

NBER WORKING PAPER SERIES

**RACE, CHILDREN'S COGNITIVE
ACHIEVEMENT AND THE BELL CURVE**

**Janet Currie
Duncan Thomas**

Working Paper No. 5240

**NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 1995**

We are grateful for financial support from NICHD 1R01-HD31722 and NSF SBR95-12670; Janet Currie also thanks the Alfred P. Sloan Foundation for support. The comments of Arthur Goldberger, Sanders Korenman, and James P. Smith have been very helpful. This paper is part of NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

© 1995 by Janet Currie and Duncan Thomas. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

RACE, CHILDREN'S COGNITIVE
ACHIEVEMENT AND THE BELL CURVE

ABSTRACT

In *The Bell Curve*, Herrnstein and Murray demonstrate that a mother's score on the Armed Forces Qualification Test (AFQT) is a powerful predictor of her child's score on the Peabody Picture Vocabulary Test (PPVT). We replicate this finding for PPVT and two related tests. However, even after controlling for AFQT, there are significant racial gaps in PPVT scores which suggest that AFQT is not all that matters. In fact, both maternal education and income are important determinants of child test scores and their influences differ dramatically with the test, the child's age, and the child's race. These racial gaps in test scores are important because, even within families, children with higher scores are less likely to repeat grades. Moreover, conditional on child test scores *and* maternal AFQT, maternal education and income affect the probability of grade repetition. We move beyond AFQT and examine the effects of individual Armed Services Vocational Aptitude Battery sub-tests on children's scores. We find that those skills that are rewarded in the labor market are not always the same skills that are associated with improved child outcomes. An understanding of the relationship between different aspects of maternal achievement and child outcomes may help unravel the complex process through which poverty is transmitted across generations.

Janet Currie
Department of Economics
University of California
405 Hilgard Avenue
Los Angeles, CA 90024
and NBER

Duncan Thomas
Department of Economics
University of California
405 Hilgard Avenue
Los Angeles, CA 90024

The intergenerational transmission of poverty from parents to children is an established fact,¹ but the causes of persistent poverty among groups remain highly controversial. On the one hand, many observers believe that the children of poor parents lack the opportunities available to other children and hence are less likely to realize their full potential.² An alternative view, most recently expressed in *The Bell Curve* by Herrnstein and Murray (1994), is that the poor are simply inherently less able than their peers, and that it is this lack of ability that is transmitted from one generation to the next.

Herrnstein and Murray focus on the relationship between two test scores: the mother's Armed Forces Qualifications Test (AFQT) and the Peabody Picture Vocabulary Test (PPVT) of her child -- both are referred to as "IQ". They show that there is a strong correlation between the two scores, and argue that "...the IQ of the NLSY mothers was much more important than their socio-economic background in determining their children's IQ" (Herrnstein and Murray, pg. 230). In a logistic regression of the probability that a white child's PPVT score falls into the lowest decile, the effects of an index of socio-economic status, maternal poverty, and home environment are "both small and statistically insignificant" when mother's IQ is included in the model (pg. 232 and pg. 618). Moreover, they argue that "IQ scores on a properly administered test mean about the same thing for all ethnic groups" (pg. 315).

In this paper, we use data from the same source as Herrnstein and Murray to re-examine the links between the test scores of mothers and children, paying particular attention to the question of what these test scores measure and whether they do in fact mean the same thing for blacks and whites.³ The first section of the paper describes the data. In addition to the PPVT that Herrnstein

¹Solon (1992) and Zimmerman (1992) provide recent estimates of the size of the correlation between parent's and children's incomes.

²There is a large literature on the effect of early education and health interventions; see, for example, Barnett (1992) for an excellent review. Currie and Thomas (1995) provide a recent evaluation of Head Start.

³There are several excellent reviews of *The Bell Curve* (c.f. Goldberger and Manski, 1995; Heckman, 1995; Hunt, 1995; Korenman and Winship, 1995). Those reviews discuss a wide range of the claims made in the book; in this paper, we examine, in detail, a specific question: how is maternal background, including test

and Murray examine, our analysis incorporates two widely used tests of children's cognitive achievement, the Peabody Individual Achievement Tests for Reading Recognition (PIAT-READING) and for Mathematics (PIAT-MATH).⁴ We find that racial gaps in test scores are much narrower for both PIATs than for the PPVT and conclude that exclusive examination of the PPVT exaggerates differences between white and black children.⁵ We also find that the PIAT-READING score is a better predictor of grade repetition than the PPVT, especially among black children. This suggests that the PPVT does not capture all relevant aspects of "general intelligence" and that it is important to analyze a broader array of test scores.

In section II, we re-examine the claim that conditional on maternal AFQT scores, socioeconomic status has no effect on children's test scores. While maternal AFQT does have a powerful effect, we show that measures of maternal circumstances such as highest grade completed and household income are also important. Moreover, maternal education appears to have a more powerful effect on black children's scores than on white children's, implying that across-the-board increases in maternal education would tend to lower the racial gap in children's test scores. And rather than assume, as Herrnstein and Murray do, that test scores are immutable, we trace out the way that the relationship between maternal characteristics and test scores varies with the child's age.

In section III, we complement our analysis of the effects of maternal AFQT on children's scores, by examining the effects of child test scores on an important child outcome: grade repetition. Grade repetition has been shown to be a significant predictor of eventual schooling attainment, and hence of wages. Our analysis shows that racial gaps in test scores are a legitimate cause for concern because, even conditional on AFQT, maternal education, and income, children's test scores

scores, related to child outcomes?

⁴The survey also includes the PIAT Reading Comprehension test. Since a child must achieve a minimum score on PIAT-READING before the Comprehension test is administered, there is a serious issue of selectivity in the sample of children who were administered the Comprehension test. We therefore ignore that test.

⁵ This point has also been made by Korenman, Miller, and Sjaastad (1995).

are correlated with grade repetition. Equally important however, is the fact that conditional on child test scores we find that measures of family background have independent effects on the probability of grade repetition. Hence the models of test scores and the models of grade repetition demonstrate that both types of indices are important predictors of child outcomes.

Finally, in section IV, we exploit the fact that the AFQT is a composite measure constructed from four of the ten tests that are routinely administered to potential military recruits.⁶ The full battery of tests covers an array of skills including, for example, verbal comprehension, numeracy, and knowledge of mechanics. Rather than assume AFQT is the "best" measure of "intelligence", we examine the effect of all ten sub-scores on children's test scores, and for comparison, on women's wages. Our results are provocative. We find that the skills that are valued in the labor market are not necessarily the most valuable in terms of child development. Moreover, the relationship between various maternal scores and children's scores varies dramatically with race, and with the child's age. Therefore, instead of focusing on the similarities between the tests (by looking at all of them as manifestations of "general intelligence"), it may be more useful to focus on the differences between them in an effort to better understand what aspects of achievement they measure, and how different types of maternal achievement are related to child outcomes.

I. The Data

Both *The Bell Curve*, and the present analysis rely on data from the National Longitudinal Survey of Youth (NLSY) and the National Longitudinal Survey's Child Mother files (NLSCM). The NLSY is a panel survey that began in 1979 with a sample of 6,283 young women and a similar number of young men who were between the ages of 14 and 21 in 1978. The survey includes a nationally representative sample of young men and women along with over-samples of blacks, hispanics, and poor whites. In our analyses, we delete the over-sample of poor whites, and all

⁶ Interestingly, the service does not actually use AFQT for job placements - different composites are constructed for different jobs (Wigdor and Green, 1991).

hispanics, so that we have nationally representative samples of black and white mothers age 26 through 33 in 1990; our analytical sample is comprised of these women and their children.

The NLSY sample has been resurveyed every year since 1979. A great deal of information is available about educational attainment and family income, two measures of socio-economic status that we will focus on. The AFQT was administered to all sample respondents in 1980. Since the respondents are different ages, we follow Herrnstein and Murray and many previous authors, and standardize the AFQT scores by creating an age-specific z-score.

Beginning in 1986, the children of the NLSY mothers have been assessed at 2 year intervals. These assessments are combined to form the NLSCM. As of 1992, the last year of data that is available, over 11,000 assessments have been completed, covering 4,300 children age 3 through 15. The NLSCM administered the PPVT to all children 3 and over, while the PIAT's were administered to children 5 and over (and hence sample sizes differ depending on the specific test). These tests are widely used and have been normed using age-specific national samples (see Baker and Mott, 1989; Baker *et al.*, 1993, and references therein for further information). Whether the child has ever repeated a grade is reported for all children 10 and over.

The NLSCM survey protocol called for the PPVT to be administered once per child until 1992, when it was administered to all children including those who had taken it previously. In principle, the PIATs were repeated as many times as the child was age-eligible for the test. We exploit the availability of all these test scores and trace out their evolution over the life course from early childhood to adolescence: the average number of times a child has taken each test ranges from 1.8 for the PPVT to 2.5 for the two PIATs.

Our sample consists of 2,164 mothers and 8,707 children who had received at least one of these three assessments by 1992.⁷ Approximately 80% of the children have a sibling in our sample,

⁷Herrnstein and Murray examine only the 1986, 1988, and 1990 waves of the NLSCM, examine children who were given the PPVT at age 6 or over, and use only one observation per child. Their final sample contains 1,784 children. There are a total of 4,607 children with PPVT scores in the 1986-1990 data and substantial differences emerge between the complete sample and the one used by Herrnstein and Murray. See the

a fact that we will take advantage of below. To summarize the essential features of the backgrounds of these children, characteristics of their mothers are reported in Appendix Table 1. Mean highest grade completed, the mean (log) of household permanent income, and the mean percentile of AFQT of the children's mothers are at the top of the table.

