The Production of Human Capital in Developed Countries: Evidence from 196 Randomized Field Experiments Web Appendices

Contents

1 Appendix A: Data Appendix

We conducted a relatively exhaustive search of all randomized field experiments in education. This appendix describes our search procedure, the selection process, how we categorized the included studies, and the information gathered from each study.

1.1 Search Procedure

This section details our systematic approach to find all field experiments in education.

What Works Clearinghouse

We began by searching all "quick reviews" and "single study reviews" in the What Works Clearinghouse (WWC). WWC was created by the U.S. Department of Education's Institute of Education Sciences in 2002. Its goal is to provide reviews of education studies, policies, and interventions in order for researchers to determine "what works" in education. Currently, WWC has over 10,500 reviews available in an online searchable database. Eligible studies are reviewed by a team of WWC's certified staff against WWC standards and assigned a rating. The highest rating of the Clearinghouse is reserved for studies that meet standards without reservations. This implies that groups compared in the study were determined through a random process, there was low overall attrition from the sample, the differential attrition across groups was low, and there were no confounding factors (that is, no factor is present that all treatment students in one group are exposed to and no students in the comparison group are exposed to. If a confounding factor is present, it would be impossible to distinguish between the effect of the intervention and the effect of the factor). Out search of WWC produced 115 randomized field experiments that met standards without reservations.

Literature Views

We then expanded our search by looking through recent education literature reviews. Specifically, we referenced Almond and Currie (2011), Fryer (2010), Heckman and Kautz (2013), Nye, Turner, and Schwartz (2006), Yeager and Walton (2011), Yoon et al. (2007), Obara (2010), Kretlow and Bartholomew (2010), Carneiro and Heckman (2003), and Heckman (1999).

Databases

Next, we conducted relatively broad searches of known databases that include education papers. Specifically, we queried ERIC, JSTOR, and EconLit. In each database, we searched for all phases generated by concatenating one element from the set of strings [“early childhood”, “education”, “housing”, “neighborhood”, “parent”, “school”, “student”, “teacher”] with one element from [“experiment”, “random assignment”, “randomization”]. For ERIC and EconLit, we collected all hits that searching for these 24 unique phrases returned. For some phrases, JSTOR’s search algorithm returned thousands of results. Due to resource and time constraints, we decided to only collect the top 200 (as determined by “relevance”) results for each phrase in JSTOR. The thousands of hits we found through the database searches are available upon request.

Narrowing the Sample

The methods described above returned over 10,000 citations to check. To conduct this laborious task, we had a team of five research assistants skim every article and select papers that explicitly mentioned a random process determining the experimental sample. Further, if a research assistant determined during their quick read that the paper was obviously not education-related or the experimental sample was post-high school, the study was screened out at this point.

Other Papers

It is important to note that we didn’t restrict ourselves just to the studies produced by the systematic search described above. When reviewing the studies produced by the search procedure above, if we noticed the original study cited a study that would pass the screening criteria, we would include the cited study in our sample for further review. Also, we used our own knowledge of field experiments and advice from our colleagues to catch potential field experiments that our above search missed. Most papers caught in these manners were unpublished working papers.

Using all the above approaches, we found 859 potential studies.

1.2 Inclusion Restrictions

This section details how we narrowed our set of studies from the 859 potential studies to the final analysis sample. Table 1 from the main text summarizes how many papers were excluded for the various exclusion reasons discussed below.

Verifiably Random Process

We only included experiments that had treatment and control groups determined by a verifiably random process. Therefore, quasi-random experiments that were determined by natural processes or studies that compared subject by post hoc matching were excluded from our analysis. Further, if an experiment did not have a control group that continued business-as-usual (i.e. the control group did not receive some sort of dosage that compromised the comparison), the experiment was excluded. Studies dropped for these reasons are labeled as "Design Issues" in Table 1.

Intent-to-Treat Analyses

We only included studies that used the initial randomization assignments to estimate the impact of an intervention. We rejected studies that attempted to use econometric/statistical techniques to correct for mobility after randomization. Studies dropped for this reason are labeled as "Design Issues" in Table 1.

Pre-College Outcomes

We only included experiments with posttreatment outcomes that were collected from children aged 0 to 18. Studies dropped for this reason are labeled as "College Sample/Outcomes" in Table 1.

Highly Developed Country

We only included experiments that took place in highly developed countries. We consider countries as highly developed if they received a classification of "Very High Human Development" in United Nations Development Programme (2010). A country is classified as "Very High Human Development" if they score in the top quartile on an index of human development that includes life expectancy, mean years of schooling, expected years of schooling, and gross national income per capita. Studies dropped for this reason are labeled as "Countries w/o Very High HDI" in Table 1.

Standardized Math or Reading Outcome

We only included experiments that reported norm-referenced reading or mathematics test scores as an outcome measure at posttreatment (e.g. scores from the Peabody Picture Vocabulary Test, the Woodcock-Johnson Test of Achievement, the Iowa Test of Basic Skills, the Stanford Achievement Test, or state tests). Note that studies that report standardized scores as an outcome for later follow-ups but not immediately

following the conclusion of the experiment are excluded from our analysis. Studies dropped for this reason are labeled as "No Standardized Reading or Math" in Table 1.

One Paper Per Experiment

We only included one paper for every experimental randomization. For some experiments, such as the Perry Preschool Project and the Milwaukee voucher program, there are multiple publications detailing the impacts of the experiment at various follow-ups, investigating the initial impacts using different analytical techniques, or just using the data to investigate new theories or statistical methods. In an attempt to not give too much weight to any one experiment, we only included one impact estimate for each experiment. If there were multiple intent-to-treat estimates for an impact, we included the study that first reported the results from the intervention. In addition, our search procedure would sometimes return multiple versions of the same paper. In this case, we would only included the most recent or published version of the paper. Studies dropped for these reasons are labeled as "Repeat Paper" in Table 1.

Other

Some papers did not provide us with enough information to enable us to determine if they should be included or to calculate the effect sizes. These papers were excluded and labeled as "Insufficient Info" in Table 1.

We were unable to locate the text for a small number of the titles that our initial search returned. If research assistants were unable to find a paper through online resources and readily available library resources, we submitted all of the information we obtained through our search to Harvard's Interlibrary Loan system.[1] If Harvard Library staff were unable to locate the title through this process, the paper was excluded and labeled as "Paper Not Located" in Table 1.

Some experiments passed all of the criteria described above, however, we excluded the paper due to the experimental sample being so specific that it did not seem comparable. For example, some studies restricted their samples to special education students with ADHD, autistic students, or delivered speech therapy to students with speech impairments. Studies dropped for this reason are labeled as "Sample Issues" in Table 1.

[1] Harvard Library has cooperative partnerships with other universities and institutions from around the world to locate copies of books or papers requested through this service.

1.3 Categorization

For ease of exposition, we divide the sample of studies into categories and sub-categories. Below we give a brief definition of the three main categories and a number of sub-categories that we reference in the paper. As noted in the main text, the assignment of studies to categories is a bit arbitrary – one could easily argue that some studies fit under multiple categories. Therefore, we provide a table at the end of this section that displays our categorization of all of the studies.

1.3.1 Main Categories

Early Childhood

Any experiment with outcomes measured before children enter kindergarten is categorized as early childhood – independent of the nature of the treatment. Therefore, this category includes experiments that investigate the impacts of preschool attendance, home-based initiatives, and different preschool models on early achievement.

Home

Home environment experiments focus on parenting, income constraints, neighborhood environment, and a student's access to educational resources in their household. Note if an experiment takes place at school and focuses on these inputs, then it is still considered a home-based experiment. For example, parenting classes that take place in a school auditorium are considered a home intervention.

School

School-based experiments target K-12 curricula, teachers, management practices, students, principals, and other school resources. Any experiment where the dosage is applied on students through a school setting – such as offering families vouchers to attend private schools or after-school programs – We categorize as a school-based intervention. Note if an experiment takes place at home and focuses on these inputs, then it is still considered a school-based experiment. For example, if tutors from the school tutor students in their living rooms, this is considered a school-based experiment.

1.3.2 Sub-Categories

Home – Parental Involvement

These experiments investigate the impact of increasing parents' involvement on their children's academic achievement. Treatments that teach parents how to be effective tutors at home, incentivize parents for various behaviors, or give parents information on effective parenting practices are included in this sub-category.

Home – Educational Resources

These experiments investigate the impact of giving children or families household resources that have potential educational returns. Treatments that provide the household of students with books, computers, or internet are included in this sub-category.

Home – Poverty Reduction

These experiments investigate the impact of increasing the income of a student's family. Treatments that increase income through tax reform, direct payments, or by increasing parents' employment (welfare-to-work programs) are included in this sub-category.

Home – Neighborhood Quality

These experiments investigate the impact of neighborhood quality on students' outcomes. Treatments that move students or families from high-poverty to low-poverty neighborhoods or increase the quality of neighborhoods by implementing community programs are included in this sub-category.

School – Student Incentives

These experiments investigate the impact of incentivizing students' educational inputs and/or outputs on students' outcomes. These incentives are financial or non-financial. Treatments that pay students or award them prizes for number of books read, grades on report cards, or scores on standardized test scores are included in this sub-category.

School – High-Dosage Tutoring

These experiments investigate the impact of high-dosage tutoring. We define high-dosage as being tutored in groups of 6 or fewer for more than three days per week or being tutored at a rate that would equate to 50 hours or more over a 36-week period.[2] Note that if the tutor is a child's parent, the experiment is classified as "Home – Parental Involvement".

School – Low-Dosage Tutoring

[2]The definition used in Dobbie and Fryer (2013) is "being tutored in groups of 6 or fewer for more than three days per week." We add to the Dobbie and Fryer (2013) definition because not all studies in our sample report days and group size.

These experiments investigate the impact of low-dosage tutoring. All tutoring programs that do not meet the thresholds described above to be considered high-dosage are labeled low-dosage. Note that if the tutor is a child's parent, the experiment is classified as "Home – Parental Involvement".

School – Teacher Certification

These experiments investigate the impact of teachers obtaining certification through alternative routes or obtaining additional certifications that are not necessary to teach. Teachers obtaining certification through programs such as Teach For America or New York City Teaching Fellows or teachers receiving National Board Certification are included in this sub-category.

