Volume Title: Targeting Investments in Children: Fighting Poverty When Resources are Limited

Volume Author/Editor: Phillip B. Levine and David J. Zimmerman, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-47581-6
ISBN13: 978-0-226-47581-3

Volume URL: http://www.nber.org/books/levi09-1

Conference Date: September 26, 2008

Publication Date: August 2010

Chapter Title: Synthesizing the Results

Chapter Authors: Phillip B. Levine, David J. Zimmerman

Chapter URL: http://www.nber.org/chapters/c11733

Chapter pages in book: (353 - 376)

13

Synthesizing the Results

Phillip B. Levine and David J. Zimmerman

13.1 Introduction

The preceding twelve chapters reviewed the evidence regarding the effectiveness of programs that could alter the subsequent poverty status of children who are poor today. These programs ranged from those that affect very young children to those who are bordering on adulthood. Each review highlighted the impact of these interventions on children's outcomes, emphasizing those that may alter economic well-being in adulthood. In some cases, the reviews indicated that the programs were not found to significantly alter children's outcomes. In others, they did seem to have an effect, but perhaps not in ways that would affect children's earnings later in life. Still others did have some promising findings, suggesting that there are ways to intervene in a child's life and potentially make a difference in reducing poverty.

Our initial goal in this chapter is to summarize all of this evidence. For those programs that are not found to alter children's outcomes in any way, any further discussion would require an investigation into the content of the program, the quality of the methodology employed in its review, or other topics that are not the focus of our analysis. Similarly, it is beyond the scope of this exercise to further consider those programs that appear to have an impact on children's outcomes in ways that may not alter their subsequent poverty status. These programs may be well-positioned to satisfy other social goals and worthy of further study or even broader implementation, but that is not the exercise we are seeking to conduct.

Phillip B. Levine is the Class of 1919 Professor and chair of department of economics at Wellesley College, and a research associate of the National Bureau of Economic Research. David J. Zimmerman is a professor of economics and Orrin Sage Professor of Political Economy at Williams College, and a research associate of the National Bureau of Economic Research.

Our goal is to evaluate the poverty-reducing impact of these interventions and it is in this dimension that we pursue additional analysis of those programs that would appear to help accomplish that task. Among those programs, the question then becomes which of these policies works the best. Since different interventions target different outcomes, a direct comparison of their benefits is difficult. The ultimate goal of this chapter is to undertake this task.

13.2 Summary of Program Effectiveness

The lessons to be learned from the dozens of programs described in the preceding twelve chapters of this volume may be difficult to decipher by the sheer volume of information presented. In this section, we will boldly attempt to summarize all of this evidence. The attempt is bold because there are many nuances involved in different specific interventions that may make some effective and others not. Labeling entire categories as successful or otherwise in some dimension requires overlooking those nuances. In each case, we use our best judgment in providing our interpretation of the big picture, but it is important to recognize that our black and white interpretations may overlook some of the relevant grays.

Tables 13.1 through 13.3 provide our summary of the effectiveness of the types of interventions reviewed in this volume. The three tables are distinguished by the child's developmental level at the time of the intervention—early childhood, middle childhood, and adolescence. The tables indicate the types of interventions attempted at each developmental stage and then summarize the outcomes studied and their overall impact. In many, if not all, of these cases, the results of several individual studies are being aggregated in our description of the impact. Any specific estimates are ballpark midpoint estimates across studies.

The summaries we provide suggest that interventions can be categorized into three groups: (a) those that do not seem to have an any impact on children's outcomes; (b) those that seem to have an impact on children's outcomes, but not in any way that may lead to subsequent poverty reduction; and (c) those that may reduce the likelihood that the child is poor later in life. The programs that fit into each category were listed earlier in the introduction to this volume. For our purposes, an outcome that can alter subsequent poverty status is one that has an impact on some dimension of human capital. Outcomes like improving test scores and increasing the likelihood of graduating from high school are measures of human capital and they are well known to improve subsequent earnings. There are some limitations of distinguishing outcomes in this way and we will discuss those in more detail later, but for now this is how we will make this distinction.

When we categorize interventions in this way, we find a variety of interventions that show promise and many that do not. Investments targeted at

high-quality center-based preschool programs have been shown to raise both short- and long-term cognitive test scores as well as to improve adult labor market outcomes. These programs are reviewed in chapter 2 by Duncan, Ludwig, and Magnuson. Model preschool programs such as the Abecedarian Program and the Perry Preschool Project, for example, show significant impacts on participating children's IQ scores. Gains of around a standard deviation are apparent for children during their participation in the program. These gains are cut roughly in half by age twelve and by another quarter by age twenty-one. Still, the Abecedarian Program reports gains of .38 standard deviations for these young adults. In addition, participants in the Perry Preschool Program were 17 percentage points more likely to graduate from high school than their counterparts in the control group. These effects were generated in an experimental setting and, while based on small samples, give credible evidence of the potential effectiveness of high quality preschool care. These programs are, however, expensive. Total per-student costs for the Abecedarian Program reached $85,000.[1] The Perry Preschool Program devoted $15,705 per student. Not surprisingly, the less expensive and less intensive Head Start and state universal pre-k programs showed smaller effects or found improvements in outcomes that are only distantly connected to labor market effects.

Given the promise of these results, it might be expected that investments in high-quality day care or after-school programs would offer similarly impressive results. That, however, was not the case, as reported by Anderson in chapter 3 and Levine and Zimmerman in chapter 5. While high-quality day care may offer some benefits, the statistical evidence is not yet compelling. Similarly, the best evidence on the efficacy of after-school programs shows no reliable evidence on improvements in test scores or graduation rates.

