

October 21, 2013.

Achievement as an Educational Good

Helen F. Ladd (presenter)

Sanford Institute of Public Policy

Duke University

Durham, NC 27708

Hladd@duke.edu

This paper is a draft of chapter 4 for a book with the preliminary title: *Educational Goods: Values, Evidence and Decision Making* by Harry Brighouse, Helen F. Ladd, Susanna Loeb, and Adam Swift. It will be presented by Helen F. Ladd ([hladd@duke.edu](mailto:Hladd@duke.edu)) at the 2013 APPAM fall conference, Washington, D.C., November 7, 2013.

In Chapter 3 we defined the term “educational goods” as the knowledge, skills, dispositions and attitudes that enable people to live flourishing lives and to contribute to the flourishing of others, and clarified the normative significance of this broad definition. In practice, however, decision makers do not have data on the full set of valued educational goods. As a substitute, they often focus narrowly on student scores on standardized tests. In the U.S. context, for example, policy makers tend to emphasize tests intended to measure student achievement in reading and mathematics.

We use the term “student achievement” in this chapter to refer to the important cognitive knowledge and skills purportedly measured by the various standardized tests generally used in the U.S. to gauge student progress. These skills are at best a subset of all educational goods and, of course, even this subset is not measured perfectly by the tests that are used. Although our focus on achievement follows the tendency for policy makers and researchers to pay attention to what can be measured and analyzed with the tools of empirical social science, the reader should not interpret it as a signal that achievement is the only relevant, or even the most important, educational good. As we emphasize in this chapter, and highlight more directly in the policy chapters in the second half of this book, it is only one of many educational goods. Even as imperfectly measured by test scores, however, student achievement is useful for illustrating some of the principles that are central to our approach, especially the distributive principles.

We begin this chapter with a brief discussion of achievement as an educational good. We then discuss the level and distribution of test scores - as a measure of achievement - across individual students. This discussion allows us to illustrate quite concretely some of the normative principles we introduced in chapter 3. We follow with an analysis of achievement gaps between groups of students, with particular attention to the process of determining which gaps are normatively significant. Finally, we build the bridge from the first part of the book to the second part where we use our normative and evidence-based framework to explore a selection of policy issues.

Achievement as an educational good

The most obvious link between student achievement as measured by test scores and subsequent flourishing is through the labor market. According to the standard economic model, high achieving students bring greater knowledge and skills to the labor market than lower achieving students. Employers reward these skills with high wages and good opportunities for learning on the job and career advancement. Low skills in turn relegate a person to the low wage labor market, which, in addition to low wages, often brings with it greater difficulty finding a job and limited opportunities for advancement.

This model is supported, for example, by the finding of a clear statistical relationship between the test scores of male teenagers on a modified version of the the Armed Forces Qualifying Test (AFQT), and the subsequent wages of those same males both when they were 26-28 and in their late 30s. The AFQT is considered a fairly broad and accurate representation of “achievement and learned skills” or “cognitive skills”. The researchers find that, compared to their lower scoring counterparts, workers with higher test scores as teenagers end up in jobs with higher wages in

their 20s and the wage gap is even larger when they are in their 30s. Stated differently, higher measured achievement is predictive of higher future earnings.ⁱ

A similar relationship emerges for the test scores of the type used in many U.S. states and school districts. In a recent study of one large U.S. city, test scores of students in grades 3-8 based on administrative records were matched to their subsequent earnings (based on tax records) when they were in their 20s. Conditional on prior year test scores and other student and control variables, the authors found that a difference of one standard deviation (a common measure of variation described below) in test scores was associated with an average difference of 11.6 percent in earnings.ⁱⁱ Various earlier studies show the same patterns.

The possession of cognitive skills and knowledge makes it possible for individuals to flourish in other ways as well, both while they are in school and as adults. Such skills enrich a person's life by opening doors to science, literature, arts and music, and, importantly, by creating opportunities for higher education and the acquisition of degrees and other accomplishments. Research shows that people with more years of schooling typically receive non-pecuniary benefits connected with their jobs such as higher job satisfaction, higher occupational prestige and lower levels of unemployment. They also receive a variety of benefits outside the labor market. As adults they have better health, are less likely to get divorced or separated, have increased trust in others, are better parents, and are less likely to be arrested and to smoke. In addition, they are likely to enjoy more leisure and to have healthier families.ⁱⁱⁱ

Cognitive skills and years of schooling are beneficial not only for the individuals who have them but for others as well. Everyone benefits from having a skilled labor force that is productive and innovative. Further, those with high cognitive ability are able to generate the technological innovations that are needed for productivity growth. In addition, people with more education exhibit higher voter turnout and civic participation, thereby contributing to the strength of our democracy.^{iv}

One must be careful, however, in attributing causation to these observed statistical relationships among cognitive skills of the type that are measured by test scores, years of schooling, and other outcomes. It may be that other types of knowledge, skills and dispositions – that is, other educational goods -- are an important part of the story. The U.S. experience with the General Educational Development Program (commonly referred to as the GED) is illustrative. The GED test was initially designed as a way for high school drop outs to show they had mastered the same cognitive skills as those who graduate from high school. With that credential, the GED holders would then gain access to the jobs and higher education opportunities available to those with high school degrees. The Nobel prize winning economist, James Heckman, however, has demonstrated the fallacy of the approach. The problem is that the GED measures only the cognitive skills that can be captured by tests and fails to measure what he calls the “non-cognitive skills” that are required not only for young people to complete high school but also to succeed in the labor market or in subsequent educational endeavors.^v

Although left unspecified by Heckman, the term “non-cognitive” skills is usually understood to include characteristics such as persistence, motivation, and self-control, some of which in fact may include a cognitive component and some of which may be more like dispositions and

attitudes than skills. Other research suggests that these “non-cognitive” skills interact with cognitive skills thereby making it difficult to separate out their separate contributions to subsequent flourishing. It appears, for example, that non-cognitive skills may promote the development of cognitive skills, but the reverse is not true. Thus, the potential for cognitive skills, especially those typically measured by standardized tests, to contribute to flourishing in the ways described above may well depend on an individual possessing other educational goods as well.^{vi}

Further while cognitive achievement undoubtedly contributes to the development of some of the capacities we identified in chapter 3 as useful guides to help decision makers to determine what knowledge, skills, dispositions and attitudes contribute to flourishing, it is more important for some than for others. Recall that that these competencies include the capacity for economic productivity, for democratic competence, to regard others as moral equals, for healthy personal relationships and for personal fulfilment. Cognitive skills are particularly relevant for economic productivity, democratic competence, and some aspects of personal fulfilment. They are less central, however, for an individual to develop strong personal relationships or to develop the capacity for treating others as moral equals.

If cognitive achievement is just one of many educational goods that contribute to flourishing either by themselves or in complicated ways, why does it get so much attention in policy discussions? The answer is clear. Compared to other educational goods, it is far easier to measure. That need not mean, however, that the tests used to measure it do a good job of accounting for even the full set of cognitive skills that contribute to flourishing. For one thing, the tests used for elementary and middle school students often focus on math and reading alone, which at best represent a subset of the relevant cognitive skills. Moreover, to keep costs at a manageable level, many states rely heavily on electronically gradable multiple choice tests. Such tests have the disadvantage that they tend to focus on basic skills rather than broader conceptual, analytical and problem solving skills that may be of greater relevance for ultimate flourishing. Finally, as we shall see in chapter 6 not only may the tests measure the wrong thing but relying on them can create incentives for teachers to avoid teaching the more valuable skills that are not tested.

Achievement and normative principles

Despite these limitations, student achievement still serves as a useful vehicle for illustrating our central normative principles. The starting point is that some children achieve at higher levels than others. This variation reflects several distinct, but interrelated, factors: differences in children’s family backgrounds and access to educational opportunities; differences in characteristics like intelligence or talent; and differences in motivation.