On average, black mothers have less education than white mothers; while the gap of 0.2 years is statistically significant, it is relatively small. Permanent income, measured as average household income over the 1978 to 1991 period, is intended to capture longer-run socioeconomic status.⁸ Racial gaps in income are large and significant: the average income of a black mother is \$19,000 compared to an average of \$32,500 for a white mother. As Herrnstein and Murray point out, black mothers also tend to have much lower AFQT scores. The score is expressed in terms of the percentile of the national distribution of AFQT scores and indicates that the average black mother was ranked at the 20th percentile, while the average white mother was ranked at the 47th percentile. This gap is equivalent to a full standard deviation difference in the average scores of blacks and whites.

By definition, the median person in a nationally representative sample is at the 50th percentile, so these numbers are low. One reason may be that more privileged women tend to delay child-bearing. By 1992, over 80% of the (nationally representative sample of black and white) women in the NLSY had given birth to one child. While our sample of mothers are relatively poor, the differences between the mothers and the full sample of women are small. For example, in our sample, the average woman is at the 35th percentile of the AFQT distribution and has 12.5 years of education while in the full sample, the average woman's AFQT score is at the 40th percentile and she has 12.7 years of schooling.

replication by Korenman and Winship (1995) for details.

⁸ Geronimus *et al.* (1994) argue that long-run averages are better measures of socio-economic status than income measured at a point in time, in part because annual income is subject to short run variations, and in part because measurement error is attenuated by averaging.

The rest of Appendix Table 1 reports z-scores of the ten Armed Services Vocational Aptitude Battery (ASVAB) sub-tests. These figures offer an initial look at how the test scores of black and white mothers differ. Herrnstein and Murray claim that "blacks and whites differ most on the tests that are the best measures of g or general intelligence" (pg. 270). Putting aside questions about the usefulness of "g", the basis for this claim is unclear. We find that the largest gaps (1.1 and 1.0 standard deviations) are on "Word Knowledge" and "Paragraph Comprehension" tests, two of the components of the AFQT. The other two components are "Math Knowledge" and "Arithmetic Reasoning". The differences in scores of blacks and whites are much smaller on these tests (.64 and .85 standard deviations) and it seems likely that these skills would also be correlated with "g". As we will see below, the same pattern of large gaps in vocabulary tests but smaller gaps in mathematics emerge in the child test scores.

The racial gaps in the ASVAB scores of women are also relatively small for "Automobile and Shop Information" and for "Mechanical Comprehension", subjects in which neither black nor white women are likely to have had much training. These patterns suggest that rather than reflecting differences in "intelligence", the racial gaps may be largest in those areas in which white women have had the greatest superiority in training. This interpretation is bolstered by a comparison of the ASVAB scores of black and white men: among them, the largest racial gaps are in the three vocational tests ("Automotive and Shop", "Mechanical Comprehension", and "Electronics Information").

Means and standard errors of the z-scores of the children's test scores at each age are shown in Appendix Table 2a.⁹ Sample sizes are given in brackets. The table indicates that there is a very wide gap between the PPVT scores of white and black children, which varies from approximately 2.2 standard deviations to 1.5 standard deviations, depending on the child's age. In contrast, the

⁹ The NLSY children tend to have mean scores above the national norms on the PIAT tests. Baker and Mott (1989) suggest that this may be due to the fact that the norms for these tests are old, and basic reading and mathematical skills may have improved in the child population.

gap in the PIAT-MATH varies from 0.7 to 0.9 standard deviations, while the gap in PIAT-READING varies from 0.2 to 0.9 standard deviations. In general, standardized PPVT scores tend to rise with age, at least for blacks, and the black-white gap declines. In contrast, PIAT scores of both blacks and whites tend to decline with age and the black-white gaps in the PIATs are larger among older children, which suggests that black children are falling further behind in the dimensions that these tests measure. Finally, Appendix Table 2b shows that black children are more likely to repeat grades than white children; by age 15, almost half black students have repeated a grade, compared to a third of white students.

Appendix Tables 1 and 2 indicate that there is, on average, a large gap between the AFQT scores of white and black mothers and between the test scores of white and black children. In part, these gaps reflect differences in socio-economic status. But that is only part of the story. A preliminary exploration of the relationship between the three child test scores and AFQT appears in Figure 1; the relationship between child test scores and household permanent income is reported in Figure 2.¹⁰

Figure 1 confirms that there is a positive relationship between children's scores and maternal AFQT among both whites and blacks. The racial gap in PPVT scores is large, even after controlling for AFQT, and it is approximately constant across the entire distribution of AFQT. If we follow Herrnstein and Murray and interpret PPVT scores as "intelligence" then the figure leads

¹⁰The figures are estimated non-parametrically and so place no restrictions on the shape of the relationship between child's test scores and maternal AFQT or income. We use the method of locally-weighted smoothed scatterplots (LOWESS, see Cleveland, 1979 and Hardle, 1990). Essentially, each observation is replaced by its predicted value based on a weighted regression of the dependent variable on the independent variable (say income) The sample in each weighted regression includes only those children from households whose incomes are in the same "neighborhood" or band. The fraction of children included, i.e. the band-width, is arbitrary and we have experimented with a range of values. The estimates in these figures include 20 percent of the sample. The weights in each of the LOWESS regressions are given by $w_j=(1-d_j^3)^3$, where d_j is a measure of the difference between the income of child j and the income of the child of interest, i . In this case, $d_j=(s_j-s_i)/(s_i^*-s_i)$, where $(s_i^*-s_i)$ is the maximum income gap between the child of interest and all other children in the 20 percent band. The weight is positive for every observation within the band and the weighting function is approximately Gaussian in shape. Since $d=0$ for the child herself, her weight is equal to one; the weights decline as the income gap increases. Children outside the band receive zero weight which is why this is a local rather than global estimator.

to the conclusion that if a white and black mother have the same intelligence, the white child will be smarter, on average, than the black child. Put another way, Herrnstein and Murray would conclude that white mothers are more efficient at transmitting "intelligence" to their offspring. However, this is true only for PPVT. When we turn to the PIAT scores in the right-hand panel of the figure, we find that there are virtually *no differences* in the scores of black and white children conditional on maternal AFQT. Thus, this figures suggests that black and white mothers are equally productive in the transmission of (PIAT-related) skills to their children.

One possible interpretation of these results is that there is more measurement error in PPVT than in the PIATs. Measurement error in PPVT could, in a linear model, yield two parallel lines like those shown in Figure 1, since children who are observed to have low PPVT scores are more likely to have had a "bad" draw on the test, while children who are observed to have high scores are more likely to have had a "good" draw. Children from families with lower socio-economic status are also more likely to have low PPVT scores. Thus, there may be a correlation between low socio-economic status and a "bad" draw on the test. Since black children are disproportionately from low socio-economic status, they are more likely to have had negative errors and so their scores will be everywhere below those of whites, who are more likely to have had positive errors. Hence, in a linear model, high random error rates will tend to generate parallel lines for blacks and whites.

But, the estimates in Figure 1 are not constrained to be linear. In our non-parametric framework, measurement error of the type described above would produce a sigmoidal curve since the probability of a negative error rises as we move towards the left tail and a positive error is more likely as we move towards the right tail. There is no evidence of this shape in the figure.

As a second test of the measurement error explanation, we re-estimated the PPVT-AFQT relation for whites, stratifying on whether or not the mother was poor in 1978. On average, children of poor women score about 0.5 standard deviations below those whose mothers were not poor in 1978. If the measurement error hypothesis is correct, the scores of poor children should lie below those of non-poor children, conditional on AFQT, just as the scores of black children lie

below those of white children. However, as shown in Appendix Figure 1(b), after controlling for maternal AFQT, the gap in scores is small. The same conclusion follows when we stratify on whether the mother was born in the South. Thus, random measurement fails to explain the racial gaps in PPVT scores in Figure 1.¹¹

Figure 2 illustrates a second problem with Herrnstein and Murray's interpretation of the relationship between AFQT and child outcomes: essentially the same patterns emerge for the relationship between the three test scores and household permanent income (except that there is a larger racial gap in the PIAT-READING score). Relationships between maternal education and child test scores (which are not shown) are remarkably similar. This similarity raises the question of whether AFQT is in fact an index of socio-economic status, ability, and background rather than strictly a measure of "intelligence" *per se*. As Heckman (1995, page 21) comments, AFQT "... may actually be a better measure of the environment facing children than the measure of environment used by Herrnstein and Murray".¹²

¹¹The relationship between PPVT and AFQT shown in Figure 1 is the child analogue of Herrnstein and Murray's figure (pg. 288) which relates AFQT to socio-economic status among adults. Their figure shows that the racial gap in maternal AFQT scores is approximately constant at all levels of socio-economic status and they conclude that lower average SES among blacks cannot explain racial gaps in AFQT scores. But their figure implies that for any given level of AFQT, blacks achieve *higher* socio-economic status than whites. We plot women's wages (measured in 1992) against AFQT in Appendix Figure 1(a) and show much the same relationship. This may reflect widespread discrimination in favor of blacks (which is constant for those whose AFQT z-score is greater than -1.5) or it may reflect systematic error in the test which is correlated with race (since AFQT under-predicts black performance). Distinguishing between these very different implications is key to the interpretation of the results, but our reading of the literature indicates that the facts have not been convincingly established. It has been argued that AFQT is a racially unbiased test of ability (c.f. Neal and Johnson, 1994). Rodgers and Spriggs (1995) take a different view and point out that if measurement error in the AFQT test is correlated with race then that can also explain the racial gap in the AFQT-wage relationship. A National Research Council investigation of these issues compares AFQT scores with hands-on performance in specific tasks in the military. They concludes that "...the data are thin and should be interpreted with caution. Blacks scored lower than nonminorities on both selection tests and job performance measures...[however] the magnitude of the average score differences between blacks and nonminorities is much larger on the AFQT (-.85 of a standard deviation) ... than on the hands-on test (-.36 of a standard deviation)...[these findings] suggest that scores on the AFQT exaggerate the size of the difference that will ultimately be found in the job performance of the two groups" (Wigdor and Green, 1991, page 12).

