardized achievement test scores assess competencies determined more by intelligence than by self-control, whereas report card grades assess competencies determined more by self-control than by intelligence. They completed “Two longitudinal, prospective studies of middle school students support predictions from this model. In both samples, IQ predicted changes in standardized achievement test scores over time better than did self-control, whereas self-control predicted changes in report card grades over time better than did IQ.”

[13] The authors of “Do social and behavioral characteristics targeted by preventive interventions predict standardized test scores and grades” assessed whether characteristics of individuals that are predictors of youth problem behavior such as substance use, delinquency, and violence also predict academic achievement. Longitudinal data from 576 students participating in the Raising Healthy Children (RHC) project were analyzed. The paper found that lower test scores and lower grades were predicted by elevated levels of attention problems, negative behavior of peers, and disruptive and aggressive behavior. Lower test scores also were predicted by early use of alcohol and cigarettes.

[14] The authors of “Why the apple doesn't fall far: Understanding intergenerational transmission of human capital” questioned if parental education actually changes the outcomes of children, suggesting an important spillover of education policies, or is it merely that more able individuals who have higher education also have more able children? They used a unique dataset from Norway and found little evidence of a causal relationship between parents' education and children's education, despite significant OLS relationships.

[15] In the paper “The contribution of neighborhood and family income to developmental test scores over the first three years of life” the effects of neighborhood and family income and family risk factors on developmental test scores at ages 1 through 3 are examined using a subsample from the Infant Health and Development Program. Moderating effects of family risk on family and neighborhood income effects revealed an interaction between family poverty and risks for scores at age 3.

[16] Bigelow wrote how standardized tests threaten multiculturalism. He hypothesized that “If multiculturalism's key goal is accounting for historical influences on current social realities, then Oregon's standards and tests earn a failing grade.”

[17] Downey tried to find the explanation of the inverse relationship between the number of siblings and children's educational performance. The author had a sample of 24,599 eighth graders from the 1988 National Education Longitudinal Study. His model found that children benefit less from certain parental resources when they have many versus few siblings.

[18] The authors of “The roles of self-esteem and the sense of personal control in the academic achievement process” proposed that academic achievement boosts self-esteem and the sense of personal control, but that only the latter influences subsequent academic achievement. They used three waves of panel data and examined the effects of academic achievement in the 8th grade on the sense of personal control and self-esteem in the 10th grade and the subsequent effects of control and esteem in the 10th grade on academic achievement in the 12th grade. They were able to present “subsequent effects of control and esteem in the 10th grade on academic achievement in the 12th grade. They present evidence that the sense of personal control affects subsequent academic achievement, but that self-esteem does not. Earlier academic achievement and parental support increase self-esteem and the sense of personal control”.

[19] In the paper “Collateral damage: How high-stakes testing corrupts America's schools” categorized the way high-stakes testing threatens the purposes and ideals of the American education system. Their analysis is grounded in the application of Campbell’s Law, which posits that the greater the social consequences associated with a quantitative indicator, the more likely it is that the indicator itself will become corrupted—and the more likely it is that the use of the indicator will corrupt the social processes it was intended to monitor.

[20] The authors of “Does money really matter? Estimating impacts of family income on young children's achievement with data from random-assignment experiments” used a random-assignment-induced variation in family income in four welfare and anti-poverty programs to identify income effects. Their results suggest that family income has a policy-relevant positive impact on the school achievement of preschool but not older children.

**Data**

I will be using the 2014 and 2007 HSPA reading and math results. The data for 2014 and 2007 HSPA results were retrieved from the State of New Jersey’s Department of Education website and these results will be my dependent variables. The dependent variables will be the test scores from the 21 counties in New Jersey. In 2014 and 2007 the numerical scores for reading and math of 199 and below indicated partially proficient, 200 to 249 indicated proficient and 250 to 300 indicated advanced proficient. For the purpose of this capstone the average score from the two subject areas will be used.