School – Teacher Incentives

These experiments investigate the impact of incentivizing teachers to improve student outcomes, move to new schools, or change teaching practices. These incentives can be financial or non-financial and can be awarded to individual teachers or a group of teachers. Treatments that pay teachers for their students' performance on standardized tests, offer teachers bonuses for transferring to low-achieving schools, or give schools financial awards based on predetermined benchmarks (which is then distributed to the teachers) are included in this sub-category.

School – General Professional Development

These experiments investigate the impact of general professional development (PD) programs. Treatments that provide teachers a summer institute or monthly seminars discussing issues such as classroom management or beneficial classroom practices, provide teachers with experienced coaches/mentors, or implement induction programs for new teachers are included in this sub-category. Note that we classify long-term packaged programs in a separate sub-category, "School – Managed Professional Development", discussed below.

School – Managed Professional Development

These experiments investigate the impact of managed PD – packaged programs that have precise training and curriculum materials that schools and districts can implement over an extended period of time in an effort to increase teacher effectiveness. Examples of managed PD include Success for All, Reading Recovery, the Alabama Math, Science, and Technology Initiative, and eMINTS.

School – Data-Driven Instruction

These experiments investigate the impact of using data to guide classroom instruction or school-wide managing practices. Treatments that provide principals with objective progress reports comparing their

teachers' performance to the performance of other teachers throughout the district or implement continuous progress monitoring in classrooms are included in this sub-category.

School – Extended Time

These experiments investigate the impact of exposing students to high quantities of schooling. Treatments that increase the length of school days, enroll students in after-school academic programs, or increase the number of days in a school year are included in this sub-category.

School – Vouchers

These experiments investigate the impact of giving families vouchers that offset some or all of the cost of private-school attendance.

School – School Choice

These experiments investigate the impact of allowing students and families to choose which public schools they attend. Typically, students apply to their choice public school (in the same district or a district close to where they reside) and if a school becomes over-subscribed, admission is determined through a random lottery. Chicago, Illinois and Hartford, Connecticut are examples of cities with school choice programs.

School – Charters

These experiments investigate the impact of charter schools on students. A charter school is a school that receives public funding but operates independently of the established public school system in which it is located. Typically, evaluations are conducted using the random admission lotteries of over-subscribed charter schools.

School – No Excuse Charters

These experiments investigate the impact of charter schools that adopt the "No Excuses" approach on students. These schools emphasize frequent testing, dramatically increased instructional time, parental pledges of involvement, aggressive human capital strategies, a "broken windows" theory of discipline, and a relentless focus on math and reading achievement. As shown in Dobbie and Fryer (2013) and Angrist, Pathak, and Walters (2013), charter schools that adhere to "No Excuses" practices are more effective at increasing students' test scores than other charter schools. Note that this sub-category is a subset of the sub-category "School – Charters".

School – Teaching Strategy

These experiments investigate the impact of changing classroom teaching practices. Treatments that implement individualized instruction, ability-grouped instruction, reciprocal teaching, or smaller class sizes are included in this category.

Note that treatments that implement packaged curricula (e.g. Accelerated Reader, Scott Foresman's Reading Street, Rainbow Reading, and Fluency Formula) or simple curriculum changes are not included in our analysis as they don't align with traditional economic choice variables in a concise way and because of the potential effects of publication bias on these types of studies. Appendix Table 4 describes all such studies we found using our search procedure outlined above.

School – Curriculum

These experiments investigate the impact of changing K-12 curricula. They mostly focus on packaged programs (e.g. Accelerated Reader, Scott Foresman's Reading Street, Rainbow Reading, and Fluency Formula), software products, or simple curriculum changes (e.g. new textbooks, new vocabulary words, and repeated reading). These studies are included in Appendix Table 4 but not described in the text nor included in the meta-analysis, as they don't align with traditional economic choice variables in a concise way and because of the potential effects of publication bias on these types of studies.

Other

Experiments that did not fit into any of the above sub-categories were categorized as "other".

1.3.3 Assigned Categories

See the table below for the main categories and sub-categories assigned to each experiment found through our search process. Note that if an experiment has multiple treatment arms, it is possible for the treatment arms to have different categorizations. Also, if a treatment has characteristics of more than one main category, the experiment is excluded from our meta-analysis in an attempt to avoid interactions of the categories.

Online Appendix Table 1: Categorization

Study	Main Category	Sub-Category
(1)	(2)	(3)
A Comparative Study of the Reading Achievement of Second Grade Pupils in Programs Characterized by a Contrasting Degree of Parent Participation (Ryan, 1964).	Home	Parental Involvement
A Mixed-Method Multi-Level Randomized Evaluation of the Implementation and Impact of an Audio-Assisted Reading Program for Struggling Readers (Lesnick, 2006).	School	Curriculum
A Multisite Cluster Randomized Field Trial of Open Court Reading (Borman et al., 2008).	School	Curriculum
A Multisite Cluster Randomized Trial of the Effects of CompassLearning Odyssey Math on the Math Achievement of Selected Grade 4 Students in the Mid-Atlantic Region (Wijekumar et al. 2009).	School	Curriculum
A Multistate District-Level Cluster Randomized Trial of the Impact of Data-Driven Reform on Reading and Mathematics Achievement (Carlson et al., 2011).	School	Data-Driven
A Randomized Experiment of a Cognitive Strategies Approach to Text-Based Analytical Writing for Mainstreamed Latino English Language Learners in Grades 6-12 (Kim et al., 2011).	School	Other
A Randomized Experimental Evaluation of the Impact of Accelerated Reader/Reading Renaissance Implementation on Reading Achievement in Grades 3 to 6 (Nunnery et al., 2006).	School	Curriculum
A Randomized Field Trial of the Fast ForWorld Language Computer-Based Training Program (Borman et al., 2009)	School	Curriculum

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
A Study of Cooperative Learning in Mathematics, Writing, and Reading in the Intermediate Grades: A Focus Upon Achievement, Attitudes, and Self-Esteem by Gender, Race, and Ability Group (Glassman, 1989).	School	Teaching Strategy
A Study on the Effects of Houghton Mifflin Harcourt's Journeys Program: Year 1 Final Report (Resendez and Azin, 2012)	School	Curriculum
Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots (Abdulkadiroglu et al., 2009). – Charters Treatment	School	Charters, No Excuse Charters
Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots (Abdulkadiroglu et al., 2009). – Pilots Treatment	School	Charters
Action Research: Implementing Connecting Math Concepts (Snider and Crawford, 1996).	School	Curriculum
Addressing Summer Reading Setback Among Economically Disadvantaged Elementary Students (Allington et al, 2010).	Home	Educational Resources
Alternative Routes to Teaching The Impacts of Teach for America (TFA) on Student Achievement and Other Outcomes (Glazerman et al., 2006).	School	Teacher Certification
An Efficacy Study on Scott Foresman's Reading Street Program: Year One Report (Wilkerson et al., 2006).	School	Curriculum
An Evaluation of a Pilot Program in Reading for Culturally Disadvantaged First Grade Students (Bowers, 1972).	School	General PD
An Evaluation of Curriculum, Setting, and Mentoring on the Peformance of Children Enrolled in Pre-Kindergarten (Assel et al., 2006). – DDM Treatment	Early	Early

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
An Evaluation of Curriculum, Setting, and Mentoring on the Peformance of Children Enrolled in Pre-Kindergarten (Assel et al., 2006). – DDN Treatment	Early	Early
An Evaluation of Curriculum, Setting, and Mentoring on the Peformance of Children Enrolled in Pre-Kindergarten (Assel et al., 2006). – LBM Treatment	Early	Early
An Evaluation of Curriculum, Setting, and Mentoring on the Peformance of Children Enrolled in Pre-Kindergarten (Assel et al., 2006). – LBN Treatment	Early	Early
An Evaluation of Reading Recovery (Center et al., 1995).	School	Managed PD
An Evaluation of Teachers Trained Through Different Routes to Certification: Final Report (Constantine et al., 2009).	School	Teacher Certification
An Evaluation of the Effects of Paired Learning in a Mathematics Computer-Assisted-Instruction Program (Turner, 1985). – Individual Treatment	School	Curriculum
An Evaluation of the Effects of Paired Learning in a Mathematics Computer-Assisted-Instruction Program (Turner, 1985). – Paired Treatment	School	Curriculum
An Evaluation of the Teacher Advancement Program (TAP) in Chicago: Year One Impact Report (Glazerman et al., 2009).	School	Teacher Incentives
An Experimental Study of the Effects of the Accelerated Reader Program and a Teacher Directed Program on Reading Comprehension and Vocabulary of Fourth and Fifth Grade Students (Knox, 1996).	School	Curriculum
An Investigation of the Effects of a Comprehensive Reading Intervention on the Beginning Reading Skills of First Graders at Risk for Emotional and Behavioral Disorders (Mooney, 2003).	School	High-Dosage Tutoring

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
An Investigation of the Effects of Daily, Thirty-Minute Home Practice Sessions Upon Reading Achievement With Second Year Elementary Pupils (Hirst, 1972).	Home	Parental Involvement
Are High-Quality Schools Enough to Increase Achievement Among the Poor? Evidence from the Harlem Children's Zone (Dobbie and Fryer, 2011).	School	Charters, No Excuse Charters
Assessing the Effectiveness of First Step to Success: Are Short-Term Results the First Step to Long-Term Behavioral Improvements? (Sumi et al., 2012).	Home, School	Parental Involvement, General PD
Assessment Data - Informed Guidance to Individualize Kindergarten Reading Instruction: Findings from a Cluster-Randomized Control Field Trial (Al Otaiba et al., 2011).	School	Data-Driven
Beyond the Pages of a Book: Interactive Reading and Language Development in Preschool Classrooms (Wasik and Bond, 2001).	Early	Early
Can a Mixed-Method Literacy Intervention Improve the Reading Achievement of Low-Performing Elementary School Students in an After-School Program? Results From a Randomized Controlled Trial of READ 180 Enterprise (Kim et al. 2011).	School	High-Dosage Tutoring
Can Interdistrict Choice Boost Student Achievement? The Case of Connecticut's Interdistrict Magnet School Program (Bifulco et al., 2009).	School	School Choice
Career Academies: Impacts on Students' Engagement and Performance in High School (Kemple and Snipes, 2000).	School	Other
Charter Schools in New York City: Who Enrolls and How it Affects their Students' Achievements (Hoxby, 2009).	School	Charters

