Notes

1. State pre-K data here and following from Barnett et al. (2013). Dollar figures are in 2012–2013 dollars.
2. This book ignores recent federal proposals by President Obama and others, both because their passage is unlikely and because such discussion instantly becomes dated.
3. Unless otherwise stated, all dollar figures in this book are stated in 2012 dollars.
4. In 2012, total federal, state, and local government tax receipts plus social insurance contributions were $3,997 billion (BEA 2012).
5. The education effects are calculated based on the age 30 analysis of the Abecedarian program (Campbell et al. 2012). The earnings effects of educational improvements are calculated using data from the Current Population Survey’s Outgoing Rotation Group on African American wages, employment rates, and weekly work hours for different education groups (Bartik 2011, Technical Appendix 4B). The extra employment effects are calculated by adding on 12 percent to the treatment group employment rate at age 21, and by moderating this difference to be 6 percent at ages 27 and above. Campbell et al. (2002) find employment rate effects at age 21 that are 12 percent above the predicted effect of increased education.
6. Bartik 2011, Technical Appendix 4B. The anticrime effects depend on how imprisonment and a prison record affect earnings. Test-score effects are based on correlations between test scores and adult earnings.
7. Realizing the power of random assignment requires that there be only modest problems stemming from imperfect adherence to random assignment or sample attrition. These conditions have been met for early childhood programs. For example, the benefits of Perry Preschool have survived a reexamination of the program by Heckman et al. (2010).
8. Although random assignment experiments are the “gold standard,” they are difficult to do, expensive, and rare. Studies with good comparison groups due to “natural experiments” should be viewed as “silver standard” evidence, as the treatment and comparison groups are likely similar in unobserved characteristics. Studies that only have controls for observable characteristics provide “bronze standard” evidence: suggestive but possibly biased.
9. CPC’s evidence is less rigorous than other natural experiments because CPC is comparing voluntary participants in CPC in treatment-group neighborhoods with a sample of all children in comparison neighborhoods, which includes both families who would have participated in CPC and others. However, this comparison is better than comparing vol-
The CPC evidence suggests that the treatment and the comparison groups are similar in observed characteristics. The CPC study is somewhere between a “silver standard” natural experiment and a “bronze standard” study that controls only for observable characteristics. However, the CPC is of great value because it is one of the few large-scale studies of pre-K with long-term evidence on adult outcomes.


11. These studies control for observed individual characteristics in estimating how test scores predict adult earnings. Chetty et al.’s (2011) results rely on random assignment to classes with different average test-score effects for classmates, which controls for unobserved preexisting characteristics.

12. All test score predictions of adult earnings effects in this book are based on Chetty et al. (2011). A study’s test score effects are first translated into change in percentiles in the overall test score distribution. Where possible, this is done using the study’s information. Other times, the study only provides effect size estimates for test scores using the standard deviation in some disadvantaged group. If the study does so, this book assumes, based on Bartik, Gormley, and Adelstein (2012), that the standard deviation of test scores for the disadvantaged group is 76.5 percent of the standard deviation for all children. If no information is provided in the study on the starting test score percentile, this book assumes, based on Weiland and Yoshikawa (2013), that disadvantaged students start 0.71 standard deviations below the overall sample mean. For determining percentage earnings effects at the overall population mean, the book uses the estimated dollar and percentage effects implied by Chetty et al.’s Appendix Table V, column 1, except that these estimates are adjusted downward by the ratio of the “leave-out mean” estimates for kindergarten entrants to the ordinary least squares (OLS) estimates in Chetty et al.’s Appendix Table XIII. If the study’s target group is disadvantaged children, effects are translated into percentage effects on their earnings by using updated data to do adjustments, as in Bartik, Gormley, and Adelstein (2012), for how these children’s future earnings are likely to compare to average income, given their parents’ income relative to average income. These updated data, based on earnings data in the American Community Survey for parents of public school first-graders, predict that children eligible for a free or reduced-price lunch will have expected adult earnings equal to 71 percent of
average earnings. Subsequent endnotes will describe this procedure as this book’s standard test score prediction of earnings procedure.

13. These test score predictions differ from Bartik, Gormley, and Adelstein (2012). Their estimates assumed that the standard deviation in the disadvantaged group was equal to the standard deviation in the overall population and that the testing end score was at the median, and also predicted percentage effects at population mean earnings. This book makes more accurate assumptions.