Investments targeted at the parents of young children yielded mixed results, also reported in chapter 2 by Duncan, Ludwig, and Magnuson. Prenatal programs such as the Elmira Nurse Family Partnership show some evidence of gains in IQ scores, but these gains are not statistically significant at the long term follow-up when the children are four years old. Children in the "at-risk" sample were, however, less likely to have been arrested or subject to alcohol or drug impairment at age fifteen. Programs targeted at improving parenting techniques similarly show gains in children's problem behavior, but do not register improvements easily linked to the child's human capital.

While programs targeted at improving parental care have shown limited success in raising a child's productive attributes, mentoring programs such as those provided by the Big Brothers Big Sisters Program have shown positive effects on participants' academic performance (Levine and Zimmerman,

1. All dollar amounts reported in this chapter refer to 2008 dollars. The value of the Consumer Price Index was 219.1 in August of 2008 (the most recent available at the time this chapter was written) on a base of 100 in 1982–1984.

Table 13.1 **Summary of early childhood intervention effects**

Intervention	Intervention/outcomes	Effect
	Child development	
Parental Education Programs	Programs seek to boost parent's general knowledge about parenting and child development. Outcomes: IQ, health, crime, child abuse, emergency room visits, alcohol and drug use.	Generally modest and statistically insignificant effects on longer-term measures of child development.
Parent Management Programs	Programs designed to teach strategies to improve children's problem behavior. Outcomes: Negative parenting, parenting stress, problem behavior.	Reduced problem behavior.
Early Education (Perry Pre-School, Abecedarian, Head Start)	Center-based early childhood intervention programs. Outcomes: IQ, test scores, behavior problems, crime, teen parent, drug use, high school graduation, college entry.	Model programs show improvements at age 3 or 4 in IQ (+ 1 sd), reading (+ .7 sd), math (+ .7 sd). IQ gains decline with age (.5 sd at age 12 and .38 sd at age 21). Also improvements in high school graduation rates in Perry (+ 17 pp). Earnings higher by $5,500 in Perry Pre-School ($2000). State programs show average gains of .14 sd for PPVT and .29 sd for math effects at age 4. Most effects smaller and often ns in Head Start at age 4. Head Start shows high school graduation + 21 pp at age 23.
	Child care	
Day care provision	Provision of child care for preschool aged children. Outcomes: Behavioral and cognitive outcomes of child, family incomes impact on child.	High-quality care may be beneficial, though no good experimental evidence. No causal link in outcomes with family income.

	Child health	
Medicaid/SCHIP Expansion	Provision of health insurance. Outcomes: Probability child uninsured, probability of child visits doctor, birth weight.	Insurance increases prenatal care but no measured effect on child health. Birth weight is linked to education, but not to interventions.
WIC	Provision of food and nutrition education for mothers. Outcomes: birth weight, gestational length.	Little evidence linking WIC to birth weight.
ADHD Drug Treatment	Treatment for Attention Deficit Hyperactivity Disorder. Outcomes: Test scores, grade retention, special education placement.	ADHD linked to GPA and test scores, but no credible evidence on long-term or short-term effects from treatment.
Asthma treatment	Treatment for asthma. Outcomes: doctor visits, school days missed, academic performance.	Little evidence linking asthma to academic achievement or income.
Dental health	Provision of fluoridation, dental sealants, or dental insurance. Outcomes: tooth loss, income.	Little evidence linking dental health to income.
Obesity	Obesity prevention programs. Outcomes: Family Income, probability of marriage, wages.	Obesity linked to outcomes, but no effective intervention.
Environmental toxins	Improvements in air quality and lead exposure. Outcomes: Absenteeism, asthma, pneumonia, infant mortality.	Links established, but interventions not evaluated.

Note: SCHIP = State Children's Health Insurance Program; WIC = Women, Infants, and Children.

Table 13.2 **Summary of middle childhood intervention effects**

Intervention	Intervention/outcomes	Effect
	After-school programs	
After-school care	Provision of after-school programs for school-aged children. Outcomes: Self-care. Maternal employment, reading test-score (SAT-9), English grade, math grade, science and social science grades, TV viewing time, homework completion, attendance.	No statistically significant effect.
QOP	Intensive intervention in high-risk adolescents. Outcomes: High-school graduation rates, attend postsecondary institution, achievement test scores, behavioral issues.	No statistically significant effect.
Mentoring	Mentoring programs. Outcomes: drug abuse, crime, academic performance.	+ .08 grade points (Big Brother Big Sister) Smaller or insignificant effects from School-Based Programs. Basal word recognition and spelling show statistically significant gains (Howard Street).
	Education reforms	
Direct dollar investments	Policies changing per-pupil expenditures in schools. Outcomes: test scores.	Mixed. Mostly no effect.

Class size reduction	Policies reducing school class size. Outcomes: Third- and eighth-grade test scores, taking college entrance exam.	+ .15 sd on composite test. + .02 probability took college entrance test.
Curriculum and instructional programs	Curricular innovations. Outcomes: Test scores.	Limited evidence. "Success for All" program increased reading test scores by .21–.36 sd.
Computer aided instruction	Integrating computers into schools' instruction programs. Outcomes: Math and reading test scores.	Mixed results. Some programs have very large effects and others find nothing statistically significant.
Teacher wages, recruitment, professional development	Policies aimed at improving teacher quality. High school graduation, college enrollment, reading, math, and science tests.	10% increase in teachers wages yields 3–4% fall in dropout rates and 1–2% increase in college enrollment. "Teach for America": + .07–.15 sd in math versus not traditionally trained teachers.
Vouchers and charter schools	Vouchers that can be used for tuition at a private school. Public funding of charter schools. Outcomes: Test scores.	Insignificant or negative effects.

Note: QOP = Quantum Opportunities Program.