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
Children At-Risk for Poor School Readiness: The Effect of an Early Intervention Home Visiting Program on Children and Parents (Necoechea, 2007).	Early	Early
Classroom Assessment for Student Learning: The Impact on Elementary School Mathematics in the Central Region (Randel et al., 2011).	School	General PD
Closing the Achievement Gap: A Structured Approach to Group Counseling (Campbell and Brigman, 2005).	School	Other
Collaboration Between Teachers and Parents in Assisting Children's Reading (Tizard et al., 1982). – Home Treatment	Home	Parental Involvement, Educational Resources
Collaboration Between Teachers and Parents in Assisting Children's Reading (Tizard et al., 1982). – School Treatment	School	Low-Dosage Tutoring
Combining Cooperative Learning and Individualized Instruction: Effects on Student Mathematics Achievement, Attitudes, and Behaviors (Slavin et al., 1984). – Curriculum Treatment	School	Curriculum
Combining Cooperative Learning and Individualized Instruction: Effects on Student Mathematics Achievement, Attitudes, and Behaviors (Slavin et al., 1984). – TAI Treatment	School	Teaching Strategy
Comer's School Development Program in Prince George's County, Maryland: A Theory-Based Evaluation (Cook et al., 1999).	School	General PD
Comparative Effectiveness of Scott Foresman Science: A Report of a Randomized Experiment in Five School Districts (Miller and Jaciw, 2007).	School	Curriculum

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
Comparing Instructional Models for the Literacy Education of High-Risk First Graders (Pinell et al., 1994). – DI Treatment	School	High-Dosage Tutoring
Comparing Instructional Models for the Literacy Education of High-Risk First Graders (Pinell et al., 1994). – RR Treatment	School	Managed PD
Comparing Instructional Models for the Literacy Education of High-Risk First Graders (Pinell et al., 1994). – RS Treatment	School	Managed PD
Comparing Instructional Models for the Literacy Education of High-Risk First Graders (Pinell et al., 1994). – RW Treatment	School	Managed PD
Computer Assisted Instruction as an Enhancer of Remediation (Hotard and Cortez, 1983).	School	Curriculum
Computer-Assisted Instruction to Prevent Early Reading Difficulties in Students at Risk for Dyslexia: Outcomes from Two Instructional Approaches (Torgesen et al., 2009). – LIPS Treatment	School	Curriculum
Computer-Assisted Instruction to Prevent Early Reading Difficulties in Students at Risk for Dyslexia: Outcomes from Two Instructional Approaches (Torgesen et al., 2009). – RWT Treatment	School	Curriculum
Costs, Effects, and Utility of Microcomputer Assisted Instruction (Fletcher et al., 1990).	School	Curriculum
CSRP's Impact on Low-Income Preschoolers' Preacademic Skills: Self-Regulation as a Mediating Mechanism (Raver et al., 2011).	Early	Early
Direct Instruction in Fourth and Fifth Grade Classrooms (Sloan, 1993).	School	General PD

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
Does Rainbow Repeated Reading Add Value to an Intensive Intervention Program for Low-Progress Readers? An Experimental Evaluation (Wheldall, 2000).	School	Curriculum
Does Reading During the Summer Build Reading Skills? Evidence from a Randomized Experiment in 463 Classrooms (Guryan et al., 2014).	Home	Educational Resources
Early College, Early Success: Early College High School Initiative (ECHSI) impact study (Berger et al., 2013).	School	Student Incentives
Early Intervention in Low-Birth-Weight Premature Infants: Results Through Age 5 Years From the Infant Health and Development Program (Brooks-Gunn et al., 1994).	Early	Early
Educational Effects of the Tools of the Mind Curriculum: A Randomized Trial (Barnett et al., 2008).	Early	Early
Effect of Early Literacy Intervention on Kindergarten Achievement (Phillips, 1990). – Home Treatment	Home	Parental Involvement, Educational Resources
Effect of Early Literacy Intervention on Kindergarten Achievement (Phillips, 1990). – Home+School Treatment	Home, School	Parental Involvement, Educational Resources, Curriculum
Effect of Early Literacy Intervention on Kindergarten Achievement (Phillips, 1990). – School Treatment	School	Curriculum
Effect of Technology-Enhanced Continuous Progress Monitoring on Math Achievement (Ysseldyke and Bolt, 2007).	School	Data-Driven

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
Effective Early Literacy Skill Development for Young Spanish-Speaking English Language Learners: An Experimental Study of Two Methods (Farver et al., 2009). – English Treatment	Early	Early
Effective Early Literacy Skill Development for Young Spanish-Speaking English Language Learners: An Experimental Study of Two Methods (Farver et al., 2009). – Transitional Treatment	Early	Early
Effectiveness of Paraeducator-Supplemented Individual Instruction: Beyond Basic Decoding Skills (Vadasy et al., 2007).	School	High-Dosage Tutoring
Effectiveness of Reading and Mathematics Software Products: Findings From Two Student Cohorts (Campuzano et al., 2009). – AN Treatment	School	Curriculum
Effectiveness of Reading and Mathematics Software Products: Findings From Two Student Cohorts (Campuzano et al., 2009). – AoR Treatment	School	Curriculum
Effectiveness of Reading and Mathematics Software Products: Findings From Two Student Cohorts (Campuzano et al., 2009). – DR Treatment	School	Curriculum
Effectiveness of Reading and Mathematics Software Products: Findings From Two Student Cohorts (Campuzano et al., 2009). – Headsprout Treatment	School	Curriculum
Effectiveness of Reading and Mathematics Software Products: Findings From Two Student Cohorts (Campuzano et al., 2009). – LT Treatment	School	Curriculum
Effectiveness of Reading and Mathematics Software Products: Findings From Two Student Cohorts (Campuzano et al., 2009). – Larson Treatment	School	Curriculum
Effectiveness of Reading and Mathematics Software Products: Findings From Two Student Cohorts (Campuzano et al., 2009). – PF Treatment	School	Curriculum

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
Effectiveness of Reading and Mathematics Software Products: Findings From Two Student Cohorts (Campuzano et al., 2009). – WE Treatment	School	Curriculum
Effectiveness of Selected Supplemental Reading Comprehension Interventions: Impacts on a First Cohort of Fifth-Grade Students (James-Burdumy et al., 2009). – Project CRISS Treatment	School	Curriculum
Effectiveness of Selected Supplemental Reading Comprehension Interventions: Impacts on a First Cohort of Fifth-Grade Students (James-Burdumy et al., 2009). – Read for Real Treatment	School	Curriculum
Effectiveness of Selected Supplemental Reading Comprehension Interventions: Impacts on a First Cohort of Fifth-Grade Students (James-Burdumy et al., 2009). – ReadAbout Treatment	School	Curriculum
Effectiveness of Selected Supplemental Reading Comprehension Interventions: Impacts on a First Cohort of Fifth-Grade Students (James-Burdumy et al., 2009). – Reading for Knowledge	School	Curriculum
Effects of a Voluntary Summer Reading Intervention on Reading Achievement: Results From a Randomized Field Trial (Kim, 2006).	Home	Educational Resources
Effects of a Volunteer Tutoring Model on the Early Literacy Development of Struggling First Grade Students (Pullen et al., 2004).	School	High-Dosage Tutoring
Effects of Academic Tutoring on the Social Status of Low-Achieving, Socially Rejected Children (Coie and Krehbie, 1984). – Mentoring Treatment	School	Other
Effects of Academic Tutoring on the Social Status of Low-Achieving, Socially Rejected Children (Coie and Krehbie, 1984). – Tutoring Treatment	School	High-Dosage Tutoring

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
Effects of Academic Tutoring on the Social Status of Low-Achieving, Socially Rejected Children (Coie and Krehbie, 1984). – Tutoring+Mentoring Treatment	School	High-Dosage Tutoring
Effects of an Early Literacy Professional Development Intervention on Head Start Teachers and Children (Powell et al., 2010).	Early	Early
Effects of Health-Related Physical Education on Academic Achievement: Project SPARK (Sallis et al., 1999). – Specialist Treatment	School	Curriculum
Effects of Health-Related Physical Education on Academic Achievement: Project SPARK (Sallis et al., 1999). – Trained Treatment	School	Curriculum
Effects of Intensive Reading Remediation for Second and Third Graders and a 1-Year Follow-Up (Blachman et al., 2004).	School	High-Dosage Tutoring
Effects of Parent Involvement in Isolation or in Combination with Peer Tutoring on Self-Concept and Math (Fantuzzo et al., 1995). – PI Treatment	Home	Parental Involvement
Effects of Parent Involvement in Isolation or in Combination with Peer Tutoring on Self-Concept and Math (Fantuzzo et al., 1995). – PI+RPT Treatment	Home, School	Parental Involvement, Teaching Strategy
Effects of Peer-Assisted Learning Strategies With and Without Training in Elaborated Help Giving (Fuchs et al., 1999). – PALS Treatment	School	High-Dosage Tutoring
Effects of Peer-Assisted Learning Strategies With and Without Training in Elaborated Help Giving (Fuchs et al., 1999). – PALS-HG Treatment	School	High-Dosage Tutoring

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
Effects of Preschool Curriculum Programs on School Readiness (Preschool Curriculum Evaluation Research Consortium, 2008). – BB Treatment	Early	Early
Effects of Preschool Curriculum Programs on School Readiness (Preschool Curriculum Evaluation Research Consortium, 2008). – CC Treatment	Early	Early
Effects of Preschool Curriculum Programs on School Readiness (Preschool Curriculum Evaluation Research Consortium, 2008). – CCorn Treatment	Early	Early
Effects of Preschool Curriculum Programs on School Readiness (Preschool Curriculum Evaluation Research Consortium, 2008). – CCwL Treatment	Early	Early
Effects of Preschool Curriculum Programs on School Readiness (Preschool Curriculum Evaluation Research Consortium, 2008). – DD Treatment	Early	Early
Effects of Preschool Curriculum Programs on School Readiness (Preschool Curriculum Evaluation Research Consortium, 2008). – DLM Treatment	Early	Early
Effects of Preschool Curriculum Programs on School Readiness (Preschool Curriculum Evaluation Research Consortium, 2008). – ELLM Treatment	Early	Early
Effects of Preschool Curriculum Programs on School Readiness (Preschool Curriculum Evaluation Research Consortium, 2008). – LB Treatment	Early	Early
Effects of Preschool Curriculum Programs on School Readiness (Preschool Curriculum Evaluation Research Consortium, 2008). – LE Treatment	Early	Early
Effects of Preschool Curriculum Programs on School Readiness (Preschool Curriculum Evaluation Research Consortium, 2008). – LFC Treatment	Early	Early