14. This is based on Leak et al.’s (2010) end-of-treatment mean effect size of 0.28, which is assumed to be for disadvantaged groups. The earnings prediction is based on this book’s standard procedure.

15. Based on Camilli et al.’s (2010) mean short-term effect size of 0.48, which is assumed to be for disadvantaged groups. The earnings prediction is based on this book’s standard procedure.

16. The IHDP was targeted at low-birth-weight babies. The IHDP estimates used in this book are from Duncan and Sojourner’s (2013) analysis for heavier or “high” low-birth-weight babies. They argue that these results can generate good predictions for the general U.S. population.

17. Low income is defined as less than 180 percent of the poverty line. The results for low-income children show strongly significant effects for IQ at ages two, three, five, and eight. The reading results at age eight only have a p-value of 0.216, and the math results at age eight have a p-value of 0.099. However, the reading and math test score effects are consistent in magnitude with the IQ effects at ages five and eight. Therefore, on the whole the results suggest that reading and math achievement are boosted at age eight.

18. These earnings predictions use this book’s standard procedure and are based on Duncan and Sojourner’s (2013) reading and math test score effects.

19. This is based on test score effects in Ladd, Muschkin, and Dodge’s (2014) preferred model, column 2 of Table 4. Estimates for both reading and math are multiplied by 11 to reflect average 2009 funding. This standard deviation effect is then averaged across reading and math tests. The starting point is assumed to be the median, given that these are all-student averages. The implied percentile effect is then combined with Chetty et al. (2011) to get percentage earnings effects.

20. As discussed in Bartik, Gormley, and Adelstein (2012), each point in Figure 2.1 shows the average test score of a group of students, where students are sorted by age. This is done to make the pattern of how test scores vary with age easier to see.

21. Regression discontinuity (RD) studies of pre-K have been criticized, but these criticisms are not convincing. Whitehurst (2013b) argues that
RD estimates also reflect different parent behavior in the pre-K year. This is true, but the main difference is that parents enroll their kid in pre-K. The comparison group in RD studies is less likely to participate in pre-K than a randomly assigned control group that enters kindergarten next year. This affects the interpretation of the RD estimates: These estimates reflect effects of pre-K versus no pre-K to a greater extent than do random assignment estimates, which reflect this pre-K program versus other pre-K. But this is more of a feature than a bug if the pre-K estimates are being compared with the program’s costs.

RD pre-K studies have been criticized because some children who enter pre-K will not enter kindergarten in the same school district, and therefore will be excluded from the treatment group (Armor and Sousa 2014). But this attrition could bias estimates in either direction. In addition, Bartik (2013) finds that RD estimates of pre-K’s effects do not much change when we restrict estimates to the same children tested at both pre-K entrance and kindergarten entrance. Furthermore, if RD studies of pre-K are biased by attrition, this should result in a jump in other observables at the age cutoff, which is not seen in most studies. Bartik, Gormley, and Adelstein (2012) and Weiland and Yoshikawa (2013) do sensitivity tests and do not find evidence to suggest significant bias. Weiland and Yoshikawa reweight the data to correct for attrition, which makes little difference.

This is based on effect sizes of 0.4 to 0.6 in regression discontinuity studies (Bartik 2013, Table 1).

The 6-to-15-percent range is derived from the Bartik, Gormley, and Adelstein (2012) result that Tulsa’s half-day program increases adult earnings of low-income children by 6 percent, and the Weiland and Yoshikawa (2013) result that earnings of low-income children will be boosted by Boston’s full-day program by 15 percent. But this is consistent with other research. The average regression discontinuity effect size estimated in Wong et al. (2008), Hustedt, Barnett, and Jung (2008), and Hustedt et al. (2010) is an effect size of 0.407. Using this book’s standard prediction procedure, the percentage gain in adult earnings implied by this 0.407 effect size is 7.4 percent.

Based on data graciously provided by Christina Weiland, the average percentile gain in test scores for free and reduced-price lunch children is 21 percentiles. This is used to predict earnings effects using this book’s standard procedure.

Based on Tulsa (Bartik, Gormley, and Adelstein 2012). Boston estimates suggest a 15 percent earnings effect (Weiland and Yoshikawa 2013).