¹²See, also, Korenman and Winship (1995) who investigate measurement of socio-economic status in some detail.

Three facts have been established in this section. The racial gap in PPVT scores is approximately constant at all levels of maternal AFQT; the gap cannot be explained by measurement error alone; and the relationship between child test scores and AFQT is remarkably similar to the relationship between scores and other measures of socio-economic status, in particular income and education. These facts suggest that differences in maternal test scores may not be the only factor underlying the gap in child scores and call into question Herrnstein and Murray's finding that PPVT scores of whites are influenced only by maternal ability. We will now test that claim directly.

II. Is it True that Only Mother's Test Score Matters?

The claim that conditional on maternal AFQT, socio-economic status does not matter is central to Herrnstein and Murray's arguments about the sources of group differences in child test outcomes if AFQT is predetermined but measures of socio-economic status, such as income, may be subject to change. If it is only the test score of the mother that matters, then attempts to close racial gaps through, for example, transfers to mothers of young children, will be futile.

In this section, we present regression models that simultaneously control for the effects of AFQT, highest maternal grade, and family permanent income.¹³ Given Figure 1, we certainly expect AFQT to matter. But, holding AFQT constant, do education and income also affect child test scores? Results are shown in Table 1 for children of various ages. We stratify by the age when the child was measured in order to deal with the fact that, as shown in Appendix Table 2,

¹³Herrnstein and Murray use a logistic framework to examine the determinants of a child scoring in the bottom decile of the national PPVT distribution. The evidence in Figures 1 and 2 suggests that discretizing the scores into this binary outcome will throw away a good deal of information; our dependent variables are, therefore, the z-scores for each test. The index of socio-economic status adopted by Herrnstein and Murray is a function of the child's grand-parents' education and occupation and the income of the mother's household in 1978 and 1979. We prefer to place maternal AFQT and background on a more equal footing and hence include the mother's own education and long run household income as determinants of her child's achievement.

children's standardized test scores vary by age in this sample.¹⁴ To account for potential differences in test environments across survey waves, we control for the year of assessment. For ease of interpretation, we show the coefficients for blacks and for whites, estimated from separate regressions, as well as the difference between blacks and whites.

Among white children, the highest grade completed by the mother is only statistically significant for the youngest children. The effect of maternal education is consistently positive for all black children -- an additional year of education would increase the child's test scores by between .13 and .19 standard deviations. The differences between whites and blacks are statistically significant in half of the regressions. Permanent income has a positive effect on the scores of both white and black children. Whereas, among whites the income effect diminishes with the child's age, the effects are largest among the oldest black children. If the permanent income of black mothers was increased to be equal to that of white mothers, then the PPVT score of an average 3 to 4 year old black children would increase by about the same amount as if the mother spent an additional year at school. Test scores of the oldest black children in the sample would rise by substantially more -- about the equivalent of two extra years of maternal schooling. As demonstrated by Herrnstein and Murray, maternal AFQT has a strong positive effect on children's PPVT scores. This impact tends to grow slightly as the child ages, especially among black children. An increase of one standard deviation in maternal AFQT scores increases the child's PPVT score by between .3 and .8 standard deviations.¹⁵

¹⁴This strategy also ensures that each child is not included more than once in a regression.

¹⁵One reason for these differences between whites and blacks may have to do with other background characteristics that are omitted from the analysis: obvious candidates are characteristics of the father. Our long run income measure should capture some of his characteristics. However, the NLSCM contains little information other than the age and education of the father and we have no measures of his cognitive achievement. We have estimated models that do include his education (and also control for whether he has ever been in the household). The most striking fact is that, for both blacks and whites, there is no change in the estimated impact of maternal AFQT and little change in the impact of education and income. The effect of income declines by about 20% for whites but increases for blacks. Maternal education declines by about 10%. Paternal education has a positive and significant influence on test scores of young children and the magnitude is similar to that of black mothers. 25% of the children in the sample do not have a father who ever resided in the household. We have, therefore, also experimented with using the characteristics of the

Results for the PIAT-READING and for PIAT-MATH tests are shown in the second and third panels of Table 1. Maternal AFQT has a strong positive effect on both test scores at all ages (although among children 5 to 7 years old, the effect on PIAT-MATH is much larger for whites than blacks). As with the PPVT, maternal education has little impact on test scores of whites but it has a positive and significant effect on black scores, particularly PIAT-READING scores. This effect is relatively constant over the life course and an increase of one year in maternal education raises the child's PIAT-READING score by between .07 and .15 standard deviations. Higher family income is also associated with higher PIAT scores. For PIAT-READING, the impact tends to fade with age among white children and is not significant among 13 to 15 year olds. But, among blacks, the role of income tends to be greatest for older children.

After controlling for these measures of background, the main effect of being black is either insignificant or positive in models of PPVT and PIAT-READING. For PIAT-MATH, black children continue to score lower at least within some age groups.¹⁶

In sum, the evidence is clear: maternal income and education do affect children's test scores, even after controlling for AFQT. But the influence of each appears to differ by age and race. The fact that the effects of background vary with age indicates that other factors come into play during the child's life which enhance or reduce the importance of maternal inputs at various stages in the child's development. We interpret this as evidence that children's test scores are not pre-determined by maternal "intelligence" alone. For all tests, there is a tendency for the impact of AFQT to be greatest among older white children whereas the role of income tends to decline with age. This finding suggests that income-related interventions are likely to be most effective among the youngest whites. For blacks, however, AFQT, education, and income all have an

mother's spouse or partner. The conclusions we draw from Table 1 are not affected.

¹⁶Males tend to score lower on the PIATs. Tests administered in later years also tend to be lower which we interpret as reflecting differences in the quality of the interview or test environment. (See Baker and Mott, 1994).

independent effect on test scores and the role of income tends to be greatest among older children suggesting that there may be scope for effective interventions at least until early adolescence.

III. Do Differences in Child Test Scores Matter?

A case can be made that we should only care about test scores if they can be shown to impact outcomes that are demonstrably linked to a child's social and economic success. In this section we address two questions: First, conditional on AFQT and other measures of family background, do child test scores affect child outcomes? Second, conditional on the child's own test score, does family background as measured by AFQT, income, and maternal education matter? Ideally, we would like to examine the relationship between child test scores and measures of well-being as an adult. However, the oldest children of the NLSCM are adolescents and so we focus on grade repetition among children ten and older. Academic performance in early grades has been shown to be a significant predictor of eventual high school completion.¹⁷ The link between high school completion and future wages is also well established.

In order to investigate the relationship between child test scores and grade repetition, we estimate linear probability models in which the probability of having repeated a grade as of the last time the child was surveyed is a function of each test score. In order to minimize the probability a child is tested after repeating a grade, we use the first test for each child. The results are shown in Table 2. The first panel shows estimates from models in which the probability of grade repetition depends on the child's PPVT z-score. In addition to PPVT, we control for the child's age and gender;¹⁸ maternal AFQT; income and education; the difference between the current age

¹⁷ For example, Grissom and Shepard (1989) found in a study of over 140,000 students from school districts in three different states that students who were retained in grade were 30% more likely to drop out of school, even when achievement on standardized tests, socio-economic status, gender, and ethnicity were accounted for.

¹⁸ The estimated effects of these covariates are consistent with those of earlier studies (Barrington and Hendricks, 1989; Cairns, Cairns, and Neckerman, 1989; Grissom and Shepard, 1989; Shepard and Smith, 1990; Ensminger and Slusarcick, 1992).

and the age at which the test was taken; and the survey year. Thus, we control for the age at which the test was taken, the child's cohort, and for any differences between the waves of the survey.¹⁹

Models estimated using Ordinary Least Squares (OLS) are shown in columns 1 through 3.²⁰ These estimates may be biased by the omission of variables that are correlated both with grade repetition and with test scores, but not captured by the measures of maternal background we have included in the model. For example, some mothers may provide more supportive and stimulating environments for their children than others with similar income, education, or test scores. In an attempt to control for these factors, models that include a fixed effect for each mother are reported in columns 4 to 6 of the table. In these models, the effect of test scores on grade repetition is identified using within-family differences in test scores between siblings. Under the assumption that all unobserved heterogeneity is fixed within families, these are consistent estimates of the impact of test scores on grade repetition.²¹ In order to facilitate comparisons between the OLS and fixed effects models, we restrict attention to the subset of children with a sibling. OLS estimates based on the full sample of children yield the same conclusions.²²

The regressions are estimated separately for blacks (column 1) and whites (column 2) and the differences between the estimates are reported in column 3. Results in the first two columns show that a one standard deviation increase in PPVT scores would be associated with between a 5.5% and 7.5% decline in the probability of grade repetition among white children and black children. Although the effect for black children is 2% higher than for whites, the difference is not statistically significant. Including fixed effects reduces the size of the coefficients on PPVT for

¹⁹To capture differences in quality of the child assessment, dummies were included equal to one if the test was administered in 1988, 1990, and 1992.

²⁰Estimates based on logits and probits yield very similar results.

²¹ If for example, a mother favored one child, and this affected test scores, then the fixed effects assumption would be violated.

²²The sample is restricted to children at least 10 years of age with at least one sibling who is also 10 or older. Children of high fertility (and poor) mothers are over-represented.

blacks and increases it for whites. PPVT remains a statistically significant determinant of grade repetition and the gap between blacks and whites is small and statistically insignificant.