We depict such variation in achievement levels across students by the bell shaped distribution in Figure 1a. Along the horizontal axis is student achievement, which in practice would be measured by test scores. For the purposes of this discussion, we simply assume that the test provides a good measure of each student’s knowledge and skills in subjects relevant for subsequent flourishing. The area under the curve between any two achievement levels represents the percentage of students whose achievement falls in that range, with the area under the full

curve representing 100 percent of the students. To keep things simple, we have drawn the curve as symmetric, which means the average achievement level is at the center of the distribution as marked on the figure by A bar. The bell shaped distribution implies that many students achieve at levels close to the average, with far fewer students having achievement levels in either of the tails of the distribution. In practice, the distribution need not be symmetric. If the relevant population of students included a disproportionate number of disadvantaged students who achieve at lower levels than their more advantaged counterparts, for example, the area under the left tail would be larger and the average would be further to the left. Similarly, if the relevant population included a disproportionate number of highly motivated students who achieve at high levels, the area under the right tail would be larger and the average would be further to the right. For the current illustrative discussion, however, the precise shape of the distribution is not important.

The other distributions in figure 1 highlight three concepts that are useful for normative discussions about student achievement. Figure 1b illustrates the meaning of a higher average achievement level (often called the mean) with no difference in how achievement is distributed around the average. As shown in that figure the whole distribution is simply further to the right. Figure 1c illustrates the concept of variation. With more students having very low or very high levels of achievement, the variation is clearly larger in the bolded distribution Figure 1c compared to the distribution in 1a. For bell shaped curves, empirical researchers often use the concept of a standard deviation to talk about this variation. In our example, the distribution in 1d has a larger standard deviation than the distribution in 1c. A useful rule of thumb is that for a bell shaped distribution, 68 percent of the observations fall within one standard deviation of the mean. The advantage of a specific measure of this form is that it allows both policy makers and researchers to compare the variations across distributions.

Finally figure 1d incorporates the concept of an adequate level of achievement. If the adequate level of achievement were deemed to be A_L we would see that well over half of all students are achieving at an adequate level or higher, or stated differently, that a relatively small proportion are achieving at levels that are inadequate. If the adequate level of achievement were deemed to be A_H in the figure, however, well over half the students would be achieving at inadequately low levels.

More is better

With these figures in mind, we can now turn to the normative significance of achievement and how it is distributed across students. Because achievement is an example of an educational good, and hence is something that we value, all else equal a higher average level of achievement as shown in figure 1b is unambiguously better than a lower level. That corresponds to our normative principle that more is better.

At the same time, some cautionary notes are in order. To the extent that higher average achievement is attained not by a rightward shift in the whole distribution as shown in figure 1b but rather by increases in student achievement at the top of the distribution alone with no gains at the bottom, or possibly even declines among the low achievers, policy makers would need to trade off the benefits of having more of this educational good against the costs in terms of one or

more of our distributional principles. Another tradeoff would arise if the increase in student achievement comes at the expense of other educational goods. Thus, for example, if the policy mechanisms used to raise student achievement adversely affected other dispositions and attitudes that also contribute to future flourishing, policy makers might want to weigh the loss in these other things against the increase in achievement.

Finally, we note the possible tradeoff with other independent values, two of which are potentially relevant in this context: childhood goods and parents' interests. While higher achievement for young children is desirable, so also are goods like play and naïve curiosity. Concern for healthy childhood development and happiness within childhood may sometimes argue against an excessive focus on achievement, either cognitive or other. A trade off with parents' interests arises whenever policy interventions designed to promote higher achievement interfere with parents' views about how their child should be raised. Concern for parents' interests does not mean that achievement is any less important. It simply means that any potential gains in achievement may need to be weighed against the value attached to any loss with respect to parents' interests.

Distributive principles

In chapter 3, we identified three main distributive principles: equality, adequacy, and priority to less advantaged, and emphasized that they can apply to the distribution of either educational goods or flourishing. Applying the principles to achievement requires first that we recognize that educational achievement is but one educational good that may compete with other such goods. But, further, the distribution of educational goods *as a whole* may differ from the distribution of flourishing. When talking about equality, adequacy or priority to the less advantaged, it is important to be clear which domain we have in mind.

Equality. Suppose policy makers place a high value on promoting an equal distribution of educational goods across individuals. Equality in the distribution of such goods need not require equality in the distribution of achievement because any differences in student achievement could potentially be offset by differences in other educational goods. Indeed, if achievement were perfectly inversely correlated with other educational goods, then observing achievement differences across individuals would not tell us much about the distribution of educational goods. We are doubtful that, in our society, this correlation holds. In fact, we suspect that the distinct, but interrelated, factors and processes that give rise to achievement differences – such as inequalities of family wealth, stress, levels of parental employment, quality of housing, neighborhoods, school and teacher quality, and health – simultaneously give rise to similar, though not identical inequalities in other educational goods. To the extent that is true, achievement differences would translate, albeit imperfectly, into differences in educational goods. At the same time, we note that in some situations a heavy emphasis on equalizing cognitive achievement might well interfere with the equalization of educational goods.^{vii}

Even with a focus on achievement alone, it may not be clear who should be equal to whom because of disagreement about the normative significance of the various determinants of student achievement. One determinant is variation in the natural talents that children are born with. Many people regard any differences in student achievement associated with variation of this type

as entirely acceptable. Others might see inequalities caused even by natural variation as undesirable from the point of view of equality but view them as justified overall because of other values. For example, if some measures to equalize achievement between students with very different levels of natural ability are very costly, or involve reducing the achievement levels of higher-achieving students, the 'more is better' principle could trump the goal of trying to eliminating inequalities due to differences in natural talents.^{viii}

Those who have no objection to achievement inequalities due to natural differences tend nonetheless to worry about the extent to which children's development and exercise of their innate capacities may be a function of other characteristics, such as their race, gender or family's socioeconomic position. Later in this chapter we discuss the factors that may make some gaps more normatively salient than others. Here we use the achievement gap between children from low-income and higher income families, which is widely regarded as salient, to illustrate the policy implications of a particular specification of the equality principle.

Students from higher income families achieve, on average, at higher levels than their counterparts from less economically advantaged families. The reasons for this pattern are well understood: at early ages children from the more economically advantaged families are exposed to more books and a wider vocabulary and to cognitively richer environments; middle class children are less likely to come to school with untreated health problems; middle class families are more able to work the school system to the advantage of their children through their choice of school and teacher assignment, and such families are able to provide a richer array of after schools and summer experiences..

A variety of policy interventions could potentially reduce inequalities of this type. To the extent that that current disparities in schooling are detrimental to economically disadvantaged students, one might require at a minimum that all children have access to equal quality schools and teachers so that the schools do not exacerbate the variation in achievement associated with home environments. In fact, though, full equality would require far more: the schools and classrooms serving low performing students from disadvantaged backgrounds would have to be of even higher quality than other schools and classrooms in order to compensate for the learning challenges that such children typically bring to school. In particular such schools might well need more teachers and higher quality teachers, as well as more nurses and counselors, and other adults who could connect needy children and their families to social and health services that middle class families take for granted.

Full equality would also require a variety of public interventions outside the traditional schooling sector for children from disadvantaged families. Such interventions might include, for example, early childhood and pre-school programs as well as afterschool and summer enrichment programs. In addition, the full equalization of educational goods might well require restrictions on how parents contribute to the educational goods of their children (for example, sending their children to private schools or spending money for enrichment activities) and may require other public interventions into family life. Thus, an equality standard might well require extensive intervention into the family lives of both high achieving and low achieving students, some of which would undoubtedly interfere with what we have called the independent value of parental rights.