Chapter 2 and its notes explain how these estimates are based on Abecedarian and NFP estimates.
27. The $18,381 cost per year of Abecedarian/Educare is from Ludwig and Sawhill (2007), updated to 2012 prices.
28. The $10,050 figure is from Gault et al. (2008), updated to 2012 prices. A pre-K program is assumed to be six hours per day, class size of 15 children with 2 teachers, and lead teacher paid public school wages.
30. Future adult earnings for disadvantaged children are calculated the same as in the test score projections of earnings. The method is an updated national version of Bartik, Gormley, and Adelstein (2012). The 2012 American Community Survey is used to calculate average earnings by gender and year for all adults, and for three other groups: parents of first-graders whose family income makes them eligible for 1) a free lunch, 2) a reduced-price lunch, or 3) no lunch subsidy. As in Bartik, Gormley, and Adelstein, the average ratios of gender and age cells for these groups are projected into the future by assuming an intergenerational correlation of earnings of 0.4 (Chadwick and Solon 2002; Solon 2002). The relative weights for free-lunch versus reduced-price-lunch students are based on Tulsa pre-K enrollment. To calculate future earnings for disadvantaged children (e.g., eligible for a free or reduced-price lunch), average future earnings are multiplied by the ratio of disadvantaged children’s future earnings to overall average adult earnings, which is projected to be 71.4 percent. For children from families ineligible for a free or reduced-price lunch, the projected ratio of future earnings to the overall average is estimated to be 130.1 percent. Future real earnings are assumed to increase by 1.2 percent per year, as in Bartik (2011), based on Social Security Administration projections. Future earnings and costs are discounted to age four for pre-K, age zero for Educare and the NFP. The social discount rate is the commonly used rate of 3 percent, as explained in Bartik (2011).

In subsequent endnotes, this procedure is referred to as this book’s standard baseline earnings prediction procedure. For reference, the sum of career earnings in the ACS without any assumed real earnings increase or discounting is $1,556,000. With a 1.2 percent projected increase from age four, average future earnings without discounting sum to $2,584,000. With the 1.2 percent projected increase plus a 3 percent discount back to age four, discounted average future earnings are $766,000. If instead we project a 1.2 percent annual real increase from birth, average future earnings without discounting sum to $2,710,000. With a 1.2 percent annual real increase plus a 3 percent discount rate back to birth, discounted average future earnings are $714,000. For disadvantaged children these figures are multiplied by 71.4 percent.
For non-disadvantaged children, these figures are multiplied by 130.1 percent.

For the benefit calculations in Table 3.1, the percentage earnings gains of 9.7 percent for full-day pre-K, 26.4 percent for Educare, and 3.2 percent for the NFP are applied to the present value of future adult earnings for disadvantaged children as of age four for pre-K, and at birth for Educare and the NFP. Cost figures are also discounted to the appropriate age for each program.

A half-day school-year pre-K program at age four has parental earnings benefits, in present-value terms, of 2 percent of the child’s earnings benefits (Bartik 2011, p. 81). An eight-hour-per-day pre-K program for 45 weeks at age four has parental earnings benefits of 4 percent of the child’s benefits (Bartik 2011, endnote 38, p. 156).

Murray (2013) makes two other criticisms. “The main problem is the small size of the samples [for these two programs]. . . . Another problem is that the evaluations of both Perry Preschool and Abecedarian were overseen by the same committed, well-intentioned people who conducted the demonstration projects.” On the first criticism, small sample size is accounted for when standard errors are calculated. As Heckman has argued, “a small sample would actually work toward not finding anything. . . . There are methods that account for the small sample size. Size doesn’t matter. It holds up” (Matthews 2013). On the second criticism, Perry and Abecedarian have been analyzed by outside researchers—Perry by Heckman et al. (2010) and Abecedarian by Barnett and Masse (2007) and Temple and Reynolds (2007).


Similar criticisms are made by Whitehurst (2013b) and the Wall Street Journal (2013).

Similar criticisms are made by the Wall Street Journal (2013).

Similar criticisms are made by Wertheimer and Vedantam (2013), FactCheck.org (2013), and Murray (2013).

The test score effects in the text are for the Peabody Picture Vocabulary test, the Woodcock Johnson III Letter-Word Identification test, and the Woodcock Johnson III Math Applied Problems test, which are the three tests administered at both the end of Head Start and the end of third grade. The source for these data is Exhibit D.1A in the Head Start third-grade follow-up report (Puma et al. 2012). The average effect size for “treatment on the treated” is 0.217 at the end of Head Start and 0.062 at the end of third grade. If the estimates are uncorrelated, the standard error of the average effect size would be 0.065 at the end of Head Start.
Notes 85

and 0.054 at the end of third grade. Because the estimates are probably positively correlated, these standard errors are understated. Therefore, the true effect size in third grade could be three times as great, or it could be negative: \(-0.06\).