Results for the two PIAT scores are presented in Panels 2 and 3 of Table 2. Both PIAT-READING and PIAT-MATH scores are associated with significant reductions in the probability of grade repetition. Once again, the effects are larger for blacks than whites and the difference is significant for PIAT-READING. Including fixed effects reduces the size of the white coefficients somewhat, but has little effect on the black coefficients: a standard deviation increase in PIAT-READING has twice as big an impact on blacks relative to whites. In answer to our first question, the findings indicate that, even within families, higher test scores are associated with a lower probability of grade repetition. Hence, racial gaps in test scores are a legitimate cause for concern.²³

Turning to the effects of background, Panel 1 of Table 2 shows that conditional on both the child's PPVT *and* maternal AFQT, maternal education and income have a negative effect on grade repetition among both white and black children. Alternatively, if AFQT is viewed as just another measure of family background, then Panel 1 shows that maternal background has a strong effect on grade repetition, conditional on the child's own score. Moreover, the effects are broadly similar for both black and white children. Panels 2 and 3 show that similar results hold for the PIATs, and that maternal education and income appear to be especially important determinants of PIAT-MATH scores. Apparently, grade repetition does not depend on child test scores alone but also on family background as measured by these characteristics.

²³Models which simultaneously include all three tests are reported in Appendix Table 3. In OLS regressions, all the scores are associated with significant decreases in grade repetition among whites and blacks, with PIAT-READING scores having the biggest impact. The PIAT-READING scores also have larger effects on blacks than on whites. When fixed effects are included, only the PIAT-READING and possibly PIAT-MATH scores remain statistically significant, and then only in the equation for blacks. However, none of the point estimates is statistically different from the OLS estimates.

Finally, the estimates suggest that PIAT-READING scores are a good predictor of grade repetition and they are apparently a better predictor than PPVT.²⁴ Recall, from Figure 1, that PIAT-READING is the test for which there is no evidence of a racial gap, conditional on AFQT. Moreover, a given change in this test score has a greater impact on black schooling performance than on the performance of whites. The latter finding can be interpreted in several ways.

First, it may indicate "decreasing returns" in the sense that the higher test scores are to begin with, the less effect increases in test scores are likely to have. We find no evidence in favor of this hypothesis: among blacks, the relationship between child test scores and probability of grade repetition is linear, while among whites, to the extent that there is a non-linear relationship there are "increasing returns".²⁵

Second, it may reflect differences in the types of schools attended by blacks and whites and/or policies regarding grade retention. If the racial differences can be explained by heterogeneity in homes, schools, or neighborhoods, then they should emerge when children are stratified on other indicators of background such as maternal education. The effects of the tests are remarkably similar for children whose mothers have graduated from high school and those whose mothers have not.²⁶ Differences in background do not appear to be a dominant explanation for the differences in the impact of the tests on white and black children.

²⁴Drawing comparisons among the tests is not entirely straightforward since PIAT-READING and PIAT-MATH scores are measured when the child is 6 and PPVT when the child is 3 years old. Timing cannot explain the differences between PIAT-READING and PIAT-MATH and it seems unlikely that it can explain the very large gap between PIAT-READING and PPVT. It also cannot explain the race differences.

²⁵We investigated this hypothesis with two different approaches. First, the non-parametric relationship between grade repetition and test scores is essentially linear. Second, when quadratics of test scores are added to the regressions, PPVT² has a significant *negative* effect on grade repetition of whites, implying increasing returns.

²⁶The impact (and t statistic) of test scores on grade repetition of children whose mothers graduated high school and those who did not are, for PPVT -.025 (2.3) and -.054 (1.8), for PIAT-READING -.067 (2.9) and -.066 (1.4) and for PIAT-MATH -.038 (1.9) and -.020 (0.4), respectively.

A third interpretation is that the difference reflects test bias in the sense that the scores underpredict performance of blacks in school. That is, conditional on test scores, a black child is less likely to repeat a grade than a white child.²⁷

IV. The ASVAB Sub-Tests and Child Outcomes

We have shown that although socio-economic status does affect child test scores and the probability of grade repetition, there is no doubt that AFQT has a powerful effect. As discussed above, AFQT is a non-linear combination of four of the ten tests administered to potential military recruits. In this section, we exploit the full breadth of information contained in the ASVAB scores, rather than relying on this summary statistic. Our aim is three-fold. First, we wish to see whether AFQT adequately summarizes all the information contained in its four components. Second, we ask whether the other six sub-tests also predict child scores. In both cases, special attention is paid to differences across age and race of the child. The third aim is to determine whether maternal income and education affect child test scores after taking into account all the information about maternal ability and achievement contained in the ASVAB sub-test scores.

Before discussing the effects of the ASVAB scores on child test scores, it is useful to take a small detour and examine the relationship between a woman's scores and her wages. Using a sample of all women in the NLSY, we regress log hourly wages (in 1992) on the ASVAB sub-test scores, dummies for each year of schooling from 10 through 17 (and above), and age controls.²⁸

Among white women, only one of the four components of AFQT, "Math Knowledge", is associated with a higher wage. Two tests that are not included in the AFQT also have an important positive influence on their wages: "Numerical Operations", and "Coding Speed", a timed

²⁷As discussed above, precisely the same issue arises in the interpretation of racial differences in AFQT: given AFQT, black women tend to perform better (earn higher wages) than whites.

²⁸In the absence of good instruments to control for selection into the labor market, we estimate wage functions that are conditional on participation. Since none of the ASVAB scores predict labor force participation, it seems unlikely that the results discussed below can be entirely attributed to selection.

accuracy test. All three tests are related to quantitative skills and the magnitudes of their effects are very similar (a standard deviation increase in each score is associated with between 7 and 8% increase in wages). These effects are not only significant but also much larger in magnitude than for any of the other sub-tests.

The wages of black women apparently do not depend on the same skills. Two of the AFQT components, "Paragraph Comprehension" and "Arithmetic Reasoning" are associated with higher wages (a standard deviation increase in the score yields between 12 and 17% higher wages). In stark contrast with wages of white women, "Math Knowledge" (another AFQT element) is associated with significantly *lower* wages. A third sub-test, "Coding Speed", has a significant positive effect on the wages of black women (which is about the half the magnitude of the other two significant tests).

Table 3 reports regressions similar to those in Table 1 except that they include all 10 ASVAB scores rather than just AFQT (and we use wider child age bands). As in Table 1, the estimated effects of the ASVAB scores should be interpreted as their effect after controlling for maternal education and household income. In each panel, estimates for whites are reported in the first row, for blacks in the second row and significant racial differences are denoted by a dagger next to the latter coefficient estimate.

We begin with the four tests that comprise the AFQT. An immediately striking result in the table is that higher maternal scores on the "Word Knowledge" test are associated with better child scores although "Word Knowledge" is unrelated to wages. Among whites, "Word Knowledge" is a powerful predictor of all three tests, especially PPVT and PIAT-READING. In contrast, maternal "Paragraph Comprehension" scores have small (and negative) effects on white children. "Arithmetic Reasoning" and "Mathematics Knowledge" are the other two tests that make up the AFQT aggregate. They have no effect on PPVT scores and a marginal positive impact on PIAT-READING scores of very young children, which declines with age. However, the impact of the quantitative tests on PIAT-MATH scores increases with age, becoming large and significant

among older white children, particularly for "Mathematics Knowledge". This age profile is the reverse of that for "Word Knowledge" which rises dramatically with age for PIAT-READING, but falls with age for PIAT-MATH. This distinction between the mother's verbal and quantitative skills apparently becomes increasingly important as the child ages. It is worth noting that "Math Knowledge", the only element of the AFQT that influences wages of white women, has no impact on PPVT (which, recall, Herrnstein and Murray call "IQ") nor PIAT-READING, but does it influence PIAT-MATH scores.

The results for blacks are quite different. Like whites, "Word Knowledge" scores have a positive influence on PPVT, but the effects on the PIAT's are much smaller; among 3 to 6 year old blacks, the effects are actually negative for PIAT-MATH. Maternal "Paragraph Comprehension" scores have no influence on white scores but do have positive effects on black test scores and these are significant in more than half the regressions. Moreover, "Word Knowledge" effects are bigger on white children for all tests but the effects of "Paragraph Comprehension" are bigger for blacks (with the differences being significant in several cases). Recall that better "Paragraph Comprehension" reaps a reward in the labor market for black women whereas "Word Knowledge" does not. Higher "Arithmetic Reasoning" scores are also associated with higher wages of black women but this test turns out to have a significant *negative* effect on the PPVT scores of their children. In fact, the two quantitative tests have a significant positive influence on the scores of black children in only one case: the PIAT-MATH scores of black 7 to 10 year olds when both tests are taken together (the p-value on the joint test is 0.02).

There is tremendous heterogeneity in the influence of even the four sub-tests that make up the AFQT. These effects differ dramatically by race and vary across the child's life course. Perhaps most remarkably, there seems to be little relationship between those skills that are valued in the labor market and those associated with cognitive achievement of children. This may be because mother's labor market returns are captured by household permanent income in these regressions.

However, if income is excluded from the regressions, it turns out that none of the discussion above is amended (although the magnitudes of estimates do change slightly).