Adequacy. Our second distributional principle, the principle of adequacy, would solve some of these conflicts. Recall that this principle requires that all students receive at least an adequate level of educational goods, with no attention to the distribution of educational goods above the threshold. Recall also that specification of such a principle involves answering the question ‘adequate for what?’ and we have suggested that this should be in terms of flourishing as an adult. So, ideally, adequacy would be specified in terms of the full bundle of educational goods, not simply in terms of student achievement as measured by test scores. In its ruling in the 1989 Rose school finance case that dealt with adequacy, the Kentucky court made an effort to include a broad set of educational goods, as have several other state courts since then. The Kentucky court defined an adequate education as one that fosters seven learning goals, including, for example, oral and written communication skills, knowledge of economic social and political systems, understanding of governmental processes, self-knowledge, and grounding in the arts, and academic or vocational skills. The intent was for each child to emerge from school ready to function fully in a complex and changing world.^{ix}

In practice, however, policy makers, especially those in the United States, often define adequacy far more narrowly in terms of achievement (as measured by test scores) alone. One justification may be the standard one, namely, that, achievement can be measured and therefore lends itself to a quantifiable standard. Another may be the belief that cognitive skills and knowledge are so central to flourishing, especially for those at risk of having very low levels of achievement, that no tradeoffs are possible with other educational goods. If children do not learn to read by third grade, for example, they are not in a position to do the reading required for learning as they progress through school. But a single minded focus on achievement may detract from other educational goods including, for example, the “noncognitive skills” referred to earlier. Especially for children at the low end of the achievement distribution, such skills could be essential for developing cognitive skills and could be more important for ultimate flourishing than test scores.

Raising the level of achievement of those at the bottom of the achievement distribution up to an adequacy standard with little or no change in what happens at the top of the distribution would concomitantly make the overall distribution of achievement more equal. How much it increases equality of flourishing will partly depend on the extent to which achievement is a positional good. Recall that a positional good is one for which the benefits depend not on how much of the good someone has but rather on his or her position in the overall distribution of the good. Thus, while the pursuit of an adequacy standard defined in terms of achievement may lead to a more equal distribution of the capacity to enjoy literature or solve math problems it may do little to equalize the capacity of people to obtain good jobs if such jobs are limited; in that case, despite the higher skills of those at the bottom of the distribution, the good jobs will continue to go to people with the relatively higher skills. Similarly, the positional aspect of education means that whether one person or group’s educational achievement is in fact adequate for them to achieve a particular level or kind of flourishing may depend on the overall distribution of achievement, not only on the absolute level of those at the bottom of the distribution.

If we set aside the positional aspects of achievement, and its implications for how to think about what counts as ‘adequate’ achievement, we can identify two main ways to meet an achievement-based adequacy standard. One is to move the whole achievement distribution so far to the right

that all children attain at least the adequate level. Although such an approach, if successful, could be appealing because it would simultaneously achieve two normative goals –adequacy and more educational goods for everyone -- the costs in the form of substantial additional investments in schooling, and hence in other consumption goods foregone, are likely to be very high. A modified version of this approach would be to intervene in ways that increase achievement at the bottom of the distribution but increase it even more at the top. An example might be television programs such as Sesame Street, the benefits of which accrue to children at all income levels, but more so to children from middle class families.^x In this case, the movement toward an adequacy standard brings with it greater inequality, thereby generating a trade- off between those two distributional principles.

An alternative more cost-effective strategy –though possibly politically difficult to implement -- would be to target resources and attention to the low end with the goal of raising the achievement of all the low-achieving students to the specified level. That strategy has the advantage of allowing policy makers to craft policies and programs specifically designed to address the learning needs of the low-achieving children and less costly in that it requires no additional investments for children already above the threshold. If the budget for educational goods was fixed, however, that strategy would require a transfer of resources from those above to those below the standard, and possibly to a reduction in the total amount of educational goods produced. As a consequence, the adequacy principle would need to be traded off against the “more is better” principle. Further, the setting of a well-defined threshold or bar for achievement raises some additional concerns in practice. Studies show that teachers may respond to specified achievement targets by focusing attention on the students closest to the threshold, ignoring both those who are far below and those who are comfortably above the standard. One might ask why children whose achievement levels are just below the standard are any more deserving of attention than those at other levels of achievement.

Priority to the less advantaged. The third distributive principle avoids the need for a threshold by asserting that distributions should improve the position of those who are worse off. In chapter 3 we advocated applying this principle to flourishing rather than to educational goods. Suppose, for a moment, though, that we could regard a person’s educational achievement as a decent proxy for her flourishing as an adult. The priority principle would then support any policy intervention that shifts the left tail of the achievement distribution to the right, even if it also reduces the total amount of achievement or increases the gap between those with lower and higher achievement levels or both.

But we should recall a point made in the previous chapter. It may be that, over time, the worst off people in the society will be better off if the society is more productive, and that greater productivity requires technological innovations that will be created only by those with high levels of cognitive ability. In that case, resources should be invested in the highly able – even if that increases inequality in achievement levels. Even though it might be possible to raise achievement at the bottom end of the distribution in the short term by putting the marginal resources there, in fact, the priority principle directs policy makers to put them at the top. So, even if achievement at the low end is not as high as it could be in the short run, the less advantaged might still experience greater flourishing – and perhaps even greater educational

achievement – in the long run. Here the idea that ‘more is good’ is combined with a concern to give distributive priority to the worst off. Inequality is permitted, perhaps even encouraged, in order to increase the size of the pie – the ‘flourishing pie’ - which in turns yields bigger pieces even to those who have least.

Achievement gaps

Achievement gaps between clearly defined groups of students – such as black and white students – receive a tremendous amount of attention in U.S. policy debates. Depending on how the gaps are measured, efforts to reduce them could be consistent with the distributive goals of either equality or adequacy. In some cases, however, what is really at stake may be the low achievement level of members of the lower performing group rather than the size of the gap between the average achievement of the lower and the higher group. In those cases, the relevant distributive criterion may be adequacy or priority to the less advantaged.

The most straightforward measure of an achievement gap is the difference between the average performance of one group of children, call them group A, and the average performance of another group, call them group B. Even if the members of group B outperform the members of group A on average -- thereby generating an achievement gap that favors group B -- some members of group A may well outperform some members of group B, because the distributions of the two groups are likely to overlap. So the focus on group averages can be misleading in so far as it attributes to the individual the average characteristics of the group.

Normatively speaking it is the wellbeing of individuals that matter – their membership in groups is relevant only in so far as the group characteristics affect them as individuals. Group characteristics may affect an individual’s wellbeing in different ways. Knowing that black students tend to achieve at lower levels on average than white students directs us towards causal factors influencing the achievement, and subsequent flourishing, of different individuals. The same applies in the case of the gap between low-income and high-income children. But, further, being, and being perceived as, a member of a group defined in, say, racial terms may itself affect an individual’s wellbeing. If it reinforces stereotypes and social expectations, the very fact that black students generally tend to achieve less highly than white students can adversely affect the lives even of those black students who perform well.

Any comparisons of this type require that achievement for the two groups be measured on the same scale. In the U.S., for example, scale scores in math and reading on the National Assessment of Educational Progress (NAEP), commonly referred to as the nation’s report card, are reported every two years for representative samples of 4th and 8th graders in every state (and now, in 20 large districts). These scale scores permit comparisons across groups of students at one point in time and over time. When test scores are not measured or reported on the same scale, the typical approach is to make the scores comparable by standardizing the units of each one to have a mean (or, equivalently, an average) of zero and a standard deviation of one, where the standard deviation is a measure of how much dispersion there is around the mean. If the distributions both have the bell shapes as shown in the figure, a difference in the means of 1 standard deviation would mean that the achievement level of the typical person in group A would be below that of about 85 percent of the members of group B.