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
Effects of Preschool Curriculum Programs on School Readiness (Preschool Curriculum Evaluation Research Consortium, 2008). – PA Treatment	Early	Early
Effects of Preschool Curriculum Programs on School Readiness (Preschool Curriculum Evaluation Research Consortium, 2008). – PC Treatment	Early	Early
Effects of Preschool Curriculum Programs on School Readiness (Preschool Curriculum Evaluation Research Consortium, 2008). – Pre-K Math Treatment	Early	Early
Effects of Preschool Curriculum Programs on School Readiness (Preschool Curriculum Evaluation Research Consortium, 2008). – RSL Treatment	Early	Early
Effects of Reading Decodable Texts in Supplemental First-Grade Tutoring (Jenkins et al., 2004). – Less Treatment	School	High-Dosage Tutoring
Effects of Reading Decodable Texts in Supplemental First-Grade Tutoring (Jenkins et al., 2004). – More Treatment	School	High-Dosage Tutoring
Effects of Targeted Intervention on Early Literacy Skills of At-Risk Students (Wang and Algozzine, 2008).	School	Curriculum
Effects of Whole Class, Ability Grouped, and Individualized Instruction on Mathematics Achievement (Slavin and Karweit, 1985). – AGAT Treatment	School	Teaching Strategy
Effects of Whole Class, Ability Grouped, and Individualized Instruction on Mathematics Achievement (Slavin and Karweit, 1985). – MMP Treatment	School	Curriculum
Effects of Whole Class, Ability Grouped, and Individualized Instruction on Mathematics Achievement (Slavin and Karweit, 1985). – TAI Treatment	School	Teaching Strategy

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
Efficacy of a Direct Instruction Approach to Promote Early Learning (Salaway, 2008).	Early	Early
Efficacy of Collaborative Strategic Reading with Middle School Students (Vaughn et al., 2011).	School	Curriculum
Empirical Evaluation of Read Naturally Effects (Christ and Davie, 2009).	School	Curriculum
Enhancing First-Grade Children's Mathematical Development with Peer-Assisted Learning Strategies (Fuchs et al., 2002).	School	High-Dosage Tutoring
Enhancing Kindergarteners' Mathematical Development: Effects of Peer-Assisted Learning Strategies (Fuchs et al., 2001).	School	Low-Dosage Tutoring
Enhancing the Efficacy of Teacher Incentives Through Loss Aversion (Fryer et al., 2012). – Gain Treatment	School	Teacher Incentives
Enhancing the Efficacy of Teacher Incentives Through Loss Aversion (Fryer et al., 2012). – Loss Treatment	School	Teacher Incentives
Evaluation of Child Care Subsidies: Findings from Project Upgrade in Miami (Layzer et al., 2007). – BELL Treatment	Early	Early
Evaluation of Child Care Subsidies: Findings from Project Upgrade in Miami (Layzer et al., 2007). – Breakthrough treatment	Early	Early
Evaluation of Child Care Subsidies: Findings from Project Upgrade in Miami (Layzer et al., 2007). – Ready Set Leap Treatment	Early	Early
Evaluation of Curricular Approaches to Enhance Preschool Early Literacy Skills (Fischel et al., 2007). – LB Treatment	Early	Early

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
Evaluation of Curricular Approaches to Enhance Preschool Early Literacy Skills (Fischel et al., 2007). – WF Treatment	Early	Early
Evaluation of Experience Corps: Student Reading Outcomes (Morrow-Howell et al., 2009).	School	High-Dosage Tutoring
Evaluation of Quality Teaching for English Learners (QTEL) Professional Development: Final Report (Bos et al., 2012).	School	General PD
Evaluation of the DC Opportunity Scholarship Program: Final Report (Wolf et al., 2010).	School	Vouchers
Evaluation of the Early Start to Emancipation Preparation Tutoring Program in Los Angeles County, CA (Courtney et al., 2008).	School	Low-Dosage Tutoring
Evaluation of the Effectiveness of the Alabama Math, Science, and Technology Initiative (AMSTI) (Newman et al., 2012).	School	Managed PD
Evaluation of the First 3 Years of the Fast Track Prevention Trial with Children at High Risk for Adolescent Conduct Problems (Bierman et al., 2002).	Home, School	Parental Involvement, Curriculum
Evaluation of the i3 Scale-Up of Reading Recovery: Year One Report (May et al., 2013).	School	Managed PD
Evaluation Research on the Effectiveness of Fluency Formula: Final Report (Sivin-Kachala and Bialo, 2005).	School	Curriculum
Experimental Estimates of Education Production Functions (Krueger, 1999).	School	Teaching Strategy

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
Experimental Evidence on the Effects of Home Computers on Academic Achievement among Schoolchildren (Fairlie and Robinson, 2013).	Home	Educational Resources
Explaining Charter School Effectiveness (Angrist et al., 2011).	School	Charters
Final Reading Outcomes of the National Randomized Field Trial of Success for All (Borman et al., 2007).	School	Managed PD
Financial Incentives and Student Achievement: Evidence from Randomized Trials (Fryer, 2011). – Chicago Treatment	School	Student Incentives
Financial Incentives and Student Achievement: Evidence from Randomized Trials (Fryer, 2011). – Dallas Treatment	School	Student Incentives
Financial Incentives and Student Achievement: Evidence from Randomized Trials (Fryer, 2011). – NYC Treatment	School	Student Incentives
Fostering Development of Reading Skills Through Supplemental Instruction: Results for Hispanic and Non-Hispanic Students (Gunn et al., 2005).	Home, School	Parental Involvement, High-Dosage Tutoring
Fostering the Development of Vocabulary Knowledge and Reading Comprehension Through Contextually-Based Multiple Meaning Vocabulary Instruction (Nelson and Stage, 2007).	School	Curriculum
Full-Day versus Half-Day Kindergarten: An Experimental Study (Holmes and McConnell 1990).	School	Extended Time
Getting Parents Involved: A Field Experiment in Deprived Schools (Avvisati et al., 2014).	Home	Parental Involvement

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
Head Start Children's Entry into Public School: A Report on the National Head Start/Public School Early Childhood Transition Demonstration Study (Ramey et al., 2000).	Home, School	Parental Involvement
Head Start Impact Study: Final Report (Puma et al., 2010).	Early	Early
Homework in Arithmetic (Koch, 1965). – full treat	School	Teaching Strategy
Homework in Arithmetic (Koch, 1965). – half treat	School	Teaching Strategy
Impact of eMINTS Professional Development on Student Achievement (Brandt et al., 2013).	School	Managed PD
Impact of Thinking Reade Software Program on Grade 6 Reading Vocabulary, Comprehension, Strategies, and Motivation (Drummond et al., 2011).	School	Curriculum
Impacts of Comprehensive Teacher Induction: Results from the Second Year of a Randomized Controlled Study (Isenberg et al., 2009).	School	General PD
Improving Reading Comprehension and Social Studies Knowledge in Middle School (Vaughn et al., 2013).	School	Curriculum
Improving Reading Fluency and Comprehension in Elementary Students Using Read Naturally (Arvans, 2009).	School	Curriculum
Improving Students' Reading Comprehension Skills: Effects of Comprehension Instruction and Reciprocal Teaching (Spörer et al., 2009). – IG Treatment	School	Teaching Strategy

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
Improving Students' Reading Comprehension Skills: Effects of Comprehension Instruction and Reciprocal Teaching (Spörer et al., 2009). – RT Treatment	School	Curriculum
Improving Students' Reading Comprehension Skills: Effects of Comprehension Instruction and Reciprocal Teaching (Spörer et al., 2009). – RTP Treatment	School	Teaching Strategy
Individualizing a Web-Based Structure Strategy Intervention for Fifth Graders' Comprehension of Nonfiction (Meyer et al., 2011).	School	Curriculum
Information and Employee Evaluation: Evidence from a Randomized Intervention in Public Schools (Rockoff et al., 2012).	School	Data-Driven
Information and Student Achievement: Evidence from a Cellular Phone Experiment (Fryer, 2013). – Incentive Treatment	School	Student Incentives
Information and Student Achievement: Evidence from a Cellular Phone Experiment (Fryer, 2013). – Information Treatment	School	Other
Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments (Fryer, 2014).	School	Charters, No Excuse Charters
KIPP Middle Schools: Impacts on Achievement and Other Outcomes (Tuttle et al., 2013).	School	Charters, No Excuse Charters
Large-Scale Randomized Controlled Trial with 4th Graders Using Intelligent Tutoring of the Structure Strategy to Improve Nonfiction Reading Comprehension (Wijekumar et al., 2012).	School	Curriculum

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
Literacy Learning of At-Risk First-Grade Students in the Reading Recovery Early Intervention (Schwartz, 2005).	School	Managed PD
Longer-Term Impacts of Mentoring, Educational Services, and Learning Incentives: Evidence from a Randomized Trial in the United States (Rodriguez-Planas, 2012).	School	Low-Dosage Tutoring, Student Incentives
Longitudinal Effects of Classwide Peer Tutoring (Greenwood et al., 1989).	School	High-Dosage Tutoring
Longitudinal Results of the Ypsilanti Perry Preschool Project: Final Report (Weikart et al., 1970).	Early	Early
Making Work Pay: Final Report on the Self-Sufficiency Project for Long-Term Welfare Recipients (Michalopoulos et al., 2002).	Home	Poverty Reduction
Mastery Learning and Student Teams: A Factorial Experiment in Urban General Mathematics (Slavin and Karweit, 1984). – Both Treatment	School	Teaching Strategy
Mastery Learning and Student Teams: A Factorial Experiment in Urban General Mathematics (Slavin and Karweit, 1984). – Mastery Treatment	School	Teaching Strategy
Mastery Learning and Student Teams: A Factorial Experiment in Urban General Mathematics (Slavin and Karweit, 1984). – Teams Treatment	School	Teaching Strategy
National Assessment of Title I Interim Report: Volume II: Closing the Reading Gap: First Year Findings from a Randomized Trial of Four Reading Interventions for Striving Readers (Torgesen et al., 2006). – CR Treatment	School	Curriculum
National Assessment of Title I Interim Report: Volume II: Closing the Reading Gap: First Year Findings from a Randomized Trial of Four Reading Interventions for Striving Readers (Torgesen et al., 2006). – FFR Treatment	School	Curriculum