Turning to the non-AFQT tests, recall that white women who score better on "Numerical Operations" and "Coding Speed" earn higher wages. Whereas "Coding Speed" has little impact on their children (except the PIAT-MATH scores of 11-15 year olds), better maternal performance in "Numerical Operations" tends to be associated with higher scores, particularly PIAT-MATH, with the impact declining as the child ages. This is the only case in which controlling for income makes a difference: when income is dropped from the regression, "Numerical Operations" scores are associated with higher PIAT-READING scores of young white children. "Coding Speed" has a positive impact on both the wages of black women and the verbal scores of their children, particularly at older ages. "Numerical Operations", which are not rewarded in the labor market, positively influence the PIAT-MATH scores of black children and, again, this effect increases with the child's age.²⁹

The three vocational tests and general science tests are of special interest since they are unlikely to be good measures of native "intelligence" conditional on the other tests, education, and income. We interpret them, instead, as measures of skills acquired at school or work. While they are not associated with wages, they do affect child scores, particularly (but not exclusively) the PPVT scores of whites and, to a less extent, blacks.³⁰

Since AFQT is a non-linear transform of four of the ASVAB sub-tests, it is straightforward to test whether each ASVAB contains additional information over and above AFQT by also including AFQT in the regressions. The ASVABs continue to have an independent effect. Many of the individual scores are statistically significant predictors of child scores, after controlling for

²⁹It has been suggested that "Coding Speed" and "Numerical Operations" should be combined into a single sub-aggregate (Bock and Moore, 1986); however, that restriction is rejected by the data in three of the PIAT regressions.

³⁰There is one anomaly: PIAT scores of white children are lower if their mothers do better on the "Auto and Shop Information" test.

AFQT and, in half of the regressions, the four ASVABs that make up AFQT are jointly significant. In view of this fact, it is clear that the AFQT aggregate is only one of a number of different possible measures of "intelligence".³¹

Two facts emerge from this analysis. First, skills that are valued in the labor market are not necessarily the most valuable skills at home (at least in terms of a child's development). Among white women, the only test scores that has a significant positive impact both on wages and on child scores is the Mathematics Knowledge test. Among black women, "Paragraph Comprehension" and "Coding Speed" scores tests positively affect wages and child test outcomes. However, "Arithmetic Reasoning" has a positive effect on wages of black women but a *negative* effect on their children's scores.

Second, the AFQT does not summarize all the information contained in the ASVAB scores. In fact, taken together, the diversity of these results by age and race suggest that the relationships between different types of maternal achievement and child outcomes is complex, that different tests contain different information about maternal skills, and that there are dramatic differences in the information contained in each test depending on whether the mother is black or white. It is very difficult to see why AFQT should be viewed as "intelligence" (as argued by Herrnstein and Murray). Rather, instead of focusing on the similarities between the tests (by looking at all of them as manifestations of "g"), it may be more useful to exploit the differences between them to better understand what aspects of achievement they measure, and how different types of maternal achievement are related to child outcomes.

Finally, returning to the question of whether maternal education and income matter when a rich array of maternal test scores are controlled, Table 3 shows that maternal education and income continue to have a significant influence on child test scores. Moreover, this influence persists well into adolescence. As reported in Table 1, permanent income affects the scores of all

³¹It is possible that by including all ten ASVAB subtests, we are introducing multicollinearity in the regressions. We have re-estimated the models including only the four ASVAB subtests used to construct the AFQT and find that this is not the case: all the main conclusions are unchanged.

black and white children with the income effects possibly increasing with age among blacks. Black children also score better on all tests if their mothers are better educated.

V. Summary and Conclusions

We have shown that racial gaps are wider for PPVT than for either PIAT-MATH or PIAT-READING scores: focusing on PPVT exaggerates the extent of racial differences. It is clear that AFQT has an important influence on PPVT and also on the PIATs. Nevertheless, even conditional on maternal AFQT, there are significant gaps in the scores of black and white children and those gaps cannot be entirely explained by random measurement error. We have also shown that these gaps matter since all three test scores are correlated with schooling attainment and even within families, children with higher scores are less likely to repeat grades.

However, in contrast with Herrnstein and Murray's finding, we show that maternal AFQT is not all that matters. Specifically, maternal education and income are important determinants of child test scores, even after controlling for AFQT. Moreover, the highest grade of the mother has a larger effect on black children than on whites suggesting that increases in schooling attainment would tend to narrow racial differences in children's test scores. Similarly, conditional on child test scores *and* maternal AFQT, we find that maternal education and income affect a child's probability of grade repetition. Hence, even if it were correct to interpret test scores as valid measures of "nature", one would have to conclude that both nature *and* nurture matter.

Finally, we show that maternal scores on different types of achievement tests have very different effects on the scores of black and white children, and that these effects vary with the children's age. This finding suggests that the tests are not all more-or-less noisy measures of "general intelligence", but that they actually capture different aspects of maternal achievement and skill acquisition. Understanding the relationships between these skills and children's cognitive attainment may help us to better understand the complex process through which poverty is transmitted from one generation to another.

References

- Baker, Paula and Frank Mott. NLSY Child Handbook, 1989, Center for Human Resource Research, The Ohio State University, Columbus, Ohio, June 1989.
- Baker, Paula, Canada Keck, Frank Mott and Stephen Quinlan. NLSY Child Handbook: A Guide to the 1986-1990 NLSY Child Data, The Ohio State University, Columbus, Ohio, 1993.
- Bock, R. Darrell and Elsie E. J. Moore. Advantage and Disadvantage: A Profile of American Youth, Hillsdale, NJ: L. Erlbaum Associates, 1986.
- Cairns, Robert, Beverly Cairns and Holly Neckerman. "Early School Dropout: Configurations and Determinants", Child Development, 60:1437-52, 1989.
- Cleveland, W.S. "Robust Locally-Weighted Regressions and Smoothing Scatterplots," Journal of the American Statistical Association, 74:829-836, 1979.
- Currie, Janet and Duncan Thomas. "Does Head Start Make a Difference?" American Economic Review, June 1995.
- Ensminger, Margaret and Anita Slusarcick. "Paths to High School Graduation or Dropout: A Longitudinal Study of a First Grade Cohort", Sociology of Education, 65, 95-113, April 1992.
- Geronimus, Arline, John Bound, and Lisa Neidert. "On the Validity of Using Census Geocode Characteristics to Proxy Socioeconomic Status", University of Michigan mimeographed, February 1994.
- Goldberger, Arthur and Charles Manski. "Review Article: The Bell Curve by Herrnstein and Murray", The Journal of Economic Literature, vxxxiii #2, June 1995.
- Grissom, James and Lorrie Shepard. "Repeating and Dropping out of School", in Flunking Grades: Research and Policies on Retention, Lorrie Shepard and Mary Smith (eds.), London: The Falmer Press, 1989.
- Hardle, Wolfgang. Applied Nonparametric Regression, New York: Cambridge University Press, 1990.
- Heckman, James. "A Review of The Bell Curve: Intelligence and Class Structure in American Life", Reason, Los Angeles, CA, March 1995.
- Herrnstein, Richard and Charles Murray. The Bell Curve: Intelligence and Class Structure in American Life, New York: The Free Press, 1994.
- Hunt, Earl. "The Role of Intelligence in Modern Society", American Scientist, July/August, 355-368, 1995.
- Korenman, Sanford, Jane Miller, and John Sjaastad. "Long-term Poverty and Child Development in the United States: Results from the NLSY", Child and Youth Services Review, v17 #1/2, 127-155, 1995.

- Korenman, Sanford and Christopher Winship. "A Reanalysis of *The Bell Curve*", NBER Working Paper # 5230, September 1995.
- Neal, Derek and William Johnson. "The Role of Pre-Market Factors in Black-White Wage Differences", NBER Working Paper #5124, May 1995.
- Rodgers, William and William Spriggs. "What Does the AFQT Really Measure: Race, Wages, Schooling and the AFQT Score", mimeo, March, 1995.
- Rosenzweig, Mark and Ken Wolpin. "Are There Increasing Returns to the Intergenerational Production of Human Capital?", The Journal of Human Resources, 29 #2, 670-693, Spring 1994.
- Shepard, Lorrie and Mary Smith. "Synthesis of Research on Grade Retention", in Educational Leadership, May 1990, 84-88.
- Solon, Gary. "Intergenerational Income Mobility in the United States", American Economic Review, June 1992, 393-408.
- Wigdor, Alexandra and Bert Green (eds.) Performance Assessment for the Workplace, vol 1., Washington D.C.: National Academy Press, 1991.
- Zimmerman, David. "Regression Toward Mediocrity in Economic Stature", American Economic Review, June 1992, 409-429.