The earnings effects of a 0.062 effect size are predicted using this book’s standard procedure. Head Start student percentage effects are based on the free lunch group only, whose future earnings are estimated to be 67.7 percent of the population mean.

Figure 4.1 lists only some of the test score predictions noted in Table 2.1. Figure 4.1 focuses on test score fading, and therefore only uses tests that were the same at the end of the program/beginning of kindergarten and at third grade. These are more general cognitive tests, such as IQ tests. Specific achievement tests tend not to be used at the end of preschool. Academic achievement tests give higher predictions of adult earnings than more general cognitive tests. This is part of the broadening of skills that occurs over time as children develop.

Test score effects used in the earnings predictions are as follows: CPC: Reynolds (2000), Table 9, average results for four methods controlling for selection bias from Reynolds and Temple (1995); Abecedarian: Campbell et al. (2001), IQ results from Table 1; Perry: Schweinhart et al. (2005), IQ results in Table 3.3; and Head Start test score effects come from the experiment’s third-grade follow-up report, as cited in a previous endnote. Earnings predictions were done using this book’s standard procedure.

Adult earnings effects based on adult outcomes for CPC come from Reynolds, Temple, White, et al. (2011); for Perry, from Heckman et al. (2010); for Abecedarian, from Campbell et al. (2012); for Head Start, from Deming (2009).

Gormley et al. (2010) find that the average effect size of Tulsa pre-K on literacy and math tests is over 70 percent greater than the effect size for Tulsa Head Start. Wong et al. (2008) argue that their average results from five states for various tests are about twice the average short-run effect sizes of Head Start in the recent experiment. This may exaggerate Head Start’s problems, as a greater portion of the Head Start control group will enroll in an alternative pre-K program.

The test score effects in the Head Start experiment had considerably faded already by the end of kindergarten (end-of-kindergarten effects predict earnings effects of 1 percent, down from 4 percent at the end of the program). This rapid fading contrasts with the CPC (end-of-kindergarten test score effects predict earnings effects of 8 percent, similar to the prediction at the beginning of kindergarten) and with Abecedarian (first-grade effects predict earnings gains of 11 percent, down
from 13 percent at the end of the program). Perry also shows short-run fading, but still predicts much larger effects than Head Start (Perry has end-of-kindergarten and end-of-first-grade test score effects that predict earnings effects of 3 percent, down from 12 percent at the end of the program). Test score effects for these predictions come from the same sources cited in the endnote for Figure 4.1.

41. This book’s standard prediction procedure implies that even with no real earnings increases, future average earnings will be $1,556,000.

42. Cascio and Schanzenbach’s (2013) analysis of Oklahoma and Georgia concludes that the benefits of these programs will exceed costs if the programs only increase average test scores on the National Assessment of Educational Progress (NAEP) by between 1.0 and 1.4 scale points. State NAEP scores frequently jump around from one test to the next by many multiples of such amounts. This “noise”—by which is simply meant jumps in test scores due to measurement error and many random changes in state characteristics—often overwhelms plausible effects of programs.

43. This statement is based on Cascio and Schanzenbach’s benefit-cost analysis using eighth-grade test scores, at a 3.4 percent discount rate, of around 3-to-1 (Table 8). Based on fourth-grade test score impacts, the benefit-cost ratio is over 7-to-1. My reanalysis of their fourth-grade results gets slightly larger benefits of over 11-to-1. This occurs because earnings effects of test scores near the mean are somewhat larger in the Chetty et al. (2011) estimates that I use than in the Chetty, Friedman, and Rockoff (2013) estimates they use. I believe Chetty et al. provides better predictions, because it links percentile gains rather than effect size gains to earnings, and Chetty et al.’s results suggest that percentile impacts are more uniformly linear across the test score distribution.