Table 13.3 **Summary of adolescent intervention effects**

Intervention	Intervention/outcomes	Effect
	Substance abuse prevention	
School-based	School-based training providing information, skill building, and normative education. Outcomes: Use of drugs, alcohol, and tobacco, wages, employment, educational outcomes.	Limited evidence on the effects of drug, alcohol, and tobacco use.
Family-based	Programs aimed at families where a parent or child is at high risk. Outcomes: Time to initiation of drug, alcohol, and tobacco use.	Mixed results.
Community-based	Programs include targeted mentoring programs, drug treatment, and criminal justice interventions. Outcomes: Time to initiation of drug, alcohol, and tobacco use, educational outcomes.	Improved educational outcomes (see after-school care), delayed initiation of drug and alcohol use.
Drug and alcohol testing	Drug testing by schools, employers, or criminal justice agencies. Outcomes: drug and alcohol use.	Mixed results.
	Pregnancy prevention	
Sex education (abstinence focus)	Education programs emphasizing abstinence. Outcomes: sexual initiation, frequency of sex, unprotected sex, pregnancy.	No observable impact.
Sex education (contraception focus)	Education programs emphasizing contraception. Outcomes: sexual initiation, frequency of sex, unprotected sex, pregnancy.	Mixed. Some evidence of delayed sexual initiation, increased condom use.
Multicomponent approaches	Comprehensive approach to teen pregnancy prevention. Outcomes: sexual initiation, frequency of sex, unprotected sex, pregnancy, academic and social behaviors.	Reduced pregnancy rates, increased condom use (Carrera Program).

	Dropout prevention and college preparatory programs	
Gear up	Comprehensive federal program tasked at equalizing access to higher education for low-income students. Outcomes: college plans, math and reading scores.	Positive effect on college plans. Mixed results on math skills. No effect on reading scores.
Upward bound	Comprehensive federal program offering a variety of services including instruction, tutoring, and counseling. Outcomes: various measures of high school academic preparation.	Increase in high school math credits (+ .2 credits). No effect on other measures of high school academic preparation. Increased likelihood of four-year college enrollment (+ 6 pp).
Talent search	Federal program providing information to low-income students on high school curricular choices and on financial aid and the college application process. Outcomes: college attendance, high school graduation, apply for financial aid.	Participants more likely (+ 12 pp) to enroll in public postsecondary institution.
Project grad	Implements math and reading curricula along with professional development for teachers. Outcomes: complete core curriculum, promoted to 10th grade.	Mixed or insignificant effects.
AVID	Attempts to enroll capable students in 5th–12th grades in rigorous curriculum. Outcomes: high school GPA, high school credit accumulation.	Higher GPA in middle school, effects not sustained in high school. Small positive effect on core credits accumulation.
Puente	Program targeted at Latino students, includes counseling designed to increase college enrollment. Outcomes: complete SAT, attend two- or four-year college, remain in college 3 + years.	Increased college attendance and persistence, through self-reports and small samples.
	College aid	
College financial aid	$1,000 increase in college financial aid. Outcomes: college attendance.	4 pp increase in college attendance.

(*continued*)

Table 13.3 (continued)

Intervention	Intervention/outcomes	Effect
STAR project	Incoming students received either peer advising and organized study groups or financial incentives based on academic performance, both of these, or neither of these. Outcomes: Grades, retention.	Modest effects. GPA increased (+ .1–.2 sigma) for first-year grades. Effect for females only with the strongest effects from services combined with incentives.
Opening doors	Entering students take blocks of classes together, receive extra tutoring, supplementary financial aid, and enhanced student services. Outcomes: credits taken and earned, pass rates, GPA, retention	Increased retention (7–8 pp), Increased pass rate (10–12 pp), more credits taken (2–3).
	Neighborhood and community initiatives	
Housing vouchers	Low-income families in public housing receive a voucher to move elsewhere. Outcomes: dropping out of high school, reading and math test scores, grades, college attendance.	Mixed results. Increases in reading and math scores in Chicago (+ .2 sd) and Baltimore. Statistically insignificant effects in Boston, Los Angeles, and New York.
	Vocational training	
Job Corps	Intensive and expensive program targeting economically challenged youth with counseling, education, training, health care, and job placement services. Residential training centers used. Outcomes: employment, earnings.	Increases in both male and female earnings. Impact largest for 22- to 24-year-olds. No significant impact for youths.
Career Academy	Provides basic skills and vocational training within high schools. Substitutes vocational courses for academic courses. Outcomes: earnings.	Increased earnings for males (+ 20%). Decreased earnings for women (–5%). Population treated less disadvantaged than Job Corps participants.

chapter 5). Participants in this program register grades that are .08 grade points higher than nonparticipants. Costs in this program are modest at about $1,500.

The evidence on investments in children's health is problematic, as reviewed by Shore-Sheppard in chapter 4. While there is good evidence that policy can be initiated that can expand health care insurance opportunities, the evidence is weaker that this improves child health. Further, the links between child health and adult labor market outcomes are not well-established. For example, effective interventions targeted at dental health, asthma, and exposure to environmental toxins have been developed, but the evidence linking these improvements to adult poverty is much weaker.

Loeb and McEwan (chapter 6) find that a variety of education interventions hold promise. While direct dollar transfers to schools show mixed results, there is evidence that reductions in class size can raise test scores. Indeed, evidence from Project STAR showed that investing in smaller class sizes from kindergarten through the third grade resulted in a .15 standard deviation gain in a composite test score. It is important to note, however, that class size reductions are expensive; Project STAR costs reached $12,145 per student. Participants in the "Success for All" reading program scored between .21 and .36 standard deviations higher on a reading test; the program cost $2,789 per student.

Other educational interventions have been found to be successful. Paying teachers higher wages has been found to improve educational performance. One convincing study concludes that increasing teacher pay by 10 percent reduced dropout rates by 3 to 4 percent and increased college enrollment by 1 to 2 percent. The "Teach for America" teacher training program has been found to be successful when the newly placed teachers are replacing teachers of average quality. Teacher training costs of $1,374 have been found to generate increases in math test scores of .07 to .15 standard deviations in these instances.

Studies of other types of educational interventions have yielded mixed results, at best. The evidence does not tend to support the effectiveness of computer-aided instruction programs, professional development programs, and investments in private school vouchers or charter schools.