An alternative approach to measuring achievement gaps is to focus on differences across groups in the percentages of students who meet some threshold, often referred to as proficiency. According to this measure, the extent to which some children exceed the proficiency standard or others fall short is not relevant. Although potentially appealing in the way it combines equality and adequacy considerations, this approach is often seriously flawed in practice in that the measured gap between groups is heavily dependent on the level of the proficiency standard. The higher the standard, the larger the gaps are likely to be. If the standard is set sufficiently low then the gap may disappear. This observation means that the magnitudes of such gaps are arbitrary, and that policy makers can manipulate the magnitudes by changing the proficiency standard. Nonetheless, measuring gaps in this way is common practice in the U.S. where one often hears statements of the form: X percent of one group of students meet the proficiency standard in contrast to only Y percent of another group.

Not all achievement gaps between groups of students matter. If group A were composed exclusively of students in grade 4, for example, and group B of students in grade 5, the observation that there is an achievement gap between the two groups has no normative significance. Indeed, in this case, the absence of an achievement gap would be disturbing given that we would expect children to perform at higher levels as they mature and progress through school. Gaps between children of different ages are not problematic because they do not have implications for inequalities in overall flourishing. Similarly, as we noted earlier, if group A were composed of children with IQs below 90 and group A of children with IQ below 90, many would not consider gaps in the observed performance between the two groups normatively problematic. Here the gaps do indeed have implications for inequalities in flourishing but, because they reflect inequalities in children's natural ability, either they are regarded as unobjectionable or productivity considerations are taken to justify them overall.

Quite different are achievement gaps between groups of children defined by their race or their parents' socioeconomic position, and possibly by other characteristics as well. These are generally regarded as troubling because they suggest that children's prospects for flourishing are being determined partly by factors that should be irrelevant. They may constitute problems with respect to any of the three distributive values we have identified, but they may also affect how members of different groups relate to one another. History is relevant here, not because a legacy of mistreatment calls for restitution but because of its implications for social relationships here and now.

The black-white achievement gap

One of the most vexing policy issues for U.S. policy makers has been the persistence of large achievement gaps between black and white children. Figure 3 shows average test scores of 13 year old black and white students, and the gap between the two groups, on the National Assessment of Educational Progress (with reading scores in Figure 3a and math scores in Figure 3b).^{xi} Average scale scores are on the vertical axis. The lower line in each graph shows the scores for black students for each year that a particular test was administered and the upper line the test scores for white students.

The reading scores for black 13 year-olds increased by 25 scale points during the full period, from 222 in 1971 to 247 in 2008, which is far more than the 7 point gain for white students whose scores increased from 261 to 268. As a result, the black-white achievement gap in reading fell from 39 to 21 points during the period. Much of the narrowing of the gap occurred during the 1970s and 1980s, however, and, after a period of widening, only recently may have begun to narrow again. The pattern is quite similar for math achievement. Once again, the 34 point increase in the scores of black students far exceeds the 16 point increase for white students. Similarly the achievement gap between the races fell from 46 points in 1973 to 28 points in 2008, with most of the narrowing occurring before the mid-1980s. The recent trends indicate that the gaps continue to be large and persistent, but the early trends indicate that they can be narrowed.

The reasons for such gaps are clear. They include this country's long history of slavery, Jim Crow policies of racial separation, and segregated schooling, with the schools for black children typically being far inferior to those for white children. Despite the reduction in school segregation in the late 1960s and early 1970s that eventually followed the 1954 *Brown V. Board* Supreme Court ruling against school segregation, the quality of schooling for black students is often inferior to that for white students. ('Quality' here is measured not by financial resources alone but also by the quality of teachers and principals.) Moreover, the achievement of black children is impeded by the disproportionately low income levels of their families, many of which are headed by single parents.

One obvious normative concern is that racial achievement gaps lead to subsequent differences in wages, and hence to inequality in the potential for members of the two groups to flourish economically. Interestingly, writing about the black-white achievement gap in the early 1970s the well-known sociologist Christopher Jencks had no illusions that closing the black-white achievement gap at that time would do much to close the wage gap because of racial discrimination in the labor market. Hence, the argument for reducing the black-white achievement gap had mainly to appeal to its contribution to equalizing other capacities that contribute to flourishing, such as democratic competence, healthy personal relationships, and personal fulfillment. Those other considerations still have force, but in his subsequent early 1990s book on the black-white test score gap, Jencks argued that, while discrimination in the labor market undoubtedly still existed, closing the achievement gap could in fact make an important contribution to closing the corresponding wage gap.^{xii}

That conclusion is consistent with more recent evidence. Based on the same longitudinal analysis of AFQT scores referred to earlier in this chapter, researchers have found that differences in the cognitive ability of blacks and whites accounted for about three quarters of the wage gap between those groups when they were in their 20s and about two thirds when they were in their 30s.^{xiii} The rest of the black-white wage gap is attributable to a variety of other factors including the possibility of employer discrimination against black employees. The data are still consistent with the presence of discriminatory behavior in that black workers typically receive lower wages than their white counterparts with the same cognitive ability. The observation that so much of the black-white wage gap appears to be attributable to differences in cognitive achievement between black and white workers, however, highlights the importance of the achievement gap. In

particular it implies that reducing or eliminating that gap while students are in school could go a long way to reducing corresponding wage differentials once they are in the labor market.

The two other distributive principles, adequacy and priority to the less advantaged group, are also relevant. Regardless of the magnitude of the gap, adequacy considerations would argue for policy interventions to raise achievement levels for black students if they are currently achieving at levels deemed inadequate. In addition, the priority principle would also justify efforts to raise their achievement in so far as, on average, black students are likely to be less advantaged than white students. If one's absolute level of achievement translated directly into one's absolute level of flourishing, then such efforts would be justified even if the measures taken made no difference to the size of the achievement gap - or even if it made the gap wider. The issue is complicated, however, by the fact that achievement is partly a positional good. The achievement levels of black students could improve but if those of whites remain, on average, superior, then black students may continue to suffer from competitive disadvantage in the labor market or find it hard to function effectively in the political arena. Thus the existence of an achievement gap in itself could adversely affect the prospects of black students, perhaps in ways that offend against not only priority to the less advantaged but also some conceptions of adequacy.

The positional aspect of achievement yields one reason to care about the achievement gap between black and white students rather than focusing entirely on the level and trends in achievement of black students. Another reason is that race has a particular salience, and historical significance, which makes inequalities in achievement particularly problematic. It is relatively easy to identify students as black or white and the history of race relations in the US, particularly as that has affected schooling, give inequalities between races distinctive significance. Together these considerations bring into focus one of the competencies we described in Chapter 3, namely the capacity to regard others as moral equals. The racial achievement gap brings with it the serious risk that white students will view black students, and black students will regard themselves, as inferior, simply on the basis of their color. Eliminating or reducing the achievement gap between black and white students will leave differences between individual children, of course, but such differences do not raise the relational equality concerns raised by systematic differences across groups defined by their race.

Notice that the size of the gap may matter in different and complicated ways depending on the value at stake. From the point of view of equality, for example, it might seem that there is a direct relationship between the size of the inequality and the extent of the normative problem. Assuming normal variance around its mean within each group, bigger gaps between group means might seem to be straightforwardly, and proportionately, worse than smaller ones. It is important to remember, however, that the relationship between achievement and wellbeing is not linear. If achievement has a strong positional aspect, then even small gaps in achievement may lead to big inequalities in the metric that is of ultimate significance. The size of the gap is likely to bear also on the concern with relational equality. The bigger the gap the more likely members of a group are to perceive themselves and others as superior or inferior. This is partly because, on the same assumption about distributions within groups, smaller gaps between group means imply more overlap between members (recall Fig 2), so that the groups will be less well defined as higher and lower achieving.