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
National Assessment of Title I Interim Report: Volume II: Closing the Reading Gap: First Year Findings from a Randomized Trial of Four Reading Interventions for Striving Readers (Torgesen et al., 2006). – SR Treatment	School	Curriculum
National Assessment of Title I Interim Report: Volume II: Closing the Reading Gap: First Year Findings from a Randomized Trial of Four Reading Interventions for Striving Readers (Torgesen et al., 2006). – WR Treatment	School	Curriculum
National Board Certification and Teacher Effectiveness: Evidence from a Random Assignment Experiment (Cantrell et al., 2008).	School	Teacher Certification
National Evaluation of Welfare-to-Work Strategies (Hamilton et al., 2001). – HCD Treatment	Home	Poverty Reduction
National Evaluation of Welfare-to-Work Strategies (Hamilton et al., 2001). – LFA Treatment	Home	Poverty Reduction
National Impact Evalutation of the Comprehensive Child Development Program: Final Report (St. Pierre et al., 1997).	Early	Early
Neighborhoods and Academic Achievement: Results from the Moving to Opportunity Experiment (Sanbonmatsu et al., 2006). – Experimental Treatment	Home	Neighborhood Quality
Neighborhoods and Academic Achievement: Results from the Moving to Opportunity Experiment (Sanbonmatsu et al., 2006). – Section 8 Treatment	Home	Neighborhood Quality
Parent Tutoring as a Supplement to Compensatory Education for First Grade Children (Mehran and White, 1988).	Home	Parental Involvement

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
Parent Tutoring in Reading Using Literature and Curriculum Materials: Impact on Student Reading Achievement (Powell-Smith et al., 2000). – CB Treatment	Home	Parental Involvement
Parent Tutoring in Reading Using Literature and Curriculum Materials: Impact on Student Reading Achievement (Powell-Smith et al., 2000). – LB Treatment	Home	Parental Involvement
Parental Incentives and Early Childhood Achievement: A Field Experiment in Chicago Heights (Fryer et al., 2015). – Cash treatment	Early	Early
Parental Incentives and Early Childhood Achievement: A Field Experiment in Chicago Heights (Fryer et al., 2015). – College treatment	Early	Early
Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores (Bettinger, 2012).	School	Student Incentives
Poverty, Early Childhood Education, and Academic Competence: The Abecedarian Experiment (Ramey and Campbell, 1991).	Early	Early
Prevention and Remediation of Severe Reading Disabilities: Keeping the End in Mind (Torgesen et al., 1997). – EP Treatment	School	High-Dosage Tutoring
Prevention and Remediation of Severe Reading Disabilities: Keeping the End in Mind (Torgesen et al., 1997). – PASP Treatment	School	High-Dosage Tutoring
Prevention and Remediation of Severe Reading Disabilities: Keeping the End in Mind (Torgesen et al., 1997). – RCS Treatment	School	High-Dosage Tutoring
Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program (Rouse, 1998).	School	Vouchers

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
Project Breakthrough: A Responsive Environment Field Experiment with Pre-School Children from Public Assistance Families (Cook County Department of Public Aid, 1969).	Early	Early
Promoting Academic and Social-Emotional School Readiness: The Head Start REDI Program. (Bierman et al., 2008).	Early	Early
Putting Books in the Classroom Seems Necessary But Not Sufficient (McGill-Franzen et al., 1999). – Books Treatment	School	Other
Putting Books in the Classroom Seems Necessary But Not Sufficient (McGill-Franzen et al., 1999). – Books+Training Treatment	School	General PD
Randomized Field Trial of an Early Literacy Curriculum and Instituional Support System (Cosgrove et al., 2006).	Early	Early
Reading and Language Outcomes of a Multiyear Randomized Evaluation of Transitional Bilingual Education (Slavin et al., 2011).	School	Curriculum
Repeated Reading Intervention: Outcomes and Interactions with Readers' Skills and Classroom Instruction (Vadasy and Sanders, 2008).	School	High-Dosage Tutoring
School Choice as a Latent Variable: Estimating the "Complier Average Causal Effect" of Vouchers in Charlotte (Cowen, 2008).	School	Vouchers
School Choice in Dayton, Ohio after Two Years: An Evaluation of the Parents Advancing Choice in Education Scholarship Program (West et al., 2001).	School	Vouchers
School Choice in New York City After Three Years: An Evaluation of the School Choice Scholarships Program (Mayer et al., 2002).	School	Vouchers

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
Segmentation / Spelling Instruction as Part of a First-Grade Reading Program: Effects on Several Measures of Reading (Uhry and Shepherd, 1993).	School	Curriculum
Spatial Temporal Mathematics at Scale: An Innovative and Fully Developed Paradigm to Boost Math Achievement Among All Learners (Rutherford et al., 2010).	School	Curriculum
Summer School Effects in a Randomized Field Trial (Zvoch and Stevens, 2012).	School	Extended Time
Supporting Families in a High-Risk Setting: Proximal Effects of the SAFEChildren Preventive Intervention (Tolan et al., 2004).	Home, School	Parental Involvement, Low-Dosage Tutoring
Teacher Behavior and Pupil Performance: Reconsideration of the Mediation of Pygmalion Effects (Alpert, 1975).	School	Teaching Strategy
Teacher Incentives and Student Achievement: Evidence from New York City Public Schools (Fryer, 2013).	School	Teacher Incentives
Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching (Springer et al., 2010).	School	Teacher Incentives
Teacher Study Group: Impact of the Professional Development Model on Reading Instruction and Student Outcomes in First Grade Classrooms (Gersten et al., 2010).	School	General PD
Teaching Children to Become Fluent and Automatic Readers (Kuhn et al., 2006). – Repeated-Reading Treatment	School	Curriculum

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
Teaching Children to Become Fluent and Automatic Readers (Kuhn et al., 2006). – Wide-Reading Treatment	School	Curriculum
Team Pay for Performance: Experimental Evidence From the Round Rock Pilot Project on Team Incentives (Springer et al., 2012).	School	Teacher Incentives
Technology's Edge: The Educational Benefits of Computer-Aided Instruction (Barrow et al., 2009).	School	Curriculum
The (Surprising) Efficacy of Academic and Behavorial Intervention with Disadvantaged Youth Results from a Randomized Experiment in Chicago (Cook et al., 2014). – BAM Treatment	School	Other
The (Surprising) Efficacy of Academic and Behavorial Intervention with Disadvantaged Youth Results from a Randomized Experiment in Chicago (Cook et al., 2014). – BAM+Tutoring Treatment	School	High-Dosage Tutoring
The Early Training Project for Disadvantaged Children: A Report After Five Years (Klaus and Gray, 1968).	Early, School	Early
The Effect of Computer Assisted Instruction in Improving Mathematics Performance of Low Achieving Ninth Grade Students (Bailey, 1991).	School	Curriculum
The Effect of School Choice on Participants: Evidence from Randomized Lotteries (Cullen et al., 2006).	School	School Choice
The Effect of Second-Language Instruction on the Reading Proficiency and General School Achievement of Primary-Grade Children. (Potts, 1967).	School	Curriculum
The Effective Instruction of Comprehension: Results and Description of the Kamehameha Early Education Program (Tharp and Roland, 1982).	School	Teaching Strategy

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
The Effectiveness of a Program to Accelerate Vocabulary Development in Kindergarten (VOCAB) (Goodson et al., 2010).	School	Curriculum
The Effectiveness of Computer Assisted Instruction of Chapter I Students in Secondary Schools (Davidson, 1985).	School	Curriculum
The Effectiveness of Extended Day Programs: Evidence from a Randomized Field Experiment in the Netherlands (Meyer and Klaveren, 2013).	School	Extended Time
The Effectiveness of Secondary Math Teachers from Teach for America and the Teaching Fellows Programs (Clark et al., 2013). – TFA Treatment	School	Teacher Certification
The Effectiveness of Secondary Math Teachers from Teach for America and the Teaching Fellows Programs (Clark et al., 2013). – Teaching Fellows Treatment	School	Teacher Certification
The Effectiveness of Team-Accelerated Instruction on High Achievers in Mathematics (Karper and Melnick, 1993).	School	Teaching Strategy
The Effects of a Language and Literacy Intervention on Head Start Children and Teachers (Wasik et al., 2006).	Early	Early
The Effects of a Negative Income Tax on School Performance: Results of an Experiment (Maynard and Murname, 1979).	Home	Poverty Reduction
The Effects of A One-Year Staff Development Program on the Achievement Test Scores of Fourth Grade Students (Cole, 1992).	School	General PD
The Effects of a Voluntary Summer Reading Intervention on Reading Activities and Reading Achievement (Kim, 2007).	Home	Educational Resources

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
The Effects of Brain Gym as a General Education Intervention: Improving Academic Performance and Behaviors (Nussbaum, 2010).	School	Other
The Effects of Computer Assisted Instruction as a Supplement to Classroom Instruction in Reading Comprehension and Arithmetic (Easterling, 1982). – Mathematics Treatment	School	Curriculum
The Effects of Computer Assisted Instruction as a Supplement to Classroom Instruction in Reading Comprehension and Arithmetic (Easterling, 1982). – Reading Treatment	School	Curriculum
The Effects of Peer-Assisted Literacy Strategies for First-Grade Readers With and Without Additional Mini-Skills Lessons (Mathes and Babyak, 2001). – PALS Treatment	School	High-Dosage Tutoring
The Effects of Peer-Assisted Literacy Strategies for First-Grade Readers With and Without Additional Mini-Skills Lessons (Mathes and Babyak, 2001). – PALS+ML Treatment	School	High-Dosage Tutoring
The Effects of Structured One-on-One Tutoring in Sight Word Recognition of First-Grade Students At-Risk for Reading Failure (Mayfield, 2000).	School	High-Dosage Tutoring
The Effects of the Home Instruction Program for Preschool Youngsters (HIPPY) on Children's School Performance at the End of the Program and One Year Later (Baker et al., 1998).	Early	Early
The Effects of Theoretically Different Instruction and Student Characteristics on the Skills of Struggling Readers (Mathes et al., 2005). – Proactive Treatment	School	High-Dosage Tutoring