Shifting to investments targeted mostly at adolescents, chapter 7 by Kilmer and Pacula tackles the topic of adolescent substance abuse. The evidence on these interventions suggests that some programs may be effective in reducing adolescents' consumption of alcohol, drugs, or cigarettes. The evidence is less compelling that these reductions in use persist through high school. It is still the case, however, that short-term program effects could generate significant changes in lifetime consumption—perhaps, for example, by altering a child's peer environment. Further, programs that delay the initiation of substance use can reduce the odds of later dependency. These findings are, however, matched with less compelling evidence on the link between

substance use and labor market outcomes. While chronic substance use may impact economic outcomes, modest levels of use may have no effect. As the authors note: "the level of consumption, duration of consumption, and timing of consumption all have important implications in terms of whether we should expect to see an impact on poverty, employment, or earnings." The evidence, as it stands, however, does not suggest a clear and effective policy prescription.

Kearney (chapter 8) reviews programs targeted at reducing teen pregnancy and concludes that these programs are not likely to reduce adult poverty. It is true that several interventions in this area have demonstrated the possibility of delaying the age at which adolescents engage in sexual activity. Some programs, such as the Carrera Program, have successfully increased participants' condom use and reduced rates of teen pregnancy. Unfortunately, however, the literature linking teen pregnancy to adult economic outcomes suggests that teen pregnancy, per se, plays a limited role in causing adult poverty. Thus, from a purely antipoverty perspective, investments in pregnancy prevention are unlikely to deliver a high return.

Educational interventions have also been tried that target students at older ages and are designed to keep them from dropping out of high school or to encourage them to attend college. Long's discussion in chapter 9 concludes that research on dropout prevention programs has not been able to convincingly demonstrate success. This is largely attributable to poor research designs and inadequate data. The one intervention that does seem to have some success is to extend mandatory school laws to require attendance through older ages. A policy of this sort is different than the other types of interventions considered in this volume, which are more apt to provide additional resources rather than impose stricter rules.

College aid programs appear to be more successful in encouraging students to continue with their schooling. Deming and Dynarski (chapter 10) report that providing aid without large administrative burdens is successful in increasing college attendance. They conclude that the best estimates indicate that offering a $1,000 aid award will increase college attendance by about 4 percentage points. They also report that evidence from the Opening Doors intervention raises the possibility that a combination of services and financial incentives also may increase retention among college students.

Programs designed to change the environment in which children (not necessarily just adolescents) live have also been studied, yielding mixed results. Burdick-Will and Ludwig review these programs in chapter 11. These programs attack the claim that a child's neighborhood can affect his or her outcomes through role models, access to resources, and other factors; moving children outside this environment, the argument goes, would improve outcomes in many dimensions, including academic outcomes, that would lead to greater human capital development and higher subsequent wages.

The studies, however, are not overwhelming in their support of the success of these policies. Some bright spots emerge from the literature. Moving certain population subgroups in some studies have been found to successfully improve children's outcomes. Some of these studies do not perfectly address the selection problem regarding who moves and findings of broad-based improvements in children's outcomes are rare.

Job training programs, broadly speaking, also have not been found to have a meaningful impact on the earnings capacity of disadvantaged youths. As Lalonde and Sullivan review in chapter 12, most programs implemented in the past that have been rigorously evaluated are not found to lead to significant earnings gains among program participants. They argue that a likely explanation for this is that the level of intensity of these programs makes success difficult. If a nine-month school year can increase wages by 10 percent, how effective could a training program be that meets for a few hundred hours? Consistent with this point, the exceptions to their findings involve Career Academies and the Job Corps programs. Career Academies provide both basic skills and vocational training within high schools. The Job Corps is a very intensive residential training program that provides a very comprehensive set of services, including counseling, training, work experience, health care, and job placement. The benefits of this program are large, but costs are as well. Career Academies seems to provide greater benefits relative to its costs. It does not target as poor a population as does Job Corps.

13.3 Effectiveness in Reducing Poverty: Methods

The preceding section shows that there are a number of areas in which policy interventions appear to alter children's outcomes in ways that can improve their longer run economic well-being. The question then becomes which types of interventions can best accomplish this task. This is a difficult question because we are rarely able to observe, say, thirty-year program impacts, which would make such an analysis rather straightforward. Even if we could, it is not obvious how to interpret evidence from interventions that took place thirty years ago.

Instead, we are typically only able to observe outcomes that occur at the time of the intervention or perhaps within a few years of its completion. This limitation means that any attempt to estimate the longer term effectiveness of program interventions will require us to simulate the subsequent earnings impact. Once we conduct that simulation, then we can compare that impact to the cost of the intervention to determine which types of programs deliver more "bang for the buck." This section will describe our methodology for conducting this analysis and provide some caveats that are important to consider in interpreting these results. The results themselves will be presented in the subsequent section.

13.3.1 Overview of Our Approach

To estimate the poverty-reducing impact of an intervention, we employ a two-stage empirical approach comparable to that used by Krueger (2003) in his analysis of class size reductions. In the first stage, we find "best estimates" of the impact that spending on a particular type of program has on contemporaneous outcomes. These outcomes, which include things like high school grades and math and reading test scores, represent indicators of the child's level of human capital. This stage was accomplished in the preceding section. The second stage relates these outcomes to subsequent earnings as an adult, a task that will be completed in this section. Combining these results enables us to estimate the impact of the intervention on adult earnings.[2]

Consider, for example, a hypothetical intervention that has been shown to be effective in improving the test scores of program participants by .2 standard deviations. Suppose we augment that finding with additional evidence that for every standard deviation improvement in test scores, students' subsequent earnings rise by 10 percent. One could combine these estimates to indicate that the program increases wages by 2 percent. Combining this estimate with an estimate of baseline lifetime earnings provides a means to determine the program's impact. If a typical program participant would earn $500,000 over his or her worklife (in present discounted value terms), then a 2 percent program impact means that participant's lifetime earnings would be estimated to increase by $10,000. This is the value that we would compare to the cost of the program.