Achievement gaps by family socio-economic status

In the U.S. context, a substantial body of research has demonstrated that children from disadvantaged households perform less well in school on average than those from more advantaged households. This empirical relationship shows up in studies using observations at the level of the individual student, school, district, state, and country, and with different measures of family socioeconomic status (SES): income-related measures such as family income or poverty; education level of the parents, particularly of the mother; and in some contexts occupation type of the parents or employment status. Studies based on U.S. administrative data often measure SES quite crudely, using eligibility for free and reduced price lunch, for example, as a proxy for low family income, and student race as a proxy for a variety of hard to measure characteristics. Research based on longitudinal surveys usually includes richer measures of family background. Regardless of the measures used and the sophistication of the methods, similar patterns emerge.

Achievement gaps between disadvantaged and more advantaged children are not unique to the U.S. They exist even in countries such as Finland, Canada, and South Korea, whose children on average perform extremely well on international tests -- and far better on average than those in the U.S.. Figure 4 illustrates this point. On the horizontal axis is a measure of a family's economic, social, and cultural status (ESCS), a measure that was constructed by the OECD on an absolute scale designed to be comparable across countries. Thus, a child who is from a low ESCS family in one country is comparable to a child who is from a low ESCS family in another country. On the vertical axis are the PISA test scores of 15 year olds grouped by ESCS category for the U.S. and the 13 countries whose students outperformed the U.S., grouped by ESCS category.^{xiv} The first point to emerge from the figure is the upward gradient of test scores across categories of family disadvantage. Average test scores for students in the fifth percentile across all the countries are about 350, far below the average of about 660 for students in the 95th percentile, and the test scores rise monotonically both overall and within each country.

The second point is that a close look at the children in the bottom 5 percent of the ESCS distribution shows that some countries do better with their disadvantaged students than does the U.S. Compared to other countries, Finland and South Korea appear to have the most success with their very low-ESCS students. This relative success largely reflects each country's strong commitment to education and to equal educational opportunity. In Finland, this commitment is rooted in the country's Lutheran heritage and the recognition that an educated population is the country's most valuable resource. In South Korea, the country's historical ties to Confucianism and current efforts to expand the economy lead parents in all ESCS groups to put tremendous pressure on their children to succeed in school. But even in those countries, large differences emerge between students from low- and high-ESCS families.^{xv}

The performance of U.S. students (see the bars at the far right in each set) follows the same pattern as the other 13 countries. Notably, however, U.S. students in families with ESCS below the median perform particularly badly relative to their low-ESCS peers in other countries, while U.S. students from more advantaged backgrounds perform reasonably well by international standards. That is, the largest shortfalls in performance among U.S. students are concentrated among those with relatively low ESCS. This suggests there is room for the U.S. to do better by its disadvantaged students.

The importance of doing so is reinforced by the observation that the U.S. has a far greater proportion of disadvantaged students than many other developed countries. To illustrate with respect to three specific countries, the percentage of students living in low-ESCS families (defined by the OECD as those more than 1 standard deviation below the mean) in the U.S. is more than 2.5 times that in Finland and Canada and 50 percent more than in the Netherlands. The same pattern is even starker if we shift the focus away from the OECD's absolute measure of disadvantage to a country-specific measure of poverty, such as the proportion of students who live in households with income less than 50 percent of the country's median income. By this measure, more than 1 in 5 children in the U.S. live in poverty, far more than the 1 in 25 in Finland, 1 in 9 in the Netherlands and 1 in 7 in Canada.

As with the black-white achievement gap, these gaps between children of unequal socio-economic status raise a number of normative questions. One difference, however, is that at least some of the considerations of relational equality that arise in the black-white case do not apply. There is not the same history of group-based assumptions about superiority and inferiority, and membership of different groups is less visible and salient in everyday life. Gaps in achievement between children raised in families of unequal socio-economic status still offend against a principle of distributive equality but their implications for social relationships are somewhat different. Partly for this reason, normative attention is more likely to focus on the level and trends in achievement of the most disadvantaged students rather than on the existence or magnitude of the gap between students from disadvantaged and advantaged homes.

If policies could be designed to increase the achievement of economically disadvantaged children with no adverse effects on that of more advantaged students two benefits would emerge. Not only would there be more achievement overall – which over time could be expected to lead to an increase in overall productivity and, ultimately, flourishing - but the immediate gain would accrue to those at the bottom of the distribution. There would still be a question about the extent to which achievement translates into overall flourishing at the individual level, both because achievement is only one kind of educational good, which may conflict with others, and because of its positional aspect. But if we set those considerations aside, both 'more is better' and 'priority to the worse off' would support such a policy. If the improvement at the bottom took children there from levels of achievement deemed inadequate to levels regarded as adequate, the distributive value of adequacy would also tell in favour of such a policy.

The situation is less clear if the policies necessary to raise achievement of the disadvantaged students come at the cost of somewhat lower achievement for more advantaged students. That might be the case, for example, if the policies operate primarily through the schools and within a fixed resource constraint. Additional resources – whether in the form of higher quality teachers, more teachers, more support staff at the school level, or more attention by the teacher at the classroom level – for disadvantaged students would then mean fewer resources for the more advantaged students. Even in this case, both the distributive principles of adequacy and priority to the less advantaged could justify the action. With respect to the latter principle, however, policy makers would also need to consider how the decline in resources for the more advantaged students would affect the achievement, and ultimately the flourishing, of disadvantaged students in both the short and long run. For example, if higher income families responded to the shift of resources by taking their children out of public schools, those families

and perhaps later their own grown children as well, may reduce their political support for the public school system, to the detriment of all children remaining in the public schools, including the disadvantaged.

Readers who are committed to the equality principle would want to go further and to equalize achievement levels across groups of students defined by their socio-economic disadvantage. For them, it is the gaps of the type between groups as depicted in Figure 4 that would matter, rather than simply the achievement levels of the less advantaged students. The difficulty in this case is that the large magnitudes of the gaps means that equalization across groups would require bringing down achievement at the top end of the distribution. That, in turn, would have negative effects on the productivity of the economy, particularly if we make the reasonable assumption that the higher achieving students from advantaged families are poised to become particularly productive and innovative contributors to the economy. To the extent that a more productive economy would lead not only to greater overall flourishing but also to an improvement in the flourishing of the least advantaged – an outcome that would depend on a variety of other decisions about how the extra product is distributed – the principle of priority to the worse off would support permitting the inequalities.

Bridge to the rest of the book

We hope that this chapter has begun to vindicate the approach presented in the previous two chapters by helping readers to understand the normative issues at stake when considering achievement levels and gaps. One needs to have in mind the full range of educational goods, and to be clear about how particular gaps – gaps between particular groups and with respect to particular indicators - relate to that full range of goods. Care is also needed when thinking about the different distributive values that are relevant to the assessment of gaps and to the measures that might be taken to address them. It is important also to take heed of a range of independent values – things that matter but are not themselves educational goods – such as childhood goods and parents’ interests.

There are several explanations for why policy makers and social science researchers tend to focus so much attention on cognitive skills of the kind (imperfectly) indicated by the kinds of test scores that yield measures of achievement or achievement gaps. One is that, as we have documented earlier, those skills demonstrably contribute to individual flourishing, both through the labor market and more directly, by making it possible for individuals to live richer lives - partly by opening up opportunities for higher education and the acquisition of further accomplishments. Another reason is that the current state of technology enables policy makers to measure student achievement, or at least some aspects of it, using tests or other forms of assessment that can readily be implemented on a broad scale, something that is far more difficult for other educational goods. Finally, student achievement is directly susceptible to schooling, which is the major policy lever for influencing the production of educational goods. Indeed many people believe that the main purpose of formal schooling is to raise achievement defined in this narrow way.