My independent variables of household income, ethnicity, married households, and education were gathered from the U.S. Census Bureau’s American FactFinder and have been recognized as having an influential impact on standardized test scores. American FactFinder provides access to data about the United States and come from several censuses and surveys. I decided to utilize the Guided Search in the website to help search for the data I required. The constant filter for my search was the geographic type: County, New Jersey as the main state and all counties selected. Each independent variable can be easily located on the website using the ID column.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Year** | **Label** | **Table, File or Document Title** | **Dataset** | **ID** |
| 2014 | average\_income | Income In the Past 12 Months (In 2015 Inflation - Adjusted Dollars) | 2011-2015 American Community Survey 5-Year Estimates | S1901 |
| 2014 | ethnicity \_ white | RACE  Universe: Total population | 2015 American Community Survey 1 - Year Estimates | DP05 |
| 2014 | married \_ household | Households and Families | 2011 - 2015 American Community Survey 5-Year Estimates | S1101 |
| 2014 | education | Educational Attainment | 2011 - 2015 American Community Survey 5 - Year Estimates | S1501 |
| 2007 | average \_ income | Income in the Past 12 Months (In 2007 Inflation - Adjusted Dollars) | 2005 - 2007 American Community Survey 3 - Year Estimates | S1901 |
| 2007 | ethnicity\_ white | RACE  Universe: Total population | 2007 American Community Survey 1 - Year Estimates | DP05 |
| 2014 | married \_ household | Households and Families | 2005 - 2007 American Community Survey 3 - Year Estimates | S1101 |
| 2007 | education | Educational Attainment | 2005 - 2007 American Community Survey 3 - Year Estimates | S1501 |

My defined variables:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Description** | **Label** | **Variable** | **Measure** | **Type** |
| Student’s Average Test Score | testscore\_avg | Average test score of reading and math HSPA | Numerical value | Dependent |
| Average Household Income | average\_income | Average income in a county of New Jersey | Dollars | Independent |
| Ethnicity | ethnicity\_white | Percentage of population identified as “White” in New Jersey | Percentage value | Independent |
| Married Households | married\_household | Percentage of population living in a married household | Percentage value | Independent |
| Education | education | Percentage of population who have achieved a bachelor’s or higher | Percentage value | Independent |

**Method**

To model my data, I will be using linear regression as my statistical method. Linear regression can “predict” the value of the dependent variable based upon the values of one or more independent variables. This statistical data analysis will be used to determine where there is a linear relationship between a dependent variable and one or more independent variables. I will be utilizing two types of linear regression: simple linear regression and multiple linear regression. Regression analysis has three major uses and they are: causal analysis, forecasting an effect, and trend forecasting. For my causal analysis I am trying to answer the following question: “What is the strength of relationship between average test score and average income?” Using regression to forecast an effect I am trying to answer: “How much more income does a household need to earn for a student to score higher in their test score?”

In a simple linear regression a single independent variable is used to predict the value of a dependent variable. In a multiple linear regression two or more independent variables are used to predict the value of a dependent variable. The only difference between a single linear regression and a multiple linear regression is the number of independent variables. However, for both there is only one dependent variable. My dependent variable of average test scores is measured in a continuous measurement scale of 0 to 300. The independent variables are also measured in a continuous measurement scale of 0 to 100 percent and dollars.

My single linear regression (SLR) will be regressing the average test scores of reading and math with average household income in New Jersey. I will have three multiple linear regressions (MLR I, MLR II, and MLR III) that will regress average test score with the independent variables factors (household income, ethnicity, married households, and education). MLR (I) will be regressing average test scores with white ethnicity and married households in the state. MLR (2) will be regressing average test scores with white ethnicity and education. MLR (3) will be regressing average test scores with white ethnicity and married households, and education. I have kept my independent variables to four variables to keep it from overfitting which can make the model inefficient. I am going to keep my model as simple as possible because statistically if the model includes a large number of variables the probability increases that the variables will be statistically significant from random effects.

**Regression Models:**

**The formula for Single Linear Regression**

|  |
| --- |
| testscore \_ avg = β 0 + β 1 (average \_ income) |

Regressing New Jersey counties average HSPA scores with the average household income.

**The formula for Multiple Linear Regression (1)**

|  |
| --- |
| testscore \_ avg =  β 0 + β 1 (average \_ income) + β 2 (ethnicity \_ white) + β 3 (married \_ household) |

Regressing New Jersey counties average HSPA scores on the average household income, controlling for the percentage of the population in the county that is white and married households.

**The formula for Multiple Linear Regression (2)**

|  |
| --- |
| testscore\_avg =  β 0 + β 1 (average \_ income) + β 2 (ethnicity \_ white) + β 3 (education) |

Regressing New Jersey counties average HSPA scores on the average household income, controlling for the percentage of the population in the county that is white and with an educational level of some college or more.

**The formula for Multiple Linear Regression (3)**

|  |
| --- |
| testscore \_ avg =  β 0 + β 1 (average \_ income) + β 2 (ethnicity \_ white) + β 3 (married \_ household) + β 4 (education) |

Regressing New Jersey counties average HSPA scores on the average household income, controlling for the percentage of the population in the county that is white, married households, and adults with an educational level of some college or more.