One complication in all of this is the timing of when the program effect occurs. Interventions targeted at young children, for instance, may increase test scores in the near term, but those effects may fade over time. On the one hand, if it is the direct skill linked to the test score that matters, then if the effect fades over time, it should have a limited impact on wages. An alternative perspective is that the test score is just a marker for a wide variety of impacts that the test may measure. Even if the specific skills measured by the test fade over time, the other attributes of the individual that have been altered by the program may linger. It is in this sense that one may prefer to use short-term test score impacts to capture the total impact of the pro-

2. More formally, we incorporate evidence from equations of the form:

$$\text{Adult Outcomes}_i = \beta_0 + \beta_1 \text{HC}_i + \beta_2 X_i + \varepsilon_i$$

$$HC_i = \gamma_0 + \gamma_1 \text{INV}_i + \gamma_2 Z_i + e_i,$$

where adult outcomes are things like earnings or poverty status, HC represents some measure of a child's human capital (e.g., test scores), INV represents investments made in the child (i.e., program participation), X and Z represent other factors that affect adult outcomes and level of human capital, respectively, and *i* indexes individuals. In this model, β_1 represents the impact of human capital on adult outcomes, and γ_1 represents the impact of the investment on human capital, so that the product of β_1 and γ_1 indicates the impact of the investment on the adult outcome.

gram. In our case, when long-term effects on final educational attainment or direct observations on wages are available, we will use those outcomes as the program impact. Where they are not, we will use the short-term test score impact. This decision is one that we will evaluate subsequently in the one instance (Perry Pre-School) where we have both short-term impacts on test scores and long-term impacts on wages.

The tasks that remain to be completed include evaluating the percentage impact on earnings and estimating baseline lifetime earnings. These steps will be accomplished in the following subsections.

13.3.2 Review of the Literature

Past research has examined the impact of academic outcomes on subsequent earnings; we will heavily rely on that research. In his survey of the evidence, Krueger (2003) concluded that a suitable summary statistic is that a 1 standard deviation increase in math or reading test scores is associated with an earnings gain of around 8 percent. Hanusheck (2006) conducted a similar survey and concluded that a 1 standard deviation increase in test scores is associated with an earnings gain of 12 percent.

Yet none of the surveyed studies informs precisely the question that we have in mind. Past studies tend to focus on test scores obtained in high school; the academic outcomes that are available to us from the interventions previously reviewed tend to be measured at younger ages. Longitudinal studies of sufficient duration to capture both early/mid-childhood academic outcomes and adult earnings are hard to come by. The British National Child Development Survey, used by Currie and Thomas (2001), gets around this problem because it includes math test scores at ages seven and sixteen and wages at age thirty-three. It would be our preference, however, to use American data that would more accurately reflect the labor market realities relevant for the children participating in the programs we consider. The data sets that have been employed are also a bit dated, focusing on educational measures from the early 1980s and examining outcomes just a handful of years after that. We will explore this relationship further in the subsequent section.[3]

We will also rely on prior research to assign a benefit to programs that are shown to increase educational attainment. The impact of educational attainment on wages is an issue that has been extensively studied. Despite the important difficulties in estimating a causal impact, researchers have introduced methods that are sufficiently convincing that a consensus has emerged. It has become commonplace to attribute a 10 percent return to every additional year of education a child receives.

3. Another potential problem with these analyses is that the wages measured tend to be from relatively young adults and the impact of greater ability may rise over time. If so, then the estimates from past studies as well as our own investigation reported later will understate the true program impact.

13.3.3 Analysis of NLSY Data

Before proceeding, we report an analysis that we have conducted that updates previous research on the relationship between test scores on subsequent wages. We do so by using data from the children of respondents to the 1979 National Longitudinal Survey of Youth (NLSY79) as well as the newer NLSY cohort, which began in 1997 (NLSY97). The NLSY79 represents a sample of over 12,000 individuals born between 1957 and 1964 who have been surveyed annually through 1994 and biennially since then. Starting in 1986, the children of the female respondents to this survey have also been tracked biennially. To date, over 11,000 children have been born to these women. Children fifteen and over participate in a "young adult" survey. We concentrate on those survey participants who were born in 1987 or earlier, making them age nineteen or older in 2006, the last year for which data are available. There are almost 5,000 individuals that satisfy this condition. We focus on those aged nineteen or over because they have passed the regular point of high school completion, so that many of them will have fully entered the labor market. Most survey respondents in our sample are between the ages of nineteen and twenty-nine.

In addition to this measure of educational attainment, these data also provide several developmental measures, including tests of cognitive ability, which were obtained for each child beginning from about age five and through age fourteen.[4] In our analysis, we take advantage of scores on the Peabody Individual Achievement Test (PIAT), which measures a child's performance in math, reading recognition, and reading comprehension. We have also aggregated across the three components of the PIAT, estimating average scores, and we will use this aggregate measure in our analysis as well. These data are available for about 4,700 of the children born in our sample of those born in 1987 or earlier.

Along with data on test scores, a wealth of other information is available for NLSY respondents and their children. In particular, wage data is obviously necessary for this analysis and it is available in these data. About 1,600 of the remaining sample are still available in the survey, have entered the labor force, and reported earnings in 2006. We also control for characteristics of the individual and his/her mother that may be related to both an individual's wage and his/her test scores. Mother's characteristics include: her age at the time of birth of the child, her completed years of schooling, her score on a test of aptitude/achievement (the Armed Forces Qualifying Test, or AFQT), the number of children she had up until the year 2000, the fraction of the child's life that the mother was married, and her average annual

4. We have also investigated whether scores on tests taken between the ages of five and nine have a different impact than scores on tests taken between the ages of ten and fourteen. The results were similar, so we chose to proceed combining the ages at which the tests were taken.

family income in the year since the child was born. We also control for a number of characteristics of the child as well, including: the child's birth order; a variable indicating whether or not the child was the first born to the mother; and whether the child was female, nonwhite, or Hispanic. Missing data on these other explanatory variables reduces the sample to about 1,400.