Yet while understandable, there are real dangers with such an approach. Heavy emphasis on cognitive skills may lead to too little focus on other abilities and dispositions, such as non-

cognitive skills, that also contribute to flourishing. These other skills may be important in their own right, as we highlighted earlier in the context of the GED program, or because of their contribution to a student's ability to develop cognitive skills. Further, too narrow a focus may lead to misleading conclusions about the effectiveness of policy interventions. Studies of the Perry Pre-School program illustrate the potential problem.

This famous program, which was initiated close to 50 years ago in Ypsilanti Michigan, continues to receive attention in the public discussion about early childhood education largely because it was a high-quality, carefully designed, model program. Even more importantly, the program was designed as an experiment in which poor families were first recruited into a study and then were randomly assigned either to the treatment group of children who had access to the program or to the control group, with intensive follow up over time of both groups. The 58 treated children in the Perry Pre-School program attended a half-day preschool every weekday and had weekly ninety-minute home visits for eight months of the year, for two years. Although the treatment children performed significantly better on cognitive tests for a year or two relative to the children in the control group, the test score effects faded out by third grade. Importantly, however, the program clearly did something positive for the children because the treated group exhibited greater flourishing as adults many years later. Specifically, as of age 40, the treated group in the Perry Preschool had higher lifetime earnings, greater educational attainment, and far less involvement with crime than the control group. To judge the program only by its impact on cognitive skills would be to miss its fostering of capacities important for flourishing.^{xvi}

Importantly, the fact that it is technically possible to administer tests of cognitive achievement to all students does not mean it is always desirable to do so. Such testing (and associated grading) still takes resources, and budgetary constraints may well lead to narrower tests than would be appropriate given the ultimate goal of flourishing. Even within the realm of cognitive skills, some are more readily testable than others. Moreover, there are real costs to the use of test scores when they are used for high stakes decisions about schools or teachers. They create incentives for teachers to narrow the curriculum and game the system through teaching to the test and other mechanisms, including, in some well documented cases, cheating.

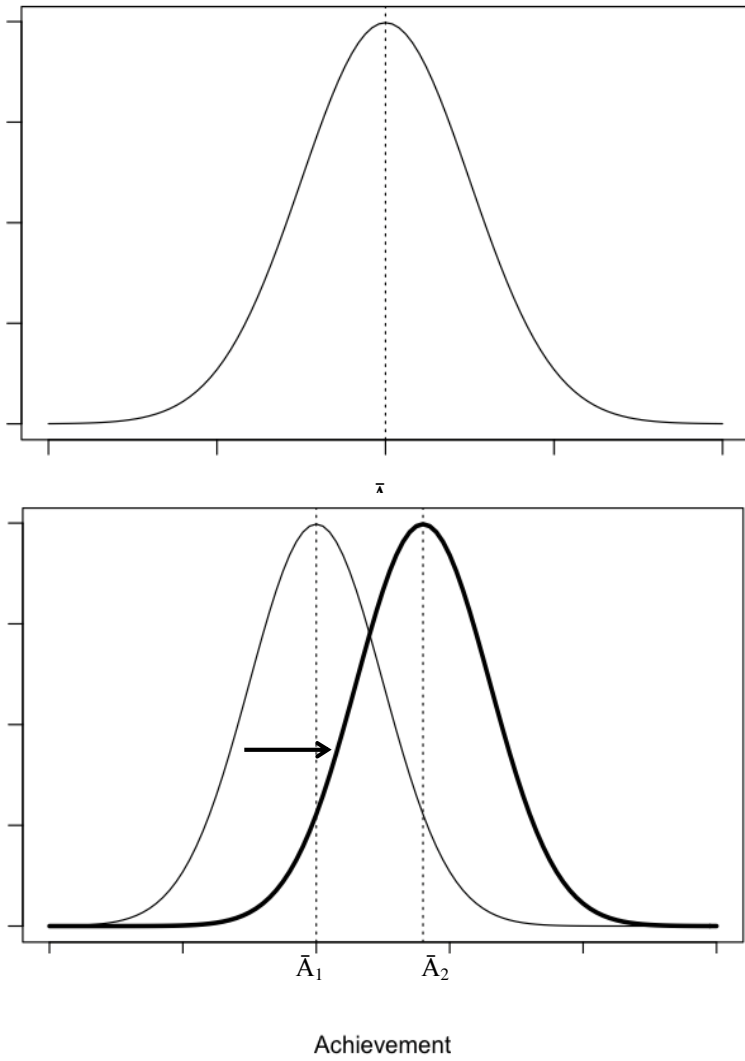
Finally, while it is true that schools may have a comparative advantage with respect to the development of cognitive skills, families and the wider social context still play an important role. The danger is that too much focus on achievement may lead policy makers to forget that educational goods, and inequalities between those who enjoy less or more of them, are not produced by schools alone. Policy makers may expect schools alone to be responsible for raising achievement, when other non-school policy interventions may be needed to offset the many barriers that impede learning for low achieving students. Such policy interventions aimed at children from disadvantaged families might include, for example, greater access to high quality pre-school programs, more access to health care and other social services, and more opportunities for enriching after school and summer programs that middle income children take for granted.

We have emphasized that trade-offs have to be made between increasing achievement overall, raising the achievement of those at the bottom of the distribution, or closing achievement gaps. But properly to assess the weight of those different considerations one must always have an eye on the fact that achievement is not the thing that really matters. This conclusion would be true

even if achievement were measured in a richer and fuller way. What matters is people's overall flourishing, and it is in terms of their contribution to *that*, and to *its* distribution, that achievement levels and gaps must be evaluated. We have indicated some ways in which the conversion of achievement into flourishing makes these judgments complex. The positional aspects of achievement are relevant here, as are the implications for relational equality of some gaps between groups.

We have not offered any judgments about how to balance the various considerations that would arise in any evaluation of, or policy response to, achievement gaps. Indeed, this chapter has not been concerned with policy at all. In contrast, the following chapters do engage directly with policy questions but there too it is not our aim to recommend any particular proposals. Instead it is to exemplify the method we are suggesting as the proper way to go about making such recommendations. We do so initially in the context of three fundamental issues that must be addressed in the design of any school system. Chapter 5 looks at the funding of schools. Chapter 6 considers decisions about who should be accountable to whom and for what. Chapter 7 addresses the distribution of authority among the various actors in the system, with particular attention to parental choice of schools and school autonomy. The thorough treatment of each of those issues would require at least a whole book, so even in those chapters we think of ourselves as no more than providing some examples of our method in action. Chapter 8 concludes by applying the method in even briefer form to policy issues of more direct interest to lower-level decision makers, including those within schools. We intend the policy examples in the second half of the book to convince the reader that the method and normative concepts are applicable to a wide range of educational decisions.