I will be utilizing the Gauss–Markov theorem which states that in a linear regression model in which the errors have expectation zero and are uncorrelated and have equal variances, the best linear unbiased estimator (BLUE) of the coefficients is given by the ordinary least squares (OLS) estimator, provided it exists.

**Assumption 1 - Linear in Parameters**

The model is in linear parameters, but does not have to be linear in the x’s: Y = β 0 + β 1 X1 + β i Xi

**Assumption 2 - Random Sample of n Observations**

An appropriately sized, random sample is used in the regression model.

**Assumption 3 – Zero Conditional Mean**

The mean of the error terms has an expected value of zero given values for the independent variables.

**Assumption 4 – No Perfect Collinearity**

The error term, *u*, is independently distributed and not correlated with any of the variables. The

variables are not correlated. The assumption of no perfect collinearity states that *there is no exact linear relationship among the independent variables*. This assumption implies two aspects of the data on the independent variables.

**Assumption 5 - Homoskedasticity**

The error terms all have the same variance and are not correlated with each other. In statistical jargon, the error terms are independent and identically distributed (iid). This assumption means the error terms associated with different observations are not related to each other.

**Data Collection:**

The data for test score will be retrieved from the State of New Jersey’s Department of Education website and gathering it from the assessment reports section. I am downloading the High School Proficiency Assessment Spring 2014 and High School Proficiency Assessment Spring 2007 data. Each school is located in one of the twenty-one county in New Jersey. The data files have columns for mean total reading and mean total math for each school. I will need to combine the reading and math score for each school and then gather the average the combine score by county. Some public schools are considered charter schools and are not linked to a county. I will need to find the county for these schools myself and add it in my data file.

Pseudo code to run in Zeppelin to combine the reading and test scores for 2014 and 2007:

% r

2014 \_ hspa < - read. csv ('2 014 hspa. csv ', header = TRUE, sep = ", ")

2014 \_ reading < - as. numeric (2014 \_ hspa $ col 1)

2014 \_ math < - as. numeric (2014 \_ hspa $ col 2)

reading \_ math \_ 2014 < - 2014 \_ reading + 2014 \_ math

% r

2007 \_ hspa < - read. csv ('2 007 hspa. csv ', header = TRUE, sep = ", ")

2007 \_ reading < - as. numeric (2007 \_ hspa $ col 1)

2007 \_ math < - as. numeric (2007 \_ hspa $ col 2)

reading \_ math \_ 2007 < - 2007 \_ reading + 2007 \_ math

For average income the pseudo code to run in Zeppelin to gather the value for each county in 2014 and 2007:

% r

avgincome \_ 2014 <- read.csv('avgincome \_2014.csv', header=TRUE, sep = "," )

avgincome \_ per\_ 2014 <- avgincome\_2014$mean \_ income

% r

avgincome \_ 2007 <- read.csv('avgincome \_2007.csv', header=TRUE, sep = "," )

avgincome \_ per\_ 2007 <- avgincome\_2007$mean \_ income

For race I will divide White alone from Total population from the data set. Pseudo code to run in Zeppelin to gather the percentage value for ethnicity for each county in 2014 and 2007:

% r

ethnicity \_ 2014 <- read.csv('ethnicity \_2014.csv', header=TRUE, sep = "," )

ethnicity \_ per\_2014 <- ethnicity\_2014$white \_ alone / ethnicity \_ 2014 $ total \_ population

% r

ethnicity \_ 2007 <- read.csv('ethnicity \_2007.csv', header=TRUE, sep = "," )

ethnicity \_ per\_2007 <- ethnicity\_2007$white \_ alone / ethnicity \_ 2007 $ total \_ population

For race I will divide married household from total population from the data set. Pseudo code to run in Zeppelin to gather the percentage value for married households for each county in 2014 and 2007:

% r

married \_ 2014 <- read.csv('married \_2014.csv', header=TRUE, sep = "," )

married \_ per\_2014 <- ethnicity\_2014 $ married \_ estimate / ethnicity \_ 2014 $ total \_ households

% r

married \_ 2007 <- read.csv('married \_2007.csv', header=TRUE, sep = "," )

married \_ per\_2007 <- ethnicity\_2007 $ married \_ estimate / ethnicity \_ 2007 $ total \_ households

To gather the percentage value for education attainment I will need to combine Population 18 to 24 years - Bachelor's degree or higher + Population 25 years and over - Bachelor's degree Population 25 years and over - Graduate or professional degree. Pseudo code to run in Zeppelin to gather the percentage value for education for each county in 2014 and 2007:

% r

education \_ 2014 <- read.csv('education\_2014.csv', header=TRUE, sep = "," )

married \_ per\_2014 <- education \_ 2014 $ column \_ x + education \_ 2014$ column \_ y + education \_ 2014 $ column \_ z

% r

education \_ 2007 <- read.csv('education\_2007.csv', header=TRUE, sep = "," )

married \_ per\_2007 <- education \_ 2007 $ column \_ x + education \_ 2007$ column \_ y + education \_ 2007 $ column \_ z

I will need to combine all these data together to be able to run a linear regression. The pseudo code to run in Zeppelin to bind these data:

% r

independent\_variables\_2014 <- rbind ( avgincome \_ 2014, ethnicity \_ 2014, married \_ 2014, education \_ 2014)

% r

independent\_variables\_2007 <- rbind ( avgincome \_ 2007, ethnicity \_ 2007, married \_ 2007, education \_ 2007)

**References**

[1] Dahl, G. B., & Lochner, L. (2012). The impact of family income on child achievement: Evidence from the earned income tax credit. *The American Economic Review*, *102*(5), 1927-1956.

[2] Anger, S., & Heineck, G. (2010). Do smart parents raise smart children? The intergenerational

transmission of cognitive abilities. *Journal of Population Economics*, 23(3), 1255-1282.

[3] Chetty, R., Friedman, J. N., & Rockoff, J. E. (2011). The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood (No. w17699). *National Bureau of Economic Research*.

[4] Chingos, M. M. (2012). Strength in Numbers: State Spending on K-12 Assessment Systems. *Brookings Institution*.

[5] Hoxby, C. M., & Turner, S. (2015). What high-achieving low-income students know about college. *The American Economic Review*, *105*(5), 514-517.

[6] Mayer, S. E. (2010). Revisiting an old question: How much does parental income affect child outcomes. *Focus*, *27*(2), 21-26.

[7] Popham, W. J. (1999). Why standardized tests don't measure educational quality. *Educational leadership*, *56*, 8-16.

[8] Patterson, C. J., Kupersmidt, J. B., & Vaden, N. A. (1990). Income Level, Gender, Ethnicity, and Household Composition as Predictors of Children's School‐based Competence. *Child development*, *61*(2), 485-494.

[9] Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: the indirect role of parental expectations and the home environment. *Journal of family psychology*, *19*(2), 294.

[10] Brooks‐Gunn, J., Klebanov, P. K., & Duncan, G. J. (1996). Ethnic differences in children's intelligence test scores: Role of economic deprivation, home environment, and maternal characteristics. *Child development*, *67*(2), 396-408.

[11] Attanasio, O., & Lechene, V. (2002). Tests of income pooling in household decisions. *Review of economic dynamics*, *5*(4), 720-748.

[12] Duckworth, A. L., Quinn, P. D., & Tsukayama, E. (2012). What No Child Left Behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. *Journal of educational psychology*, *104*(2), 439.

[13] Fleming, C. B., Haggerty, K. P., Catalano, R. F., Harachi, T. W., Mazza, J. J., & Gruman, D. H. (2005). Do social and behavioral characteristics targeted by preventive interventions predict standardized test scores and grades? *Journal of School Health*, *75*(9), 342-349.

[14] Black, S. E., Devereux, P. J., & Salvanes, K. G. (2003). *Why the apple doesn't fall far: Understanding intergenerational transmission of human capital* (No. w10066). National Bureau of Economic Research.

[15] Klebanov, P. K., Brooks-Gunn, J., McCarton, C., & McCormick, M. C. (1998). The contribution of neighborhood and family income to developmental test scores over the first three years of life. *Child development*, 1420-1436.

[16] Bigelow, B. (1999). Why Standardized Tests Threaten Multiculturalism. *Educational Leadership*, *56*(7), 37-40.

[17] Downey, D. B. (1995). When bigger is not better: Family size, parental resources, and children's educational performance. *American Sociological Review*, 746-761.

[18] Ross, C. E., & Broh, B. A. (2000). The roles of self-esteem and the sense of personal control in the academic achievement process. *Sociology of Education*, 270-284.

[19] Nichols, S. L., Berliner, D. C., & Noddings, N. (2007). Collateral damage: How high-stakes testing corrupts America's schools.

[20] Duncan, G. J., Morris, P. A., & Rodrigues, C. (2011). Does money really matter? Estimating impacts of family income on young children's achievement with data from random-assignment experiments.