TABLE 1: Determinants of child test z-scores by age

Age of child:		3-4	5-6	7-8	9-10	11-13	14-15
PPVT							
AFQT	White	0.523 [8.51]	0.449 [5.88]	0.476 [5.16]	0.513 [6.60]	0.574 [6.12]	0.823 [4.84]
	Black	0.367 [4.45]	0.326 [3.59]	0.539 [4.94]	0.672 [7.13]	0.580 [5.14]	0.486 [3.64]
	Difference	-0.157 [1.53]	-0.123 [1.03]	0.063 [0.44]	0.159 [1.30]	0.006 [0.04]	-0.337 [1.60]
Mother's education	White	0.076 [2.99]	-0.003 [0.08]	0.064 [1.57]	0.084 [2.22]	-0.004 [0.10]	0.021 [0.27]
	Black	0.139 [4.08]	0.186 [4.87]	0.181 [3.87]	0.136 [3.43]	0.129 [2.84]	0.170 [3.07]
	Difference	0.063 [1.48]	0.188 [3.59]	0.117 [1.90]	0.052 [0.95]	0.133 [2.05]	0.149 [1.63]
ln(Perm income)	White	0.456 [4.96]	0.383 [3.15]	0.403 [2.90]	0.443 [3.55]	0.408 [2.59]	0.167 [0.53]
	Black	0.340 [3.49]	0.256 [2.21]	0.261 [1.94]	0.342 [2.91]	0.503 [3.59]	0.496 [3.07]
	Difference	-0.116 [0.87]	-0.127 [0.75]	-0.143 [0.74]	-0.101 [0.58]	0.095 [0.44]	0.329 [0.98]
PIAT-READ							
AFQT	White	.	0.352 [7.55]	0.348 [6.53]	0.429 [6.36]	0.508 [5.73]	0.581 [5.00]
	Black	.	0.279 [4.87]	0.437 [7.22]	0.466 [6.16]	0.431 [4.97]	0.326 [3.27]
	Difference	.	-0.073 [0.99]	0.089 [1.10]	0.037 [0.36]	-0.077 [0.62]	-0.255 [1.69]
Mother's education	White	.	0.025 [1.15]	0.031 [1.24]	0.045 [1.40]	0.052 [1.26]	0.061 [1.16]
	Black	.	0.120 [5.00]	0.092 [3.60]	0.075 [2.43]	0.107 [3.04]	0.154 [3.89]
	Difference	.	0.096 [2.95]	0.061 [1.72]	0.030 [0.68]	0.054 [1.00]	0.093 [1.43]
ln(Perm income)	White	.	0.520 [7.06]	0.430 [5.18]	0.365 [3.30]	0.265 [1.77]	0.260 [1.24]
	Black	.	0.268 [3.81]	0.262 [3.53]	0.257 [2.71]	0.244 [2.26]	0.447 [3.76]
	Difference	.	-0.251 [2.47]	-0.169 [1.52]	-0.107 [0.73]	-0.021 [0.12]	0.187 [0.79]
PIAT-MATH							
AFQT	White	.	0.371 [7.53]	0.288 [6.06]	0.358 [5.72]	0.356 [4.71]	0.396 [3.85]
	Black	.	0.177 [2.81]	0.370 [6.53]	0.443 [6.37]	0.333 [4.38]	0.286 [3.47]
	Difference	.	-0.193 [2.42]	0.082 [1.12]	0.085 [0.91]	-0.023 [0.21]	-0.111 [0.86]
Mother's education	White	.	0.021 [0.92]	0.036 [1.65]	0.010 [0.35]	0.038 [1.07]	0.004 [0.08]
	Black	.	0.104 [3.91]	0.069 [2.86]	0.021 [0.74]	0.108 [3.56]	0.053 [1.62]
	Difference	.	0.083 [2.38]	0.033 [1.00]	0.011 [0.26]	0.071 [1.51]	0.049 [0.89]
ln(Perm income)	White	.	0.274 [3.56]	0.381 [5.16]	0.438 [4.30]	0.326 [2.54]	0.154 [0.84]
	Black	.	0.260 [3.34]	0.173 [2.49]	0.215 [2.46]	0.217 [2.31]	0.377 [3.83]
	Difference	.	-0.014 [0.13]	-0.208 [2.05]	-0.223 [1.66]	-0.109 [0.69]	0.223 [1.12]

Notes: Regressions include controls for child's age, gender and for survey year test was administered.

TABLE 2: Test scores and grade repetition

	Black (1)	OLS White (2)	Difference (3)	Fixed effects Black (4)	White (5)	Difference (6)
<u>PANEL 1</u>						
PPVT	-0.075 [7.34]	-0.056 [5.41]	-0.019 [1.32]	-0.057 [2.74]	-0.044 [2.02]	-0.013 [0.44]
AFQT	-0.027 [1.07]	-0.043 [2.03]	0.016 [0.49]	.	.	.
Mother's education	-0.017 [1.75]	-0.018 [1.97]	0.001 [0.05]	.	.	.
$\ln(\text{Perm income})$	-0.044 [1.47]	-0.057 [1.66]	0.013 [0.27]	.	.	.
(1) Male	0.051 [1.76]	0.116 [4.11]	-0.094 [2.35]	0.070 [1.49]	0.122 [2.44]	-0.052 [0.74]
Age	0.049 [4.68]	0.032 [2.88]	-0.002 [0.11]	0.048 [3.55]	0.047 [3.12]	0.001 [0.05]
$\chi^2(\text{background})$	4.31 [0.01]	7.17 [0.00]	0.18 [0.91]	.	.	.
$\chi^2(\text{education \& income})$	3.15 [0.04]	3.91 [0.02]	5.49 [0.00]	.	.	.
R^2	19.28 [0.00]	15.70 [0.00]	18.07 [0.00]	1.52 [0.00]	1.39 [0.00]	1.48 [0.00]
	0.16	0.17	0.18	0.75	0.77	0.76
<u>PANEL 2</u>						
Piat-Read	-0.138 [10.98]	-0.091 [7.52]	-0.048 [2.70]	-0.129 [6.15]	-0.053 [2.15]	-0.075 [2.29]
AFQT	-0.016 [0.69]	-0.051 [2.53]	0.035 [1.12]	.	.	.
Mother's educ	-0.014 [1.45]	-0.017 [1.87]	0.003 [0.24]	.	.	.
$\ln(\text{Perm income})$	-0.047 [1.65]	-0.048 [1.42]	0.001 [0.02]	.	.	.
(1) Male	0.013 [0.47]	0.107 [3.85]	-0.080 [1.95]	0.043 [0.95]	0.104 [2.10]	-0.061 [0.91]
Age	0.029 [2.85]	0.030 [2.80]	0.010 [0.68]	0.029 [2.18]	0.049 [3.17]	-0.020 [0.99]
$\chi^2(\text{background})$	3.56 [0.01]	8.23 [0.00]	0.67 [0.57]	.	.	.
$\chi^2(\text{education \& income})$	2.93 [0.05]	3.23 [0.04]	5.57 [0.00]	.	.	.
R^2	0.21	0.20	0.22	0.77	0.78	0.78
<u>PANEL 3</u>						
Piat-Math	-0.088 [6.84]	-0.075 [6.27]	-0.013 [0.71]	-0.085 [3.96]	-0.054 [2.30]	-0.031 [0.97]
AFQT	-0.043 [1.76]	-0.054 [2.63]	0.011 [0.35]	.	.	.
Mother's educ	-0.019 [1.94]	-0.021 [2.27]	0.002 [0.12]	.	.	.
$\ln(\text{Perm income})$	-0.057 [1.90]	-0.068 [1.99]	0.011 [0.24]	.	.	.
(1) Male	0.036 [1.24]	0.116 [4.14]	-0.065 [1.58]	0.064 [1.38]	0.123 [2.48]	-0.059 [0.86]
Age	0.047 [4.52]	0.036 [3.32]	0.016 [1.03]	0.034 [2.49]	0.048 [3.08]	-0.014 [0.65]
$\chi^2(\text{background})$	8.01 [0.00]	11.61 [0.00]	0.13 [0.00]	.	.	.
$\chi^2(\text{education \& income})$	4.51 [0.01]	5.43 [0.00]	9.43 [0.00]	.	.	.
R^2	0.15	0.18	0.18	0.75	0.78	0.77
Sample size	920	754	1674	921	754	1675