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
The Effects of Theoretically Different Instruction and Student Characteristics on the Skills of Struggling Readers (Mathes et al., 2005). – Responsive Treatment	School	High-Dosage Tutoring
The Effects of Training Parents in Teaching Phonemic Awareness on the Phonemic Awareness and Early Reading of Struggling Readers (Warren, 2009).	Home	Parental Involvement
The Efficacy of an Early Literacy Tutoring Program Implemented by College Students (Allor and McCathren, 2004).	School	High-Dosage Tutoring
The Enhanced Reading Opportunities (ERO) Study Final Report: The Impact of Supplemental Literacy Courses for Struggling Ninth-grade Readers (Somers et al., 2010).	School	Curriculum
The Evaluation of Charter School Impacts: Final Report (Gleason et al., 2010).	School	Charters
The Evaluation of Enhanced Academic Instruction in After-School Programs: Final Report (Black et al., 2009).	School	Extended Time
The Impact of a Literature-Based Program on Literacy Achievement, Use of Literature, and Attitudes of Children from Minority Backgrounds (Morrow, 1992). – Home+School Treatment	Home, School	Parental Involvement
The Impact of a Literature-Based Program on Literacy Achievement, Use of Literature, and Attitudes of Children from Minority Backgrounds (Morrow, 1992). – School Treatment	School	Curriculum
The Impact of Challenging Geometry and Measurement Units on Achievement of Grade 2 Students (Gavin et al., 2013).	School	Curriculum

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
The Impact of Collaborative Strategic Reading on the Reading Comprehension of Grade 5 Students in Linguistically Diverse Schools (Hitchcock et al., 2011).	School	Curriculum
The Impact of Elementary Mathematics Coaches on Student Achievement (Campbell and Malkus, 2011).	School	General PD
The Impact of Indiana's System of Interim Assessments on Mathematics and Reading Achievement (Konstantopoulos et al., 2013).	School	Data-Driven
The Impact of Parental Training in Methods to Aid Beginning Reading on Reading Achievement and Reading Attitudes of First-Grade Students. (Peeples, 1996).	Home	Parental Involvement
The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement (Garet et al. 2008). – Institute Series Treatment	School	General PD
The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement (Garet et al. 2008). – Institute Series+Coaching Treatment	School	General PD
The Influence of Massive Rewards on Reading Achievement in Potential Urban School Dropouts (Clark and Walberg, 1968).	School	Student Incentives
The Missouri Mathematics Effectiveness Project: An Experimental Study in Fourth-Grade Classrooms (Good and Grouws, 1979).	School	Curriculum
The Potential of Ubran Boarding Schools for the Poor: Evidence from SEED (Curto and Fryer, 2014).	School	Charters, No Excuse Charters

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
The Prevention, Identification, and Cognitive Determinants of Math Difficulty (Fuchs et al., 2005).	School	High-Dosage Tutoring
The Reading Connection: A Leadership Initiative Designed to Change the Delivery of Educational Services to At-Risk Children (Compton, 1992).	School	High-Dosage Tutoring
The Relationship Between Supplemental Computer Assisted Mathematics Instruction and Student Achievement (Manuel, 1987). – Apple Treatment	School	Curriculum
The Relationship Between Supplemental Computer Assisted Mathematics Instruction and Student Achievement (Manuel, 1987). – CCC Treatment	School	Curriculum
Towards Reduced Poverty Across Generations: Early Findings from New York City's Conditional Cash Transfer Program (Riccio et al., 2010).	Home	Parental Involvement
Transfer Incentives for High-Performing Teachers: Final Results from a Multisite Randomized Experiment (Glazerman et al., 2013).	School	Teacher Incentives
Two-Year Impacts of a Universal School-Based Social-Emotional and Literacy Intervention (Jones, et al., 2011).	School	Curriculum
Using Enrichment Reading Practices to Increase Reading Fluency, Comprehension, and Attitudes (Reis et al., 2008).	School	Curriculum
Using Knowledge of Children's Mathematics Thinking in Classroom Teaching: An Experimental Study (Carpenter et al., 1989).	School	General PD
Using Television as a Teaching Tool: The Impacts of Ready to Learn Workshops on Parents, Educators, and the Children in their Care (Boller et al., 2004).	Early	Early

Online Appendix Table 1 (continued)

Study	Main Category	Sub-Category
(1)	(2)	(3)
When Less May Be More: A 2 Year Longitudinal Evaluation of a Volunteer Tutoring Program Requiring Minimal Training (Baker et al., 2000).	School	Low-Dosage Tutoring
When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Centers Program (James-Burdumy et al., 2005).	School	Extended Time

Notes: This table presents the main categories and sub-categories assigned to each treatment from papers we found that passed our inclusion criteria. These categories are described in the text and Online Appendix A. Note that if a treatment fit into multiple main categories, the treatment was not included in our meta-analysis. Further, curriculum studies were not included in our meta-analysis.

1.4 Data Collected

For every randomized field experiment found using the search procedure described above, we calculated the impact (in standard deviations) of the intervention on standardized math and reading outcomes and collected data on key demographic and implementation features of the experiment. This section details all of the information we collected for each experiment.

1.4.1 Effect Sizes

We calculated estimates of the pooled effect sizes on reading and/or math test scores in standard deviations for each experiment that passed our inclusion restrictions. Studies reported results in a variety of ways and we had to manipulate these results in order to have comparable impacts across all experiments. Below are some of the common calculations we performed.

Scale Scores

If impacts were presented as scale score points on a test, we would divide the coefficient by the standard deviation given in the summary statistics. If no standard deviation was given in the paper, we would instead use the standard deviation from a national or norming sample.

Multiple Measures

When a study reported math or reading impacts for multiple standardized measures, we would average the impacts across all standardized measures for each subject.

Subsamples

When a study reported impacts by subsamples (e.g. by grade, by race, by cohort, etc.) and did not report pooled estimates, we would report the weighted average of the impacts across the given subsamples.

Hedge's g

When a study only reported means and standard deviations, we used this information to calculate a statistic known as Hedge's g and its corresponding standard error (see Hedges 1981 and Lipsey and Wilson 2000). In cases where studies reported impacts but did not provide enough information to estimate impacts and standard errors in standard deviation units (a common example of this was a paper reporting the impact but failing to provide any standard errors or p-values), we would instead calculate Hedge's g with reported means and standard deviations.

Standard Errors

Unfortunately, without having access to the micro-data, it was not possible to calculate the appropriate standard errors for every effect size. In an attempt to not overstate the significance of an effect size, we were overly conservative when calculating standard errors that were not already reported in a study. For example, when calculating Hedge's g, we used the number of units randomized to calculate the standard errors. Although Slavin et al. (1984) had a sample of 504 students, randomization was done at the school level (N = 6) and hence the standard errors reported in our tables were large.

In cases where p-values were given for impacts instead of standard errors, we would assume the p-value was calculated using a normal distribution and back out an estimate of the standard error.[3]

Annual Impacts

For comparability across all studies, we only used annual impacts in our meta-analysis. When a study lasted for multiple years and only reported cumulative impacts, we would divide the study by the length of the intervention to estimate annual impacts. For standard errors in this case, we divided the cumulative standard error by the square root of the length of the intervention.[4]

Other

We documented all assumptions and calculations we made for each study and these files can be obtained upon request. Unique cases that did not utilize some combination of the methods above were rare and were dealt with on a case by case basis. Note that if there was not enough information presented in a paper for us to make credible assumptions, the study was excluded.

1.4.2 Demographic Variables

For each demographic variable described below, if the paper did not provide enough information for us to determine the quantity of interest, we recorded the value as missing. Any assumptions made to calculate these variables were recorded in a text field included in the final column of the dataset.

Age

The age range of students in the experiment.

Grade

[3] Usually not enough information was given for us to estimate the degrees of freedom of a t-distribution.

[4] This follows from $\text{Var}\left(\sum_{i=1}^{N} X_i\right) = \sum_{i}^{N} \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}\left(X_i, X_j\right)$ where X_i represents an impact in year i. Assuming equal variances and non-negative covariances, it follows that estimates of annual standard errors have an upper bound of $\text{SE}\left(\sum_{i=1}^{N} X_i\right) / \sqrt{N}$ (when $\sum_{i \neq j} \text{Cov}\left(X_i, X_j\right) = 0$). We chose this calculation for our estimates as it is the most conservative under these assumptions.

The grade range of students in the experiment.

Note that since studies typically only report grade or age, not both, we usually only have data on one or the other. In our analysis, for studies that are missing age data, we impute the age assuming the typical age-grade mapping of the American education system. We assign to every grade the age students typically turn in the middle of that grade (6 in kindergarten, 7 in first grade, and so forth). If preschool/early childhood experiments do not report an age, I assign them an average age of 4.5 years. Our age results are similar when using beginning of year age for the imputations.

English-Language Learner

An indicator that is 1 if a majority of the students in the experimental sample are English-Language Learners and 0 otherwise.

Disadvantaged

An indicator that is 1 if a majority of the students in the experimental sample are disadvantaged and 0 otherwise. There is unfortunately no industry standard for what constitutes "disadvantaged" and papers wildly differed in their definitions and amount of information presented to the readers. Therefore, for this data collection, we considered a sample disadvantaged if authors emphasized that a majority of students in their sample came from an environment that would lead readers to believe they are substantially below average with respect to poverty (e.g. points out they are "low-socioeconomic status", "at-risk", "low-income", majority have "free or reduced-price lunch status", etc.) or presented summary statistics that enable the reader to easily reach this conclusion.

Black

An indicator that is 1 if a majority of the students in the experimental sample are black and 0 otherwise.

Hispanic

An indicator that is 1 if a majority of the students in the experimental sample are Hispanic and 0 otherwise.

Low-Ability

An indicator that is 1 if an experiment targets students of low-ability and 0 otherwise. There is unfortunately no industry standard for what constitutes "low-ability" and papers wildly differed in their definitions and amount of information presented to the readers. Therefore, for this data collection, We considered a

sample low-ability if authors emphasized that a majority of students in their sample were substantially below average on typical achievement measures (e.g. points out a majority are "behind grade-level", only included the two worst students from every classroom, "Normal Curve Equivalent scores are below 30", etc.) or presented summary statistics that enable the reader to easily reach this conclusion.

1.4.3 Implementation Details

For each implementation variable described below, if the paper did not provide enough information for us to determine the quantity of interest, we recorded the value as missing. Any assumptions made to calculate these variables were recorded in a text field included in the final column of the dataset.

First Year of Experiment

The year the experiment was first implemented in the field.

Year of Publication

The year the paper was published. For unpublished work (such as working papers and dissertations), we report the year of the most recent version found.