Figure 1. Achievement levels. (figures 1a to 1d)



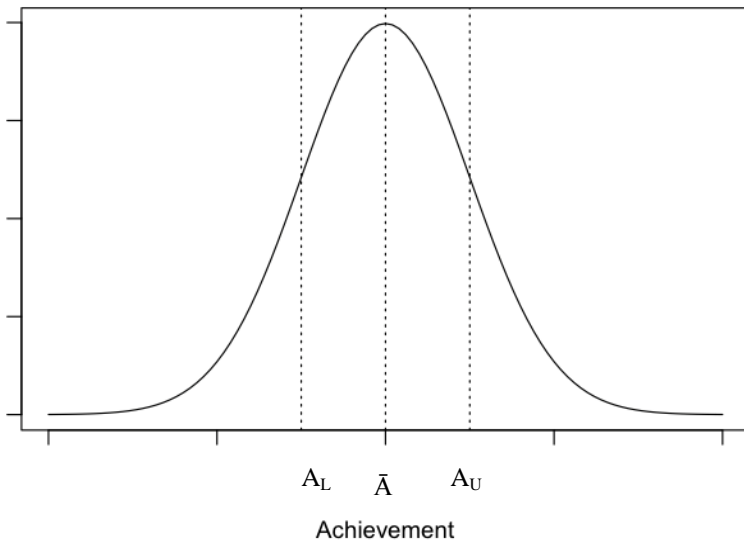


Figure 2. Achievement Gaps

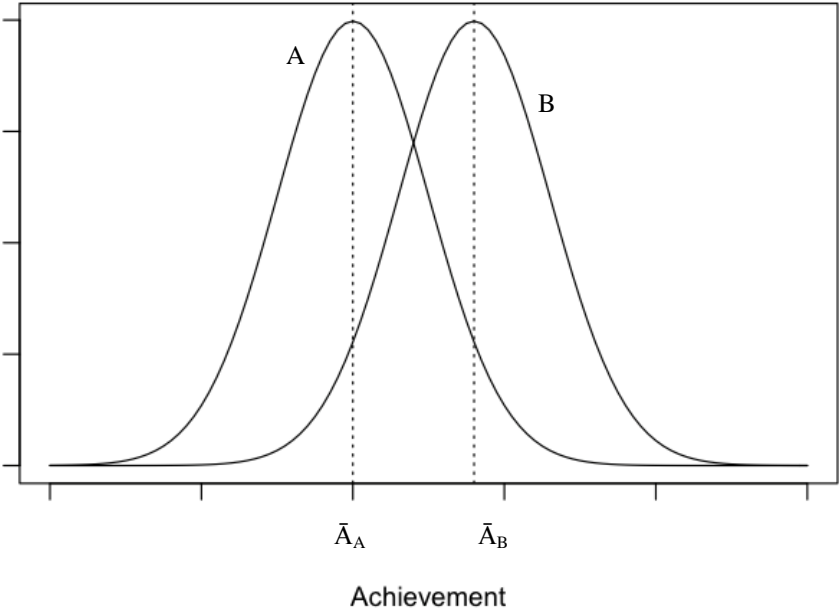


Figure 3a. Trend in white and black reading average scores for 13 year old students.

*Scores on the revised assessment format are reported beginning in 2004.

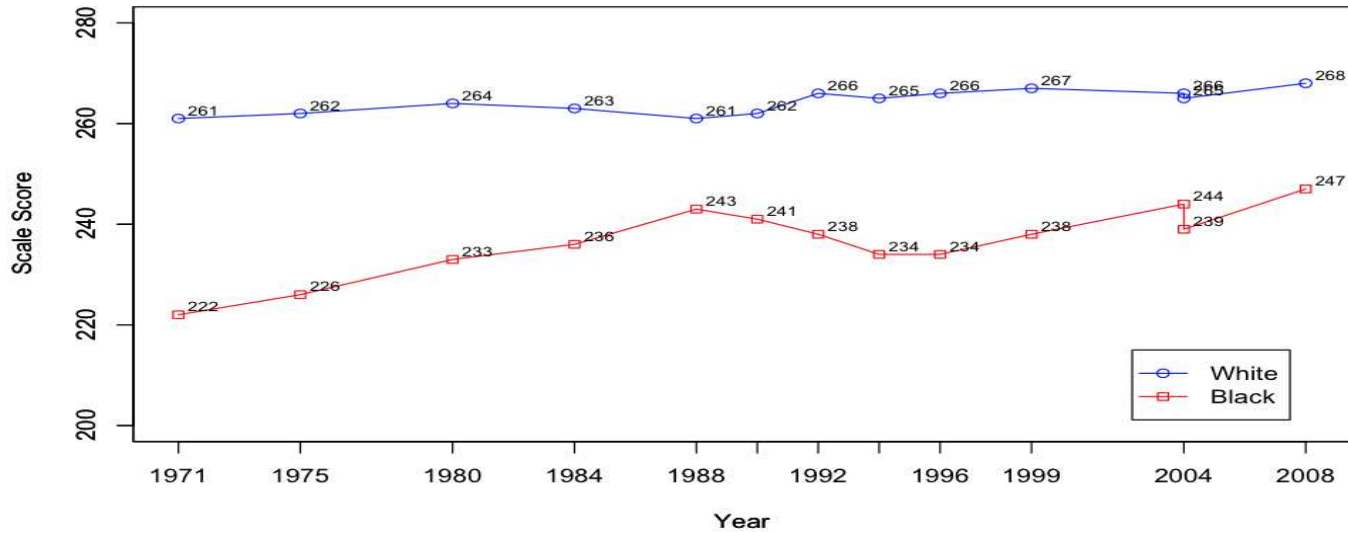


Figure 3b. Trends in white and black average NAEP scores for 13-year old students. *Scores on the revised assessment format are reported beginning in 2004. **Scores from 1973 are extrapolated data