TABLE 3: Maternal ASVAB Scores and Child Test Scores

		PPVT			PIAT-READING			PIAT-MATHEMATICS		
		3-6	7-10	11-15	3-6	7-10	11-15	3-6	7-10	11-15
Word Knowledge	White	0.261 [3.71]	0.361 [4.30]	0.303 [2.43]	0.148 [2.15]	0.322 [5.49]	0.547 [5.22]	0.217 [3.00]	0.102 [1.91]	0.145 [1.58]
	Black	0.114 [1.62]	0.296 [3.46]	0.503 [4.83]	0.043 [0.67]	0.129 [†] [2.28]	0.174 [†] [2.21]	-0.143 [†] [2.00]	0.046 [0.86]	-0.054 [0.81]
Paragraph Comprehension	White	0.019 [0.33]	-0.062 [0.87]	-0.133 [1.32]	-0.055 [0.98]	-0.060 [1.20]	-0.106 [1.21]	-0.077 [1.31]	-0.054 [1.19]	-0.102 [1.33]
	Black	0.131 [2.16]	0.061 [0.86]	-0.025 [0.30]	0.073 [1.30]	0.163 [†] [3.49]	0.124 [†] [1.94]	0.140 [†] [2.28]	0.084 [†] [1.93]	0.019 [0.35]
Arithmetic Reasoning	White	0.016 [0.24]	-0.012 [0.14]	0.093 [0.77]	0.127 [1.93]	0.015 [0.25]	-0.004 [0.04]	0.046 [0.66]	0.071 [1.35]	0.110 [1.23]
	Black	-0.194 [2.23]	-0.082 [0.78]	-0.368 [†] [3.05]	0.104 [1.29]	0.066 [0.98]	-0.141 [1.53]	0.097 [1.09]	0.107 [1.69]	0.014 [0.18]
Math Knowledge	White	0.007 [0.10]	-0.093 [1.12]	0.185 [1.38]	0.077 [1.16]	0.067 [1.14]	0.125 [1.10]	-0.027 [0.39]	0.094 [1.74]	0.236 [2.36]
	Black	-0.009 [0.10]	0.027 [0.27]	-0.031 [0.25]	-0.004 [0.05]	-0.078 [1.14]	-0.110 [1.16]	-0.028 [0.32]	0.094 [1.48]	0.117 [1.45]
Numerical Operations	White	0.076 [1.38]	0.054 [0.82]	0.007 [0.07]	0.075 [1.36]	0.168 [3.59]	0.039 [0.47]	0.202 [3.55]	0.078 [1.83]	-0.026 [0.36]
	Black	0.030 [0.58]	0.061 [1.02]	0.079 [1.14]	0.048 [1.01]	0.059 [1.52]	0.051 [0.94]	0.065 [1.24]	0.119 [3.29]	0.151 [†] [3.28]
Coding Speed	White	0.036 [0.76]	0.057 [0.99]	0.115 [1.34]	0.019 [0.39]	-0.025 [0.61]	0.085 [1.16]	-0.004 [0.08]	0.055 [1.48]	0.169 [2.65]
	Black	0.071 [1.39]	0.048 [0.82]	0.148 [2.09]	0.041 [0.88]	0.105 [†] [2.75]	0.093 [1.71]	0.019 [0.36]	-0.002 [0.05]	0.009 [†] [0.20]
General Science	White	0.161 [2.40]	0.083 [0.98]	0.046 [0.38]	0.058 [0.88]	0.047 [0.79]	-0.011 [0.11]	0.064 [0.92]	0.079 [1.47]	-0.155 [1.75]
	Black	-0.007 [0.09]	0.046 [0.52]	0.066 [0.61]	-0.062 [0.90]	-0.014 [0.24]	0.047 [0.58]	0.074 [0.98]	0.055 [1.01]	0.160 [†] [2.26]
Auto & Shop Information	White	0.037 [0.55]	0.042 [0.48]	-0.061 [0.50]	-0.050 [0.73]	-0.147 [2.44]	0.002 [0.02]	-0.150 [2.11]	-0.092 [1.66]	0.054 [0.60]
	Black	0.157 [1.80]	0.150 [1.43]	-0.029 [0.23]	0.062 [0.76]	-0.115 [1.64]	0.115 [1.17]	0.081 [†] [0.91]	0.063 [0.96]	0.040 [0.48]
Mechanical Comprehension	White	0.027 [0.45]	0.232 [3.14]	0.086 [0.76]	-0.005 [0.09]	0.042 [0.80]	-0.009 [0.09]	0.099 [1.57]	0.098 [2.04]	0.035 [0.42]
	Black	0.068 [0.87]	0.044 [0.47]	0.030 [0.27]	0.070 [0.97]	0.086 [1.39]	-0.012 [0.14]	0.017 [0.21]	-0.086 [†] [1.49]	0.005 [0.07]
Electronics Information	White	-0.001 [0.01]	-0.011 [0.15]	0.218 [2.12]	0.011 [0.20]	-0.036 [0.71]	-0.095 [1.08]	0.091 [1.53]	-0.040 [0.85]	0.051 [0.67]
	Black	0.136 [1.92]	0.097 [1.20]	0.028 [0.29]	0.073 [1.13]	0.080 [1.48]	0.036 [0.49]	0.019 [0.28]	0.019 [0.37]	0.015 [0.24]
Mother's education	White	0.049 [2.32]	0.092 [3.31]	0.015 [0.39]	0.020 [0.91]	0.034 [1.74]	0.050 [1.52]	0.018 [0.78]	0.019 [1.07]	0.029 [1.01]
	Black	0.162 [†] [5.99]	0.141 [4.50]	0.127 [†] [3.53]	0.117 [†] [4.67]	0.068 [3.35]	0.104 [3.83]	0.097 [†] [3.51]	0.034 [1.79]	0.062 [†] [2.69]
ln(Perm income)	White	0.420 [5.57]	0.415 [4.41]	0.336 [2.32]	0.504 [6.68]	0.366 [5.45]	0.229 [1.87]	0.236 [3.02]	0.382 [6.26]	0.257 [2.42]
	Black	0.267 [3.46]	0.292 [3.29]	0.407 [3.82]	0.254 [†] [3.56]	0.228 [3.87]	0.290 [3.60]	0.237 [3.02]	0.185 [†] [3.35]	0.296 [4.30]

Notes: † denotes significant difference between effect on blacks and whites.

Appendix Table 1
Means and [standard errors] of maternal characteristics

	All	Black	White	Difference
Mother education	12.54 [0.04]	12.42 [0.06]	12.63 [0.06]	-0.21 [0.08]
Permanent income (thousands \$)	26.70 [0.33]	19.42 [0.39]	32.50 [0.43]	-13.08 [0.60]
ln(perm income)	10.03 [0.01]	9.71 [0.02]	10.28 [0.01]	-0.58 [0.02]
AFQT Percentile	35.21 [0.56]	20.02 [0.57]	47.35 [0.74]	-27.33 [0.97]
AFQT z-score	-0.45 [0.02]	-1.02 [0.02]	0.00 [0.02]	-1.02 [0.03]
ASVAB (z scores)				
Word knowledge	-0.42 [0.02]	-1.04 [0.03]	0.07 [0.02]	-1.11 [0.04]
Paragraph comprehension	-0.27 [0.02]	-0.83 [0.03]	0.17 [0.02]	-1.00 [0.04]
Arithmetic reasoning	-0.48 [0.02]	-0.95 [0.02]	-0.11 [0.03]	-0.85 [0.03]
Math knowledge	-0.39 [0.02]	-0.75 [0.02]	-0.11 [0.03]	-0.64 [0.03]
Numerical operations	-0.22 [0.02]	-0.69 [0.03]	0.16 [0.03]	-0.86 [0.04]
Coding speed	-0.09 [0.02]	-0.60 [0.03]	0.32 [0.03]	-0.91 [0.04]
General science	-0.53 [0.02]	-1.02 [0.02]	-0.13 [0.02]	-0.90 [0.03]
Auto & shop information	-0.73 [0.01]	-1.13 [0.02]	-0.41 [0.02]	-0.72 [0.02]
Mechanical comprehension	-0.64 [0.02]	-1.04 [0.02]	-0.31 [0.02]	-0.72 [0.03]
Electronic information	-0.64 [0.02]	-1.11 [0.02]	-0.27 [0.02]	-0.84 [0.03]
Sample size	2164	961	1203	.

Appendix Table 2A
Child z-scores of PPVT, PIAT-Reading and PIAT-Mathematics Tests by race and age

	PPVT			PIAT-Reading			PIAT-Mathematics		
	Black	White	Diff	Black	White	Diff	Black	White	Diff
All children	-1.98 (0.03) [3526]	-0.20 (0.02) [3942]	-1.78 (0.04)	-0.04 (0.02) [4352]	0.52 (0.02) [4266]	-0.57 (0.03)	-0.57 (0.02) [4390]	0.20 (0.02) [4317]	-0.78 (0.03)
Age (years)									
3-4	-2.48 (0.05) [951]	-0.29 (0.04) [1366]	-2.19 (0.07)						
5-6	-1.83 (0.06) [637]	-0.12 (0.05) [785]	-1.71 (0.08)	0.37 (0.04) [1138]	0.54 (0.03) [1408]	-0.16 (0.05)	-0.50 (0.04) [1172]	0.21 (0.03) [1453]	-0.71 (0.05)
7-8	-1.88 (0.07) [504]	-0.18 (0.06) [594]	-1.70 (0.10)	0.07 (0.04) [1089]	0.59 (0.04) [1196]	-0.52 (0.05)	-0.43 (0.04) [1091]	0.26 (0.03) [1194]	-0.69 (0.05)
9-10	-1.69 (0.06) [665]	-0.09 (0.06) [610]	-1.60 (0.08)	-0.14 (0.05) [926]	0.55 (0.05) [843]	-0.69 (0.07)	-0.63 (0.04) [924]	0.21 (0.04) [850]	-0.84 (0.06)
11-12	-1.79 (0.07) [495]	-0.17 (0.07) [406]	-1.62 (0.10)	-0.44 (0.05) [673]	0.44 (0.06) [508]	-0.89 (0.08)	-0.67 (0.05) [683]	0.13 (0.05) [509]	-0.80 (0.07)
13-15	-1.83 (0.09) [274]	-0.30 (0.13) [181]	-1.53 (0.15)	-0.50 (0.06) [526]	0.28 (0.09) [311]	-0.78 (0.11)	-0.82 (0.05) [520]	0.07 (0.07) [311]	-0.89 (0.09)

Notes: Diff=Black-White. Means, (standard errors) and [number of observations]

Appendix Table 2B
Probability child has ever repeated a grade by race and age

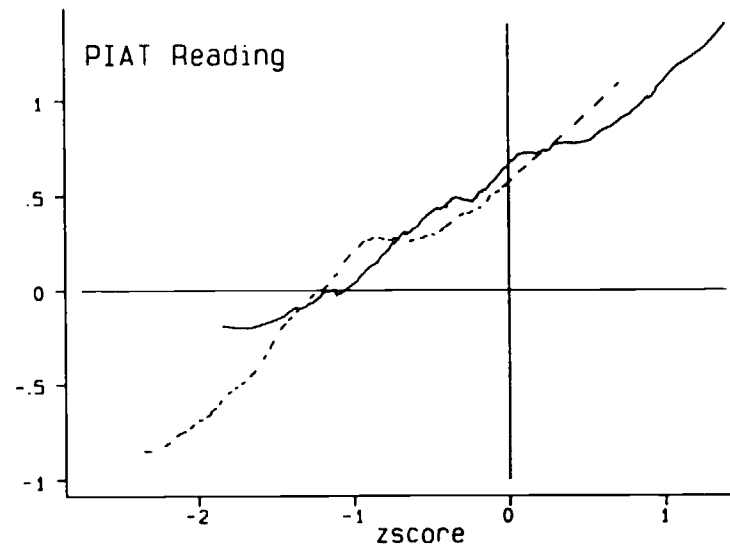
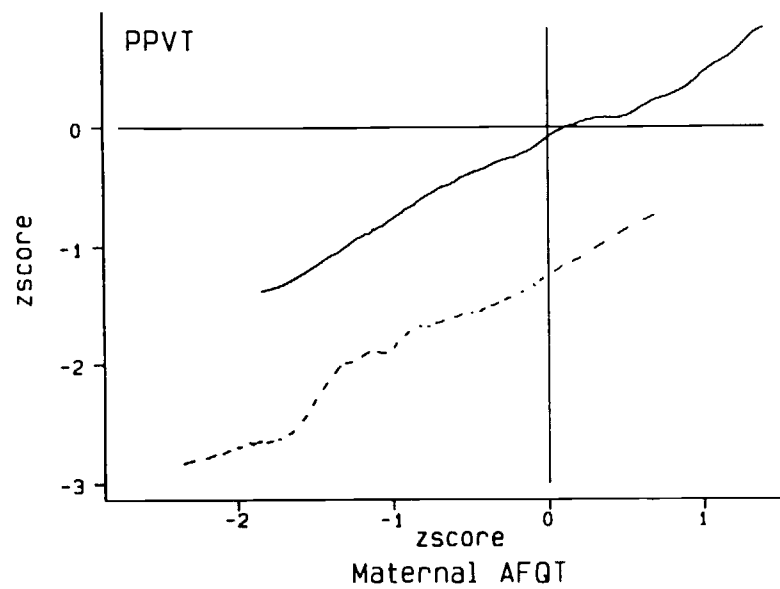
	Black	White	Difference
All children	0.34 (0.02) [926]	0.23 (0.02) [772]	0.11 (0.02) .
Age (in years)			
10	0.20 (0.03) [169]	0.13 (0.02) [189]	0.06 (0.04) .
11-12	0.29 (0.02) [344]	0.20 (0.02) [300]	0.09 (0.03) .
13-15	0.45 (0.02) [413]	0.33 (0.03) [283]	0.12 (0.04) .