Length of Treatment

The amount of time the average cohort was exposed to the experiment. During the data collection process, we record this in the same unit the author uses (days, weeks, months, semesters, years, etc.) and differentiate between academic years and calendar years. In the cleaned data, I convert all lengths to calendar years using typical assumptions (e.g. a calendar year is equivalent to 12 months, 52 weeks, or 365 days; an academic year is equivalent to 9 months; a semester is half of an academic year; a summer recess is 3 months; etc.). See the code for all conversions used.

Location

The location of the experiment. Authors varied in the amount of detail provided to readers. Although some would present the exact name of the schools, school-districts, cities, or states that their experiment was implemented in, some were more vague and only gave characterizations such as "a small rural school", if anything at all. Using the information the authors present, I created 7 location indicators (1 if an experiment took place in the given location and 0 otherwise):

- U.S.A. Northeast Region – An experiment took place in Maine, Massachusetts, Rhode Island, Connecticut, New Hampshire, Vermont, New York, Pennsylvania, New Jersey, Delaware, or Maryland.

- U.S.A. Southeast Region – An experiment took place in West Virginia, Virginia, Kentucky, Tennessee, North Carolina, South Carolina, Georgia, Alabama, Mississippi, Arkansas, Louisiana, or Florida.
- U.S.A. Southwest Region – An experiment took place in Texas, Oklahoma, New Mexico, or Arizona.
- U.S.A. Midwest Region – An experiment took place in Ohio, Indiana, Michigan, Illinois, Missouri, Wisconsin, Minnesota, Iowa, Kansas, Nebraska, South Dakota, or North Dakota.
- U.S.A. West Region – An experiment took place in Colorado, Wyoming, Montana, Idaho, Washington, Oregon, Utah, Nevada, California, Alaska, or Hawaii.
- U.S.A. National – An experiment was a national evaluation that took place across many regions.
- Foreign – An experiment took place in a country other than the U.S.A.

Note that if an experiment was not designed to be a national evaluation but spanned multiple regions, the experiment would have two or more regional indicators with a value of 1.

Number of Standardized Math Constructs

The number of standardized math outcome measures authors collect at posttest.

Number of Standardized Reading Constructs

The number of standardized reading outcome measures authors collect at posttest.

Randomization Unit

The unit researchers used for their randomization (e.g. student, school, family, classroom, teacher, etc.).

Number of Randomization Units

The number of students/schools/families/classrooms/etc. that were randomly assigned to treatment or control at the beginning of the experiment.

Number of Units Randomized into Treatment

The number of randomization units that were randomly assigned to treatment at the beginning of the experiment.

Number of Units Randomized into Control

The number of randomization units that were randomly assigned to control at the beginning of the experiment.

Number of Students in Sample at Randomization

The number of students who were randomly allocated to treatment or control at the beginning of the experiment. Note that in many cases, students were assigned to treatment or control because they were part of a larger unit that was randomized (e.g. schools, classrooms, families, etc.).

Number of Students in Sample at Post-Test

The number of students present at the end of the experiment and who have non-missing standardized math or reading outcomes.

Subject Focus of Experiment

If the focus of the experiment is math achievement , reading achievement, both math and reading, or not subject related.

Type of Publication

Whether the experiment was written up in a peer-reviewed journal, a dissertation, an unpublished working paper, a government-funded publication, a firm-funded publication, or "other".

Tutoring Hours

For experiments that had a tutoring element, the number of hours of tutoring a student would receive in 36 weeks if the rate of tutoring continued at the same pace.

Individual Tutoring

For experiments that had a tutoring element, an indicator that was 1 if tutoring was one-on-one and 0 otherwise.

2 Appendix B: Life-Cycle Model

This appendix describes the life-cycle model used in Section 4 to investigate the long term impacts of the best education experiments found through our literature search. The model draws heavily from the Social Genome Model (SGM) described in Winship and Owen (2013). However, due to the lack of source code available – even upon request – and the limited description in the available guide, our procedure for creating the dataset and running the simulation may slightly diverge from the methods in Winship and Owen (2013). We describe our procedure below so that any deviations are apparent. Further, our source code and data are available online so researchers can easily adapt the model to their own needs.

2.1 Model

The model is identical to the model described in Winship and Owen (2013) and is reiterated here.

The simple theoretical model assumes that cognitive and non-cognitive skill formation varies across an individual's life and is dependent on the stock of skills in previous stages of life. Specifically, Winship and Owen (2013) define six different life-stages: circumstances at birth (CAB), early childhood (EC), middle childhood (MC), adolescence (AD), transition to adulthood (TTA), and adulthood (AH). The model then assumes that every outcome in a given stage depends on all revealed outcomes from the stages preceding it. Formally, given a vector of circumstances at birth, CAB_i, for individual i, each outcome in the vector of early childhood outcomes, EC, is modeled as

$$EC\ Outcome_i = \beta_0^{ec} + \beta_{cab}^{ec} CAB_i + \varepsilon_i^{ec}.$$

Similarly, each of the MC outcomes is given by

$$MC\ Outcome_i = \beta_0^{mc} + \beta_{cab}^{mc} CAB_i + \beta_{ec}^{mc} EC_i + \varepsilon_i^{mc}.$$

For the adolescent life-stage we have

$$AD\ Outcome_i = \beta_0^{ad} + \beta_{cab}^{ad} CAB_i + \beta_{ec}^{ad} EC_i + \beta_{mc}^{ad} MC_i + \varepsilon_i^{ad}.$$

Outcomes when transitioning to adulthood would be

$$TTA\ Outcome_i = \beta_0^{tta} + \beta_{cab}^{tta} CAB_i + \beta_{ec}^{tta} EC_i + \beta_{mc}^{tta} MC_i + \beta_{ad}^{tta} AD_i + \varepsilon_i^{tta}.$$

And finally, adult outcomes are modeled as

$$AH\ Outcome_i = \beta_0^{ah} + \beta_{cab}^{ah} CAB_i + \beta_{ec}^{ah} EC_i + \beta_{mc}^{ah} MC_i + \beta_{ad}^{ah} AD_i + \beta_{tta}^{ah} TTA_i + \varepsilon_i^{ah}.$$

Where β_ψ^λ are the partial correlations of realized outcomes from the ψ life-stage ("0" represents an intercept) with the given LHS outcome in the λ life-stage.

2.2 Simulation

2.2.1 Data

Unfortunately, as discussed by Winship and Owen (2013), there is not yet a dataset with rich enough information that follows an individual from birth through adult outcomes. Therefore, in order to conduct simulations using the above model, we combine two well known public datasets: the National Longitudinal Survey of Youth 1979 (NLSY79) and the NLSY79 Child and Young Adult survey (CNLSY). The NLSY79 follows a nationally representative sample of 12,686 men and women who were between the ages of 14 and 22 when they were first interviewed in 1979. The sample was interviewed annually through 1994 and then biennially thereafter. The CNLSY follows all children born to the female respondents in the NLSY79.[5] These children were first interviewed in 1986 and then biennially thereafter.

Combining these two datasets together, we have a rich set of outcomes for each life-stage discussed above. From the CNLSY, we observe CAB, EC, MC and AD outcomes. From the NLSY79, we observe TTA and AH outcomes. Importantly, a subset of the CAB and AD outcomes exist in both datasets that allow us to link the datasets together in the simulation described below. Table 3 details the specific variables that were used for each life-stage and what datasets they were available in. The variables include a mix of cognitive skills (e.g. standardized test scores), non-cognitive skills (e.g. self esteem and hyperactivity indices), and important life outcomes (e.g. teen birth, drug use, and graduation).[6] Tagsets that can be used to download the raw data from the Bureau of Labor Statistics and the code used to create these variables

[5] As of the most recent survey with data available (2012), there were 11,512 CNLSY respondents ranging in age from 1 to 42.

[6] Note that in order to protect respondents' identities, the NLSY79 top-coded all income variables and the numerical cutoff for the top-code varied over the years. We first convert all income values to 2010 dollars and then re-top-code all income variables to the minimum (in 2010 dollars) top-code that the NLSY79 ever used.

from the raw data are available on this paper's companion website.[7]

We restrict the NLSY79 sample to only include the 6,111 respondents in a cross-sectional sample designed to represent the non-institutionalized civilian segment of people living in America at the time of the first interview. This drops respondents that were in a supplemental minority and poor sample (5,295 respondents) or in a supplemental military sample (1,280 respondents). Further, we limit our NLSY79 sample to respondents with valid race information.[8] Similarly, we restrict the CNLSY sample to only include children of the NLSY79 sample we defined above and who had valid race information themselves.[9] All analyses described below use unweighted data.

2.2.2 Imputation

For all outcomes, we use a simple procedure to impute values for respondents with missing values.[10] Within a dataset, we sort outcomes within each life-stage in increasing order of percent of missing responses. Starting with the youngest life-stage available in a given dataset, we then use the outcomes with no missing responses[11] as explanatory variables in a linear model[12] to predict the missing values of the outcome in that life-stage with the least amount of missing responses. We then include this newly imputed variable in the set of explanatory variables and use this set to predict the next outcome in that life-stage. We continue this procedure until all missing values in the youngest life-stage are imputed. We then include all of these variables as explanatory variables to predict the missing values of the variable in the second youngest life-stage with the least amount of missing responses. We add this imputed variable to the set of explanatory variables and continue this procedure until all values in a dataset have been imputed. After this procedure, we round all binary and categorical variables to the nearest integer.[13][14] This procedure is done separately for both of the CNLSY and NLSY79 datasets.

See Winship and Owen (2013) for a discussion of the imputation methods. Online Appendix Table 2

[7] Raw data from the Bureau of Labor Statistics can be found here https://www.nlsinfo.org/investigator/.

[8] This drops an additional 13 respondents from the sample.

[9] 5,791 children in the CNLSY had mothers from our analysis sample. Of these, 3 were dropped because they were missing race information.

[10] Missing in this case means the respondent did not have a valid response for any ages around the given life-stage. See Table 3 and the code generating the variables for the specific age ranges for each variable.

[11] In the CNLSY, race, gender, and mother's age at first birth have no missing responses. In the NLSY79, race and gender have no missing responses.

[12] Unlike Winship and Owen (2013), all outcomes are predicted using OLS. For example, this means that binary outcomes are predicted using linear probability models instead of logit or probit models

[13] The binary and categorical variables are gender, race, marital status of parents, low birth weight, high school grad status, criminal conviction, teen parent, lives independently from parents, marijuana use, other drug use, early sex, suspension, fighting, hitting, damaging property, religious service attendance, school clubs, and college completion.