Fitzpatrick (2008) has sometimes been cited (FactCheck.org 2013, Wall Street Journal 2013) as showing that Georgia’s program does not work, but her results suggest a benefit-cost ratio modestly greater than one for Georgia’s pre-K. If one combines Fitzpatrick’s assumption of a 0.09 effect size for 40 percent of Georgia’s child participants in pre-K and zero effects for the other 60 percent with this book’s standard test-score-to-earnings prediction procedure, the present value of earnings effects is 1.3 times costs, estimated at $3,652 in Cascio and Schanzenbach (2013). Fitzpatrick does not get this result because she assumes a fixed dollar effect of test score increases on hourly wages, which seems implausible.

44. Cascio and Schanzenbach (2013) state that the most precise test score results for fourth-grade math are only significantly different from zero at the 20 percent level.
45. A possible sign of such biases is some discrepancy between results for the smaller sample with parental consent and the full sample (Lipsey et al. 2013b). The discrepancy concerns one of the few outcomes for which parental consent is not needed, which is whether the child is retained in kindergarten. The smaller sample finds that Tennessee pre-K reduces retention in kindergarten from 6 percent to 4 percent. The full sample finds that Tennessee pre-K reduces retention in kindergarten from 8 percent to 4 percent. If Tennessee pre-K truly has no effect on a child’s performance as of the end of kindergarten, it is strange that the full sample finds that the odds of retention are cut in half.

46. It is noteworthy that Tennessee’s end of pre-K results, with an effect size of 0.24, imply an adult earnings effect of 4 percent, which is low for a full-day program compared to the 10 percent boost for low-income children estimated in Tulsa (Bartik, Gormley, and Adelstein 2012) and the 15 percent boost implied by results for Boston (Weiland and Yoshikawa 2013).

47. In these studies of how teacher-child interactions affect learning, differences between students are controlled for using only observable variables, which might leave some bias. However, many researchers would be inclined to believe that teacher-child interactions make a difference, because such an effect seems so plausible.

48. Based on Sabol et al. (2013). The lowest-quality level is below 2.5 on the 7-point Classroom Assessment Scoring System (CLASS) scale of teacher-child interactions, and the highest quality is 5.5 or above on the CLASS scale (see supplementary materials for Sabol et al. 2013). The five cognitive test results in Table S4 were averaged to get an average effect size increase of the highest quality “level four” versus “level one” of 0.322. The earnings effects of a 0.322 effect size increase in test scores at the end of pre-K are calculated using this book’s standard procedure.

49. This is calculated by multiplying the 6 percent earnings effect by the present value of future adult earnings for children from families eligible for a free or reduced-price lunch, predicted using this book’s standard procedure. The resulting present value of the 6 percent gain is around $33,000 per child, which, multiplied by 15 children per class, would come out to around $495,000.

50. For example, using the results in Burchinal, Kainz, and Cai (2011), which show an effect size for cognitive outcomes that averages 0.0833 for a one-standard-deviation increase in CLASS instructional support in a low-income sample, and using this book’s standard prediction procedure, a one-standard-deviation increase in CLASS instructional support will raise earnings in this sample by 1.5 percent, which has a present value of $120,000 summed over 15 students in a class. Using results
from Keys et al. (2013) for a more mixed income sample—results that show that CLASS instructional support increases this sample’s cognitive test scores by an effect size of 0.05—we find that a one-standard-deviation increase in CLASS instructional support will increase this sample’s earnings by 1.1 percent, which has a present value of $131,000 summed over 15 students in a class. (The smaller percentage increase yields a larger present value because this sample’s adult earnings will be more typical of the population.)


52. All of these programs used certified teachers. The available data on class sizes suggest they were moderate to small. Perry averaged 13 students to 2 teachers, Abecedarian 14 students to 2 teachers, and Chicago Child–Parent Center pre-K averaged 17 students to 2 teachers.

53. The increased costs are calculated by updating the figures from Table 1 in Gault et al. (2008) to 2012 dollars. As detailed in Bartik, Gormley, and Adelstein (2012), the figures in Chetty et al. (2011) suggest that a one percentile increase in test scores increases future adult earnings in dollar terms for all groups by 0.495 percent of mean overall earnings. Average future earnings are predicted using this book’s standard baseline earnings prediction procedure, described in a previous endnote. The present value as of age four of these average earnings is $766,000. The $2,500 increment to costs is 0.33 percent of average earnings. Therefore, the needed percentile test score increase is 0.67 percent, which equals 0.33 percent divided by 0.495. Such a $2,500 increment to earnings would be about 0.5 percent of the present value at age four of future earnings for disadvantaged children, which is calculated to be $547,000.