We use these data to estimate regression models where the dependent variable is the natural log of the wage. The results of our analysis are reported in the top panel of table 13.4. Each row presents the results of a separate regression model that differs depending upon the specific test score included as the key explanatory variable. Here we see that a 1 standard deviation increase in a number of different types of test scores increase adult wages on the order of about 6 to 10 percent. The results are reasonably robust to the specific type of test and to the fact that these tests were given to younger children than those examined in earlier studies.

We further examined this assertion by using a more recent cohort of data, the NLSY97. There are about 9,000 respondents in the NLSY97, all of whom were between the ages of twelve and sixteen as of December 31, 1996. Respondents are surveyed biennially; the most recent data available is from 2006, when they are twenty-two to twenty-six. Wage data for those in the labor market are available in 2006, along with data on an AFQT-type test like that administered in the NLSY79.[5] The NLSY97 is a data set similarly rich like the NLSY79, including information on age, gender, race, ethnicity, household structure and mother's and father's educational background. We include all of these variables as controls and estimate comparable models to those reported earlier using the children of the NLSY79.

The results of this exercise are reported in the middle panel of table 13.4. The results are remarkably similar to those from the NLSY79. In these data, a 1 standard deviation increase in test scores increase wages by 8.9 percent.

Based on this evidence and the simplicity of maintaining a single estimate for the estimated impact of test score effects, we will use a value of 10 percent as the return on a 1 standard deviation increase in test scores in our calculations. We arrived at this conclusion by combining our own evidence with the reviews discussed previously by Krueger (2003) and Hanusheck (2006). They concluded that an estimate of 8 percent and 12 percent, respectively, are appropriate values. Our preferred value of 10 percent is the midpoint of those conclusions and it is generally consistent with our findings as well.

We also conducted one additional analysis using NLSY data to enable us to convert the impact on high school grades to subsequent wages. The data that we use for this exercise is the original NLSY79 cohort (the parents of

5. Notes from the NLSY97 indicate this AFQT-type variable is "similar to the AFQT score generated by the Department of Defense, (but it) reflects work done by NLSY program staff and is neither generated nor endorsed by (the Department of Defense)."

Table 13.4 **Estimated impact of cognitive test scores and high school grades on wages**

Human capital measure and data set	Coefficient/(standard deviation)
NLSY79: Young adults	
Impact of 1 sd increase in aggregate test score	9.23 (1.44)
Impact of 1 sd increase in math	9.35 (1.41)
Impact of 1 sd increase in reading recognition	6.42 (1.07)
Impact of 1 sd increase in reading comprehension	6.03 (1.37)
NLSY97	
Impact of a 1 sd increase in AFQT-type test score	8.91 (1.10)
NLSY79: Original sample	
Impact of 1 point increase in high school GPA	15.87 (1.52)

Notes: Standard deviation in parentheses. Each cell represents estimates from different regression models, which are multiplied by 100 so that they may be interpreted as a percentage effect. These estimates are obtained from regression models that also control for the following characteristics. In the NLSY79 young adult sample: characteristics of the mother (age at birth of the child, educational attainment, aptitude/achievement test score, number of children, the fraction of the child's life the mother was married, and the log of the mother's average family income since the child was born) and of the child (birth order; whether or not the child was firstborn; gender, race, and ethnicity). In the NLSY97: age, gender, race, ethnicity, household structure, and mother's and father's educational attainment. In the NLSY79 original sample: age, gender, race and ethnicity, household structure while growing up, and mother's and father's educational attainment. Standard errors are adjusted to correct for heteroskedasticity.

the respondents used in our earlier exercise). We use wage data available in 2006, when respondents are between the ages of forty-one and forty-seven.

The most important characteristic of these data for the purposes of the analysis conducted here is that usable high school transcript data is available for almost 9,000 NLSY respondents, respectively. Since the courses each respondent took come from their high school transcript and are not self-reported, these data should be quite reliable. Because of sample attrition between 1979 and 2006 and the fact that not all NLSY respondents were employed in 2006, the sample size available for this exercise is about 4,600 respondents.

We use these data to estimate regression models where the dependent variable is the natural log of the hourly wage that the individual received in 2006. The results of our analysis are reported in the bottom panel of table 13.4. Here we see that a one-point increase in a student's GPA is associated with a 15.87 percent increase in adult wages.

13.3.4 Net Present Value Calculations

The last component that we need to simulate the impact on lifetime earnings associated with an intervention is a net present value calculation. Consider a program that increases test scores by 1 standard deviation and we are willing to attribute an 8 percent wage effect associated with that program based on our earlier discussion. To what number do we apply that 8 per-

cent and how does that alter lifetime earnings? In our analysis, we follow the general approach that Krueger (2003) used in taking wage data from the March Current Population Survey, in our case aggregating data from the 2006 through 2008 surveys.[6] We estimate the age-earnings profile, and then calculate the net present value of lifetime earnings, assuming that an individual hitting the labor market today would face the same wage profile as the current age-earnings profile. We assume a discount rate of 3 percent, as Krueger does.

One important difference in our calculations is that we use data from high school graduates only. In Krueger's case, he was dealing with a broad-based class-size intervention that would affect all students, so using the age-earnings profile of the average worker would be appropriate. In our case, we are dealing with interventions that largely target a lower income population, so the average worker's age-earnings profile is likely to be too high a starting point. It is difficult to know what the right education level is to assume for this population in this exercise; we have chosen those who have graduated from high school and not gone on to college. Based on this age-earnings profile and a 3 percent discount rate, we calculate that a typical worker would earn $555,000 over the course of his life (in 2007 dollars).