[14] The continuous variables are maternal educational attainment (we impute grade), maternal age at birth, maternal age at first birth, family income, mother's AFQT score, cognitive stimulation score, emotional support score, PPVT score, math achievement, reading achievement, antisocial behavior, hyperactivity, GPA, self-esteem index, and gender role attitudes.

presents summary statistics for each outcome pre- and post-imputation.

Online Appendix Table 2: Imputation Statistics

	Before Imputation			After Imputation		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: CNLSY						
Male	0.517	0.500	5,788	0.517	0.500	5,788
Black	0.140	0.347	5,788	0.140	0.347	5,788
Hispanic	0.108	0.310	5,788	0.108	0.310	5,788
Other Race	0.042	0.201	5,788	0.042	0.201	5,788
Mom Age at Birth	26.257	6.030	5,787	26.259	6.032	5,788
Mom Age at First Birth	23.020	5.482	5,788	23.020	5.482	5,788
Mom Married at Birth	0.810	0.392	5,526	0.806	0.395	5,788
Family Income (Birth)	2.978	2.270	5,139	2.927	2.328	5,788
Low Birth Weight	0.078	0.268	5,180	0.077	0.266	5,788
Mom's AFQT Score	45.829	28.847	5,520	45.664	28.682	5,788
Cognitive Stimulation Score	0.000	0.999	4,593	-0.066	1.002	5,788
Emotional Support Score	0.000	0.999	4,551	-0.062	1.017	5,788
PPVT Score	0.000	1.000	2,836	-0.045	1.019	5,788
Math Achievement ($\approx$ Age 5)	-0.068	0.996	4,563	-0.091	0.997	5,788
Reading Achievement ($\approx$ Age 5)	-0.070	1.003	4,553	-0.105	0.981	5,788
Antisocial Behavior ($\approx$ Age 5)	0.009	1.002	4,930	0.019	0.989	5,788
Hyperactivity ($\approx$ Age 5)	0.021	1.004	4,944	0.046	1.012	5,788
Math Achievement ($\approx$ Age 10)	-0.007	1.005	4,430	-0.008	0.991	5,788
Reading Achievement ($\approx$ Age 10)	-0.005	0.999	4,433	-0.005	0.986	5,788
Antisocial Behavior ($\approx$ Age 10)	-0.002	1.003	4,665	0.002	0.979	5,788
Hyperactivity ($\approx$ Age 10)	-0.005	1.001	4,726	-0.009	0.987	5,788
High School Grad Status (Age 19)	0.886	0.318	3,730	0.888	0.315	5,788
GPA	2.964	0.784	4,318	2.985	0.792	5,788
Criminal Conviction	0.166	0.372	3,727	0.157	0.364	5,788
Teen Parent	0.180	0.385	3,126	0.177	0.381	5,788
Lives Independently (Age 19)	0.230	0.421	3,550	0.215	0.411	5,788
Math Achievement ($\approx$ Age 14)	0.007	0.996	4,140	0.015	0.969	5,788
Reading Achievement ($\approx$ Age 14)	0.009	0.993	4,143	0.008	0.964	5,788
Family Income ($\approx$ Age 14)	62718.443	43660.289	4,764	63243.321	44494.305	5,788
Marijuana Use	0.341	0.474	4,063	0.336	0.473	5,788
Other Drug Use	0.069	0.253	3,774	0.065	0.246	5,788
Early Sex	0.219	0.413	3,489	0.187	0.390	5,788
Suspension	0.129	0.335	5,086	0.134	0.340	5,788
Fighting	0.080	0.272	3,743	0.072	0.259	5,788
Hitting	0.203	0.403	3,745	0.188	0.391	5,788
Damaging Property	0.095	0.293	1,362	0.082	0.275	5,788
Self-Esteem Index	-0.001	1.004	3,661	-0.004	1.005	5,788
Religious Service Attendance	3.041	1.688	3,749	3.036	1.410	5,788
Gender Role Attitudes	2.063	0.496	3,382	2.056	0.488	5,788
School Clubs	0.707	0.455	2,369	0.668	0.471	5,788
Panel B: NLSY79						

Male	0.491	0.500	6,098	0.491	0.500	6,098
Black	0.118	0.323	6,098	0.118	0.323	6,098
Hispanic	0.078	0.268	6,098	0.078	0.268	6,098
Other Race	0.053	0.224	6,098	0.053	0.224	6,098
Mom Age at Birth	25.816	6.376	5,286	25.799	6.398	6,098
Mom Age at First Birth	21.642	4.613	4,377	21.714	4.550	6,098
High School Grad Status (Age 19)	0.764	0.425	5,942	0.761	0.426	6,098
GPA	2.630	0.881	4,064	2.454	0.929	6,098
Criminal Conviction	0.099	0.299	3,022	0.108	0.310	6,098
Teen Parent	0.144	0.351	6,096	0.143	0.351	6,098
Lives Independently (Age 19)	0.423	0.494	5,217	0.421	0.494	6,098
Math Achievement ($\approx$ Age 19)	0.000	0.999	5,754	-0.004	0.994	6,098
Reading Achievement ($\approx$ Age 19)	0.000	0.999	5,754	0.000	0.997	6,098
Family Income ($\approx$ Age 19)	59755.565	36149.078	5,008	59485.324	36070.492	6,098
Marijuana Use	0.504	0.500	4,382	0.503	0.500	6,098
Other Drug Use	0.225	0.418	4,375	0.225	0.418	6,098
Early Sex	0.105	0.306	5,703	0.104	0.305	6,098
Suspension	0.210	0.407	5,860	0.210	0.407	6,098
Fighting	0.230	0.421	4,407	0.237	0.425	6,098
Hitting	0.356	0.479	4,411	0.359	0.480	6,098
Damaging Property	0.175	0.380	4,380	0.176	0.381	6,098
Self-Esteem Index	0.000	0.999	3,918	0.023	1.012	6,098
Religious Service Attendance	3.020	1.674	5,986	3.020	1.660	6,098
Gender Role Attitudes	1.879	0.553	6,006	1.878	0.554	6,098
School Clubs	0.652	0.476	5,713	0.646	0.478	6,098
Family Income ($\approx$ Age 29)	54453.909	35224.495	5,654	54580.951	35744.767	6,098
College Completion (Age 29)	0.230	0.421	5,886	0.233	0.423	6,098
Lives Independently ($\approx$ Age 29)	0.885	0.319	5,851	0.884	0.321	6,098
Family Income ($\approx$ Age 40)	69373.287	44033.432	4,968	66322.348	42926.644	6,098

Notes: This table reports the means, standard deviations, and number of observations before and after imputation for each outcome used in the life-cycle simulation.

2.2.3 Running the Simulation

ESTIMATING COEFFICIENTS

Using the two imputed datasets and the equations above, we are able to estimate the coefficients for each outcome in a life-stage. However, an issue arises in linking the life-stages across these two data sources. Due to the age of respondents at first interview in the NLSY79, the data from earlier life stages is not as rich as in the CNLSY. Therefore, the NLSY79 does not contain all of the CAB, EC, MC, and AD variables that the CNLSY has. In order to overcome this, we define a set of linking variables, LINK, that contains all outcomes that are available in both the NLSY79 and the CNLSY.[15] We can then estimate the following two equations in the NLSY79 dataset to obtain coefficients for each TTA and AH outcome:

$$TTA\ Outcome_i = \beta_0^{tta} + \beta_{link}^{tta} LINK_i + \varepsilon_i^{tta}$$
$$AH\ Outcome_i = \beta_0^{ah} + \beta_{link}^{ah} LINK_i + \beta_{tta}^{ah} TTA_i + \varepsilon_i^{ah}.$$

GENERATING THE BASELINE

Once we have estimated the model with the modification necessary to link the two datasets together, we can then use the coefficients to generate a synthetic baseline by predicting the values of all variables included in our simulation for the CNLSY respondents. Starting with the set of CAB variables in the CNLSY dataset, we predict the set of EC variables using the coefficients from our estimations. Using the newly predicted variables, we then predict the set of MC variables and afterwards predict the set of AD variables. Using the CAB and predicted EC, MC, and AD variables, we then use the coefficients from the linking equations to predict the TTA and then the AH variables.

For the continuous variables in the EC, MC, and AD, we add in the residuals from the regression that estimated the partial correlations for the given variable. This leaves the predicted continuous variables in these life-stages identical to their values in the imputed datasets. As described in Winship and Owen (2013), this attempts to capture realized unobservables. For all binary and categorical variables, we again round them to the nearest integer.

PROPAGATING THE EFFECT

Given the impact of an intervention at some life-stage, we can again use the estimated coefficients from all life-stages to propagate the effects of the intervention through to adult outcomes. For example, if we have an intervention that increases the reading scores of all students at age ten by 0.5σ, we can simulate the

[15] See Table 3 for a list of these variables.

long term impacts of this intervention by increasing age ten reading scores of all students in the sample by 0.5σ and then predicting post intervention values of the AD variables using the partial correlations we found in the imputed CNLSY dataset. We then use the set of predicted LINK variables and partial correlations found in the NLSY79 to predict the post intervention TTA variables./footnoteNote that the LINK variables that occurred before age ten will be unchanged as they occurred before the intervention. Using these predicted TTA values and the same set of predicted LINK variables, we again use partial correlations from the NLSY79 to predict post intervention values of family income at age 40.

We can use similar procedures to simulate the long-term effect of multiple impacts across different life-stages or impacts that target certain subsamples. Further, we can simulate the long-term effects of interventions that target any measures that are included in our model.

Similar to when we generated the synthetic baseline, for all variables included in the CNLSY, we again add in the residuals from the regressions that estimated the partial correlations.

Calculating the Impact

After running the procedure described above for a given intervention impact (or multiple impacts at once), we then have a dataset of baseline values and post intervention values of all variables in our model. Comparing a post-intervention estimation of an outcome to the baseline estimation would then provide us with an estimated impact of the intervention on the given outcome. For example, we could compare post intervention family income at age 40 to baseline family income at age 40 to estimate the impact on age 40 family income. Further, we could also investigate the impact an intervention would have on any outcome included in our data (e.g. high school graduation rates, adolescence test scores, drug use, teenage pregnancies, family income at age 29).