54. The cost difference assumes lead teachers are paid public-school wages. Calculations are done similarly to the calculations for increased credential requirements, and they use similar sources.

55. These calculations are presented and explained in Bartik (2011, pp. 137, 152–153).

56. For Tulsa’s full-day pre-K, percentile test-score gains for middle-class children are 89 percent of those for lower-income children (Bartik, Gormley, and Adelstein 2012). For half-day pre-K, the ratio is 88 percent. In Boston, percentile test-score gains for middle-class children are 71 percent of those for lower-income children (Bartik’s calculations, based on Weiland and Yoshikawa 2013).
57. Table 5.1 is based on Bartik, Gormley, and Adelstein (2012). However, estimated gains and earnings figures use this book’s standard baseline earnings prediction procedure.

58. These Tulsa results average results for both free and reduced-price lunch children using the same overall weights. These Tulsa results, from Bartik, Gormley, and Adelstein (2012), can be challenged because they cannot control for differences in unobserved characteristics between families that opt for half-day versus full-day pre-K. However, similar diminishing returns to a longer pre-K day are found in a random assignment experiment in New Jersey (Robin, Frede, and Barnett 2006). See Bartik (2011, p. 145) for more discussion.

59. These diminishing returns are based on the kindergarten test score effects of children in the Chicago Child-Parent Center who participated in two years versus one year (Reynolds 1995). Bartik (2011, p. 146) explains the calculations. Reynolds, Temple, White, et al. (2011) report similar diminishing returns based on an overall benefit-cost evaluation of two-year and one-year participants in CPC. Other studies are more pessimistic about the returns to adding a second year of pre-K. The supplement to Reynolds, Temple, Ou, et al. (2011) finds no evidence of any annual earnings or educational attainment differentials between the two-year group and the one-year group. Arteaga et al. (2014) also find no advantages for educational attainment or economic status between the two-year group and the one-year group. Finally, the meta-analysis in Leak et al. (2010) also suggests diminishing returns, although the extent differs with whether weights are used in the regressions. With weights, a two-year program only has an effect size larger by 0.02 than the baseline one-year program at 0.21, a 10 percent increase. Without weights, the prediction is that the two-year program will have an effect size that is higher by 0.11, about a 50 percent increase.

60. Based on Bartik, Gormley, and Adelstein (2012), but updating to 2012 national earnings using this book’s standard procedure, a half-day pre-K program for children whose families are eligible for a free or reduced-price lunch will have a present value of future earnings gains of $32,742. The costs of a program with a child-to-teacher ratio of 15-to-2, paying certified teacher wages, will be $5,418 in 2012 dollars, based on Gault et al. (2008). The resulting ratio of earnings benefits to costs is 6.04. The full-day program, summarized in Chapter 2, has earnings benefits of $52,920, costs of $10,050. The net additional benefits of going from a half-day to a full-day program are $20,178, the net additional costs are $4,652, and the resulting incremental benefit-cost ratio is 4.36.

61. The review by Leak et al. (2010) is also consistent with this finding. As summarized by Duncan and Magnuson (2013, p. 115), “Analysis of the
meta-analytic database shows that . . . effect sizes were neither larger nor smaller for children who started programs at younger ages.” Because these younger-age programs will have smaller class sizes, this implies lower benefit-cost ratios.

62. For an example of this perspective, see Dinesh D'Souza, as discussed in Bartik (2011, p. 319).

63. To be exact, $1,245,000 is the estimated dollar value of increased earnings from college graduation over a career without any secular wage increases or discounting. This college graduation premium is $2,168,000 with annual 1.2 percent real wage increases, and $571,000 if this $2.2 million is discounted at 3 percent annually.

64. Spillover wage benefits may not be uniform across all other groups. Moretti’s estimates suggest greater nominal wage benefits for non-college graduates, while Diamond, in contrast, shows that the combined changes in local amenities and housing prices may disproportionately help college graduates. At the national scale, one would expect much of the overall increase in prices to disappear, while the effects on productivity would remain.

65. My 2011 book included some spillover effects, but I made relatively conservative assumptions that forced these spillover effects to be smaller than I now think is plausible.

66. There may also be spillover effects of local skills, not only on productivity levels but on productivity growth, as argued in Dickens and Baschnagel (2008).