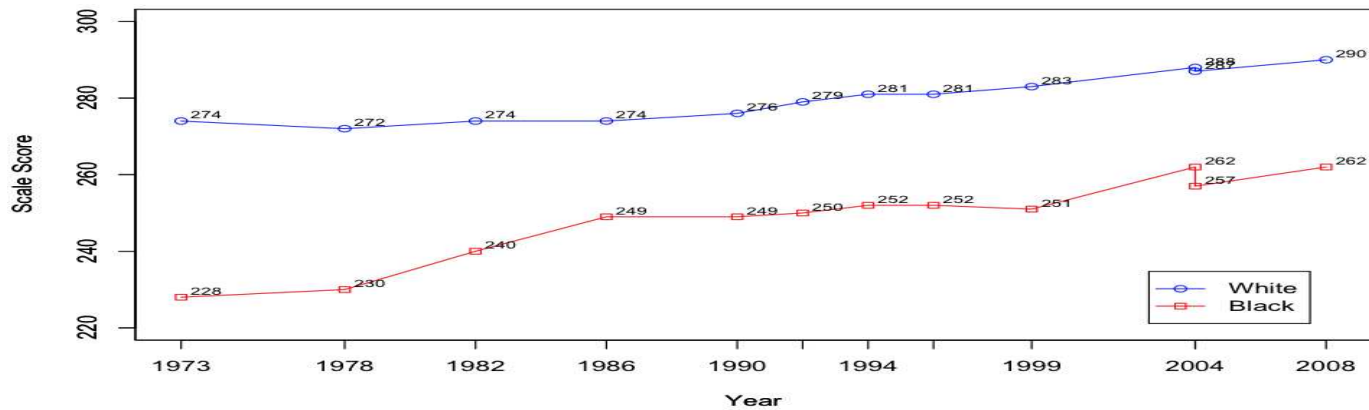
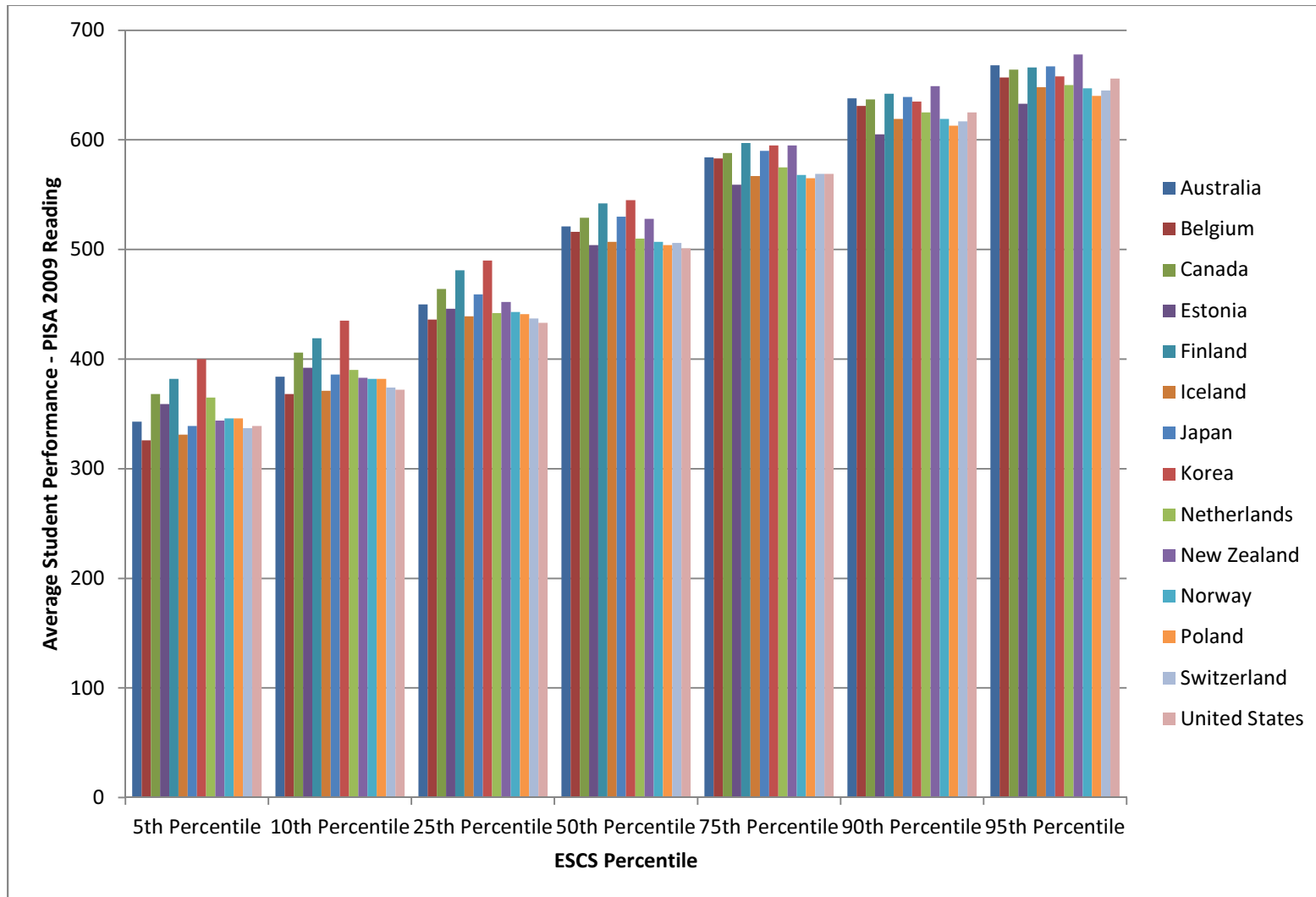


Figure 4. PISA Reading Scores by ESCS Percentile, 14 Countries.



End notes.

ⁱ The Armed Forces Qualifying Test (AFQT) is a general measure of trainability and is the primary criterion of enlistment eligibility for the Armed Forces. It covers a range of topics including general science, arithmetic reasoning, word knowledge, paragraph comprehension, numerical operations, coding speed, auto and hop information mathematics knowledge, mechanical comprehension and electronics information. (Neal, Derek A. and William R. Johnson. 1996. "The Role of Premarket Factors in Black-White Wage Differences." *Journal of Political Economy* 104 (5): 869-895.) The study referred to in the text is based on the national longitudinal study of youth (NLSY) which uses a modified version of the AFQT. See , Richard J. Murnane John B. Willett, and Katie Buckley, "Recent Trends in the Role of Cognitive Skills in Accounting for the Black-White Wage Differential," Draft 2012. (To be checked for permission to cite).

ⁱⁱ See Raj Chetty, John N. Friedman, and Jonah E. Rockoff. 2011. *The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood* NBER Working Paper 17699, appendix Table 4.

ⁱⁱⁱ See Philip Oreopoulos and Kjell G. Salvanes, Priceless: The Nonpecuniary Benefits of Schooling, *Journal of Economic Perspectives*, Vol 25, no. 1, 2011, pps 159-184 and Robert H. Havemen& Barbara L. Wolfe, Schooling and Economic Well-Being: the role of Nonmarket Effects, *The Journal of Human Resources*, vol 19, issue 3, 1984, pp. 377-407.

^{iv} See Thomas S. Dee, "Are there Civic Returns to Education?" *Journal of Public Economics*, 2004, 88: 1697-1720; Kevin Milligan, Enrico Moretti, and Philip Oreopoulos, "Does Education Improve Citizenship? Evidence from the United States and the United Kingdom." *Journal of Public Economics* 2004 88: 1667-1695. Also see Sidney Verba, Kay Lehman Schlozman, and Henry E. Brady. *Voice and Equality: Civic Voluntarism in American Politics*. Cambridge, MA: Harvard University Press, 1995.

^v See James J. Heckman & Yona Rubinstein, "The Importance of Noncognitive Skills: Lessons from the GED Testing Program," *The American Economic Review*, 91(2), May, 2001, pp. 145-149.

^{vi} See James j. Heckman, James J., Jora Stixrud, and Sergio Urzua, "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics* 24 (3), 2006: 411-482; Flavio Cunha & James J. Heckman. 2007. "The Technology of Skill Formation." Chicago, IL. (check this reference)

^{vii} Note possible point to add based on article Julie Berry Cullen, Steven D. Levitt, Erin Roberson, and Sally Sadoff, "What Can be Done to Improve Struggling High Schools," *Journal of Economic Perspectives*, vol 27(2), Spring 2013, pp. 133-152. .

^{viii} See Gordon Marshall and Adam Swift, "Meritocratic Equality of Opportunity: Economic Efficiency, Social Justice, or Both?", *Policy Studies* 1997.

^{ix} Add reference to KY case.

^x CHECK this statement and add reference.

^{xi} The source is The Nation's Report Card: Trends in Academic Progress in Reading and Math 2008. <http://nces.ed.gov/nationsreportcard/pubs/main2008/2009479.asp#section2> . NAEP reports two sets of results, one at the national level over a long time period for students at ages 9, 13, and 17 and one for a shorter time period at the national and state level for 4th and 8th graders. We use the Long Term Trend data for 13 years olds to illustrate the trends since the early 1970s. Check range and standard deviations of the scores.

^{xii} Christopher Jencks and Meridith Phillips. *The Black-White Test Score Gap*. Washington, DC: Brookings Institution Press. 1998.

^{xiii} See Neal and Johnson (1996) and Murnane et al. (2012). Calculated as $44.4 - 15.6 = 28.8$. $28.8 / 44.4 = 0.65$.

^{xiv} PISA stands for the Program of International Student Assessment. These scores were standardized as of 2000 to have a mean of 500 and a standard deviation of 100.

^{xv} For further discussion, see Helen F. Ladd, "Education and Poverty: Confronting the Evidence," *Journal of Policy Analysis and Management*, 31(2), Spring 2012, 203-227. For more information on Finland see Pasi Sahlbert, *Finnish lessons: What can the world learn from educational change in Finland?* (New York: Teachers College Press, 2011 and on South Korea, see B. M. Ahn, "Education in South Korea: National treasure or national headache?", paper presented at the annual conference of the Association for Education Finance and Policy, Seattel, WA, 2011.

^{xvi} See L. Schweinhart, J. Montie, Z. Xiang, W.S. Barnett, C.R. Belfield & M. Nores. *Lifetime effects: The High/Scope Perry preschool study through age 40*. Ypsilanti: High Scope Press, 2005.