13.3.5 Caveats

We want to be clear that we recognize the limitations of a simulation of the nature we just described. Without long-term follow-up studies to the program evaluations, we have no other method of estimating earnings impact thirty or more years after the intervention. The approach we are using enables us to simulate that impact, but clearly that approach is not perfect and we want to clarify at least some of the important oversights that we make when we implement it.

The first main problem that we face using our approach is that we overlook standard errors in our estimates. Every estimate that we use comes with a standard error and we ignore all of those. The underlying problem is even worse than that, however. One way to view sample size is to consider the number of observations in a particular evaluation and those sample sizes are often relatively small (100 observations?). Another way to view sample size is to consider the number of interventions studied. One can easily imagine that there is a distribution of impacts associated with any particular type of intervention. Things like the quality of the program administrator and differences in the populations being treated could easily alter the results of any program. In larger multisite evaluations, differences in estimated impacts across sites are commonplace. Yet our simulation process ignores this variability in projecting the earnings impact of each intervention. Without the

6. These data provide wage information for 2005 through 2007; we treat them as if the results are in 2006 dollars to adjust for inflation.

ability to implement formal methods of statistical inference, it is important to keep in mind the imprecision that affects all of our estimates. If two estimates are "close," we will do our best to resist the temptation to rank them.

The second main problem is that we ignore the role that noncognitive factors may play in improving children's life outcomes. By focusing exclusively on educational outcomes, we miss the many other important determinants of lifetime earnings. Earnings may just as easily be influenced by "soft skills," like the ability to get along with others, communicate effectively, and to act appropriately in a work environment, as they are influenced by one's cognitive ability. Our approach does not take any of those other factors into account.

It is possible, however, to at least provide a preliminary gauge regarding the role of those other factors based on the results of the long-term follow-ups available from the Perry Pre-School intervention. We can use the short-term impacts on cognitive ability from that intervention, simulate longer term outcomes, and then compare those simulated longer term outcomes with actual longer term outcomes available in the data. Results from Perry Pre-School indicate that the intervention increased students' IQ scores at age four by .87 standard deviations (see the discussion in chapter 2 by Duncan, Ludwig, and Magnuson). Based on the results of our analysis and those of previous studies, we assume that a 1 standard deviation increase in test scores would increase wages by 10 percent. This means that Perry Pre-School would increase wages by about 8.7 percent. From a base lifetime earnings of $555,000 (in 2007 dollars), this wage increase would amount to an additional $48,285 over the working life of each participant.

Duncan, Ludwig, and Magnuson (chapter 2) also report from their review of the evidence that the present discounted value of lifetime earnings of program participants is about $60,000 higher than that of control group members in 2007 dollars. This means that the estimate we obtain by simulating a lifetime increase on the basis of the returns to improved cognitive ability is about 80 percent of the observed lifetime wage increase on the basis of the experimental findings. Although there are implicit standard errors associated with both numbers that are difficult to incorporate into an analysis like this, it suggests that the improvement in cognitive ability accounts for a large share of the overall wage increase.[7]

7. We can also conduct a similar exercise, simulating lifetime earnings effects on the basis of the increase in educational attainment. Evidence from the Perry evaluation indicates that the program increased the likelihood of high school graduation by 17 percentage points. Although we do not have access to continuous educational attainment measures, we hypothesize that this means that 17 percent of the participants received two more years of education. Valuing a year of education at 10 percent, this means that lifetime earnings for program participants should have risen, on average, by 21 percent for 17 percent of the participants, for an expected value rise in earnings of 3.57 percent. Again, using a base level of lifetime earnings of $555,000, this amounts to a $19,813 increase in 2007 dollars. Again, this suggests that this standard human capital measure can explain a sizable portion of the observed wage increase ($60,000 in 2007 dollars), but a sizable portion still remains unexplained.

13.4 Effectiveness in Reducing Poverty: Results

The goal of our exercise is to identify the programs that provide the greatest earnings impact relative to the cost of the intervention.[8] To standardize across interventions, we convert all of our results into the earnings impact per $1,000 expenditure. Those that yield the biggest impact on that scale are the ones that should be supported more strongly. But there is another option that must be considered that is not presented in this table or earlier in the report. One possible intervention would be simply to give the $1,000 to the children either directly or in the form of a deposit in a savings account that cannot be used until adulthood. If a programmatic intervention cannot yield greater lifetime earnings than this alternative $1,000 gift, then it is not clear that the program should be supported. But any increase in earnings beyond the $1,000 that a program may generate can be thought of as a return on the investment in the program. In other words, if the $1,000 investment generates $1,200 in higher earnings for the participant, then one could view this as a 20 percent rate of return.

The results of this analysis are presented in table 13.5. This table displays only those programs that are found to have impacts on the types of human capital measures that are likely to be related to higher lifetime incomes, as reported in tables 13.1, 13.2, and 13.3. We report the impact of specific interventions or categories of interventions in those instances where we have summarized the impact of a number of interventions of the same type. The first column of the table reports the impact on human capital. These results are taken directly from tables 13.1, 13.2, and 13.3, which have been taken from the relevant chapters where the program is discussed. The second column converts the impact on human capital to the impact on lifetime earnings, implementing the methodology described earlier. Each human capital measure is converted into a percentage wage effect using the results of prior research and our own analysis and then converted into a lifetime earnings impact by multiplying by our estimate of lifetime net present value of earnings ($555,000 in 2007 dollars). Column (3) reports the cost of the intervention, also measured in 2007 dollars.

Our calculations suggest that almost all of the programs that are estimated to improve a child's human capital result in subsequent lifetime earnings gains that are greater than the costs of the program. One might argue that any of these interventions show enough promise to receive support from those seeking to invest in children in a way that will improve their economic well-being.