67. The Chicago Child-Parent Center reduced students in special ed from 25 percent to 14 percent (Reynolds, Temple, White, et al. 2011). Perry reduced some special ed services, for mental impairment, from 35 percent to 15 percent, but increased some other special ed services. Overall years in special ed went from 5.2 years to 4.0 years (Schweinhart et al. 2005). Bagnato, Salaway, and Suen (2009) found that Pennsylvania’s pre-K program reduced special ed services in the treatment group to 2.4 percent, versus 18 percent in the comparison group.

68. This averages test-score effects across five tests in Table 3 of Henry and Rickman (2007).

69. Universal pre-K costs are based on Gault et al. (2008). Annual full-time child care/pre-K costs are Abecedarian figures from Ludwig and Sawhill (2007), which in turn are based on Masse and Barnett (2002). Annual home-visiting/parenting costs per child are from the NFP (http://www.nursefamilypartnership.org/assets/PDF/Fact-sheets/NFP_Benefit_Cost.aspx). Oklahoma’s state pre-K participation rate of 74 percent comes from Barnett et al. (2013). Assumptions for Educare on targeting to the poor and 75 percent take-up are the same assumptions as made by
Ludwig and Sawhill (2007). Assumptions for home-visiting/parenting are based on Isaacs (2007), who implies that 8.9 percent of all children are currently first births to families below 185 percent of the poverty line; Isaacs (2007) also assumes a 75 percent take-up rate. The number of children at each age under five comes from 2012 figures of the U.S. census.

70. Government tax figures are for 2012, from BEA (2012). “Contributions for social insurance” are included in taxes.

71. Under plausible assumptions, the direct earnings effects on former participants of the $79 billion proposal would boost the present value of earnings by $207 billion, about 2.6 times the cost. This assumes that in the universal pre-K program, 47 percent of pre-K participants will be needy children who will get the higher benefit-cost ratio of 5.3, and the other 53 percent will be middle-class children who will get a lower benefit-cost ratio of 4.7 to 1 (see Table 5.1). The current $6 billion spent on state pre-K is assumed to be split 85 percent to needy children, 15 percent to middle-class children. We then add in the incremental costs and benefits of the Educare and NFP proposals using Table 3.1.

72. The CBO classifies households into income quintiles based on before-tax income, adjusted for household size.

73. Average after-tax and transfer real income growth for all households from 1979 to 2007 was 67.5 percent, compared to 41.0 percent for the lowest-income quintile and 34.8 percent for the middle-income quintile. Dividing 1.675 by 1.41 yields 1.19, and dividing 1.675 by 1.348 yields 1.24, so the last year’s income of these two groups needs to be blown up by 19 percent and 24 percent to match average overall income growth. These figures come from Supplemental Data files, Table 6, at http://www.cbo.gov/publication/43373.

74. For the average lowest-income-quintile household in 2007, total labor income was 61.1 percent of after-tax and transfer market income. So an earnings boost of 19 percent / 0.611 = 31 percent would be required to make up for lost income growth relative to the average household. For the middle-income quintile, 77 percent of those families’ after-tax and transfer market income was labor income, so a labor income boost of 24 percent / 0.77 = 31 percent would be required to make up for lagging real income growth.

75. A recent review argues that “multiple recent studies suggest a highly promising route to quality in preschool education: providing support for teachers to implement evidence-based curricula and instruction through coaching and mentoring” (Yoshikawa et al. 2013, p. 15).

76. For the half-day program considered in Bartik (2011), the ratio of the present value of earnings to program costs was 2.78 from a state per-
spective, 3.79 from a national perspective. The ratio of the state benefits to the national benefits is 73 percent = 2.78 / 3.79.

77. If universal pre-K increases property values by six times annual program costs, annual property taxes raised would be 8 percent of annual program costs (Bartik 2011, Table 7.3).

78. This 4 percent figure may seem low considering that K–12 covers 13 grades and pre-K corresponds to one grade. However, consider the following: The gross cost is $31 billion, with the $25 billion net cost due to states already spending $6 billion on pre-K for four-year-olds; average public school expenditure divided by fall enrollment count in 2010–2011 was $12,048 (National Center for Education Statistics 2013), almost 20 percent more than the cost per child of universal pre-K of $10,050; the assumed enrollment in universal pre-K at age four is 3.0 million, which is well below one-thirteenth of total K–12 enrollment of around 49 million, both because of lower births in recent years and because of the assumption that only 74 percent of all students will take up universal pre-K given the presence of Head Start as well as parents choosing other alternatives.