Notes: Difference =Black-White. Means, (standard errors) and [number of observations]

Appendix Table 3: Test scores and grade repetition
Simultaneous impact of all three tests

	OLS			Fixed effects		
	Black (1)	White (2)	Difference (3)	Black (4)	White (5)	Difference (6)
PPVT	-0.035 [3.14]	-0.026 [2.40]	-0.009 [0.59]	-0.004 [0.20]	-0.020 [0.84]	0.015 [0.47]
Piat-Math	-0.020 [1.41]	-0.032 [2.30]	0.013 [0.64]	-0.041 [1.72]	-0.025 [0.93]	-0.016 [0.45]
Piat-Read	-0.113 [7.76]	-0.064 [4.68]	-0.039 [1.97]	-0.118 [4.82]	-0.041 [1.48]	-0.077 [2.07]
AFQT	-0.002 [0.07]	-0.034 [1.64]	0.030 [1.02]	.	.	.
Mother's educ	-0.012 [1.25]	-0.017 [1.85]	0.003 [0.25]	.	.	.
ln(Perm income)	-0.036 [1.26]	-0.038 [1.12]	-0.007 [0.41]	.	.	.
(1) Male	0.018 [0.64]	0.110 [3.97]	-0.064 [1.57]	0.046 [1.02]	0.085 [1.69]	-0.039 [0.57]
Age	0.034 [3.27]	0.028 [2.53]	0.017 [1.13]	0.025 [1.40]	0.018 [0.89]	0.007 [0.26]
Tests for joint significance						
χ^2 (all tests)	4.310 [0.01]	23.60 [0.00]	1.77 [0.15]	14.40 [0.00]	2.57 [0.06]	1.93 [0.12]
χ^2 (all tests equal)	9.44 [0.00]	2.00 [0.14]	1.25 [0.29]	4.61 [0.01]	0.14 [0.87]	1.40 [0.25]
χ^2 (all maternal background)	1.67 [0.17]	4.73 [0.00]	1.14 [0.33]	.	.	.
χ^2 (maternal education & income)	1.90 [0.15]	2.69 [0.07]	3.18 [0.04]	.	.	.
F (all covariates)	22.440 [0.00]	15.990 [0.00]	26.090 [0.00]	1.720 [0.00]	1.460 [0.00]	1.620 [0.00]
R ²	0.23	0.22	0.23	0.77	0.79	0.78
Sample size	916	754	1671	917	754	1671

Figure 1: Child test scores and maternal AFQT



White ———
Black - - - -

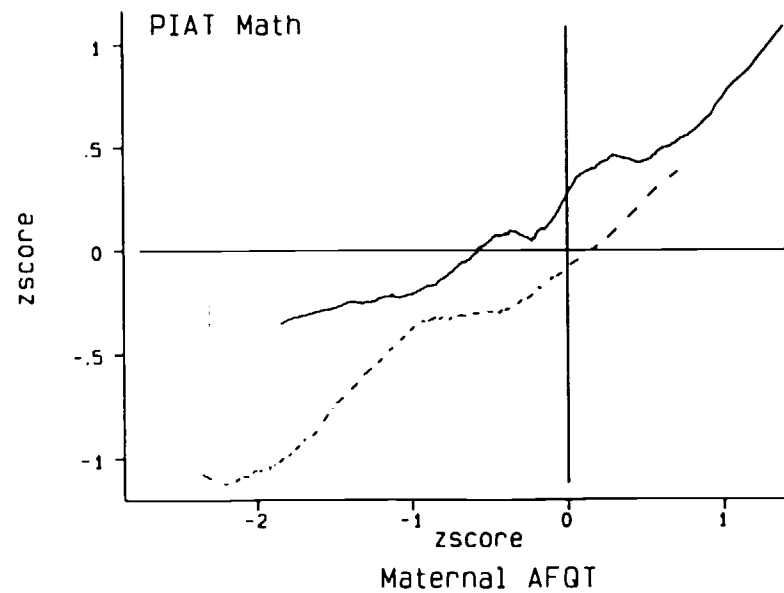
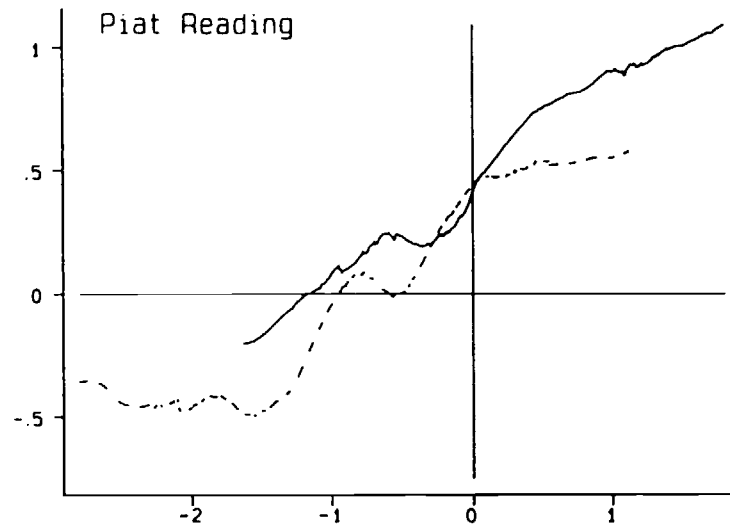
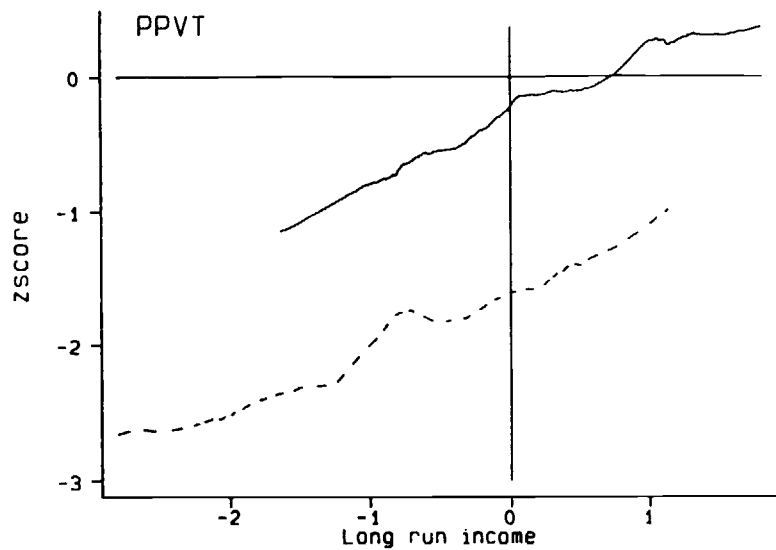
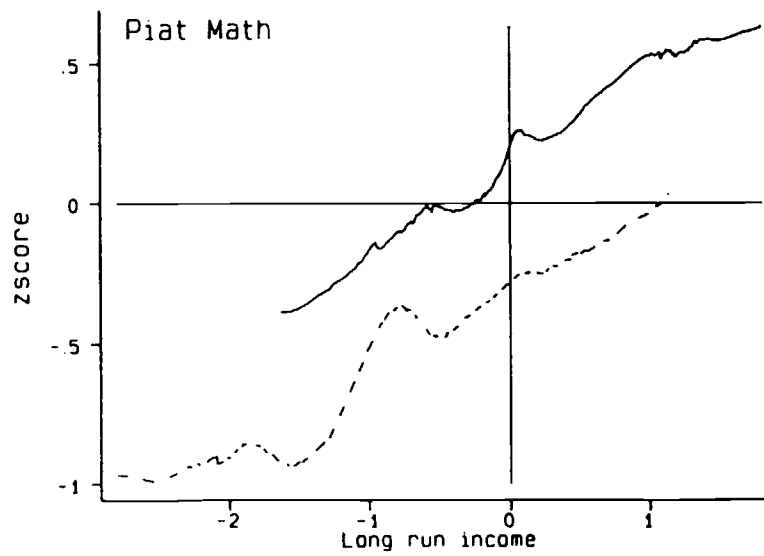


Figure 2: Child test scores and long run income



White ———
Black - - - -



Appendix Figure 1