8. Another present discounted value calculation is required in comparing benefits and costs since there is a significant time lag between the time of the intervention and labor market entry. That time lag also differs for different types of interventions. A $1,000 intervention that takes place at age five is "more expensive" than a $1,000 intervention at age sixteen because more time elapses between the earlier intervention and labor market entry during which that money could have been invested elsewhere.

Table 13.5 **Effectiveness of childhood interventions on adult earnings (all 2007 dollars)**

Intervention	Impact on human capital (1)	Impact on lifetime earnings (2)	Cost (3)	Earnings impact per $1,000 cost (4)
Early childhood education				
Perry Pre-School	Earnings measured directly	$60,000	$15,700	$3,822
Abecedarian	1 year increase in average educational attainment	$55,500	$90,000	$611
Head Start	.21 pp HS grad	$24,476	$8,000	$3,060
State Programs	.29 sd on math test	$16,095	$6,100	$2,639
Mentoring programs				
Big Brother Big Sister	.08 grade points	$7,046	$1,480	$4,761
Education reforms				
Class size reduction (Project STAR)	.15 sd on test scores	$8,325	$12,145	$685
Curriculum and instruction (Success for All)	.29 sd on reading test	$16,095	$2,789	$5,771
Teacher wages (10% increase)	3.5% reduction in HS dropout and 1.5% increase in college enrollment	$5,775	$4,440	$1,307
Teacher training (Teach for America)	.11 sd increase on math test	$6,105	$1,374	$4,443
College aid				
$1,000 Reduction in Cost	4 pp increase in college attendance	$4,662	$1,000	$4,662
Vocational training				
Job Corps	$1,695 annual earnings increase for men	$40,355	$21,000	$1,922
Career Academies	$2,088 annual earnings increase	$49,712	$2,800	$17,754

Indeed, some of the relative returns are impressive. Evidence from a recent evaluation of Career Academies indicates that those programs provide a very large return on their investment. Eight years following the intervention, treatment group members had an annual earnings gain of $2,088 (Kemple and Wilner 2008). Extrapolating this estimate to a present discounted value of lifetime earnings gains (at a 3 percent discount rate) generates $49,712 in returns. The incremental cost of the program was just $2,800 ($700 per year for four years), leading to a return of over $17,000 per $1,000 investment. This is a very large return on the investment and certainly would require further verification to further solidify the magnitude of this effect. The scale of the evaluation, though, was rather large, enhancing its precision. Perhaps the cost estimates are understated in that they are reported as incremental and program expansion may require greater outlay of upfront costs. Nevertheless, these results are encouraging.

Job Corps presents a different profile. The benefits of the program are very large, but so are the costs. In an absolute sense, Job Corps generates the second largest improvement in lifetime earnings, but at $21,000 in cost per participant, the returns per dollar invested are smaller than some other programs. In reality, it is difficult to imagine a program with costs this large to be offered to more than a relative handful of participants because more widespread implementation would wipe out governmental or private foundation outlays.

The Big Brothers Big Sisters mentoring program is another one of the programs with impressive results. We find that a cost of $1,480 generates almost five times that in terms of higher lifetime wages of program participants. Providing additional college aid also has a large effect, similarly generating about five times more earnings than the additional aid provided. The Teach for America teacher training program and Success for All reforms to curriculum and instruction also have very large effects. Although these returns are quite large, it is important to remember that they are based on a single intervention, suggesting that the standard errors on these estimated returns may be quite large.

Early childhood education programs also appear to be strong performers, if not at the highest level of returns. One favorable interpretation about the impact of these programs is that there is a number of this type of intervention, including some large-scale evaluations (Head Start), and they yield relatively similar returns. This may say something about the implicit standard error being small on programs like these relative to, say, Big Brother Big Sister, which is based on a single, relatively small intervention. Placing your bets on something that is more likely to pay off, even at a smaller rate, may have a higher average level of benefit.

The remaining highlight from this table is college aid. The evidence from this literature suggests that a $1,000 investment in college aid pays off with a return of over 4.6 times that cost in the form of higher lifetime

earnings of recipients. This conclusion is based on a number of studies that explore different types of interventions lowering the cost of attending college, strengthening our belief in the reliability of this estimate. As long as these aid programs are not administratively burdensome on families (as the Pell Grant is), they appear to generate large returns in what would appear to be a lower risk investment.

We can categorize these programs, taking into consideration both the magnitude of the estimated return per $1,000 cost and the implicit standard error related to the number and size of the studies generating these estimated returns. Three programs jump out as the ones that show the most promise in terms of poverty reduction. College aid (with a simple interface) appears to be the intervention with one of the largest returns and one of the lower standard errors. We would place this intervention in this category. Similarly, early childhood education also belongs in this group. Although the returns are not estimated to be quite as large as some of the other interventions, there are a number of different types of studies that all support the existence of a sizable impact. Career Academies round out this group. These programs show very large returns and have been evaluated using credible methods and large samples. Our trade-off between risk and return would lean in the direction of supporting programs like these.

Other interventions certainly show promise. The estimated returns from Success for All and Teach for America are certainly impressive and suggest that reforming curriculum and improving teacher training may be very effective ways to proceed. The same may be said of mentoring programs like Big Brother Big Sister. Nevertheless, our risk averse nature suggests that we would be willing to support these programs much more strongly if there were more evidence supporting these findings.

References

Currie, J., and D. Thomas. 2001. Early test scores, socioeconomic status, school quality, and future outcomes. *Research in Labor Economics* 20:103–32.

Hanushek, E. A. 2006. School resources. In *Handbook of the economics of education,* ed. E. Hanushek and F. Welch, 865–908. Amsterdam: Elsevier.

Kemple, J. J., and C. J. Willner. 2008. *Career academies: Long-term impacts on labor market outcomes, educational attainment, and transitions to adulthood.* New York: MDRC.

Krueger, A. B. 2003. Economic considerations and class size. *Economic Journal* 113 (485): F34–F63.