

The Challenges of Measuring School Quality: Implications for Educational Equity

Helen Ladd

Susanna Loeb

Section 1: Introduction

Nearly all countries, including the United States, view elementary and secondary education as so important for the well being of both individuals and society that they make schooling compulsory through some age, whether that be 14 as in many developing countries or 16 or 18 as in various U.S. states. In addition, there is a worldwide consensus that all students, but especially those in primary school, should have access to free, publicly financed schools with no required school fees. In practice many countries, especially developing countries with limited resources, do not meet this latter requirement, and often permit schools to charge fees that in some cases can be substantial (see Ladd and Fiske 2008: ch. 16). The policy throughout the United States has been clear: public schools, including both traditional schools and publicly funded charter schools, are not permitted to require parents to pay school fees for their children to enroll in the school. Compulsory schooling, supported by full public funding, reflects the observation that elementary and secondary education provides not only private benefits to those who attend school and their families but also public benefits to the broader society.

Among the private benefits are consumption benefits to the enrolled students of being in a safe, engaging and potentially enjoyable school environment; consumption benefits to their parents in the form of child care and satisfaction in their children's development; and, importantly, investment benefits in the form of future returns to students in the form of higher paying jobs, better health and a more fulfilling life (Card 2001; Haveman and Wolfe 1984). These private benefits – both the consumption and the investment benefits – can also be categorized as intrinsic or extrinsic. Intrinsic benefits arise when education is valued for its own sake such as the pleasure of being able to solve a complex problem or appreciate artistic expression, and extrinsic benefits arise when education serves as an instrument for the attainment of other valued outcomes such as the higher income for working parents that is facilitated by having children in school, or the potential for the recipients of education to seek higher paying jobs and fulfilling careers than would otherwise be possible. Regardless of the classification, it is clear that education provides a variety of different types of private benefits, many of which accrue long after the students have been in school.

If the only benefits were private, one might expect families to pay for a significant part of their child's education, as is typically the case at the higher education level. Even in the case of exclusive private benefits, however, there would be a compelling argument for making education compulsory and providing public support. The argument rests on the government's responsibility for protecting the interests of vulnerable groups, in this case children, who are not in a position to protect their own interests. Thus, it can be viewed as both unfair and undesirable for children whose families invest little in their education – regardless of whether that reflects limited resources or weak preferences for education

(see essays by the Suárez-Orozcos and Harris in this volume for the debate) – to be kept from gaining access to the skills and orientations needed to lead a productive life and to unlock their potential.

Clearly, however, the benefits to schooling accrue to more than just the child and the child's parents.

Among the public benefits of schooling are short run benefits for others that arise from keeping idle children off the streets and away from crime or other antisocial behaviors, and the longer run benefits of having an educated citizenry capable of participating in the democratic system and a workforce that is productive and innovative. These longer run benefits accrue not only to the residents of the local community in which the children live, but also to the broader society. Low educational investments in students in one jurisdiction have spillovers to other jurisdictions because people move across jurisdictions, citizens participate in the political life of the nation as well as that of their local community, and the productivity of one geographic area of the country can affect overall productivity.

Even without government financing of education, families would have an incentive to invest in the education of their children in order to generate future benefits for themselves. Parents gain directly from their children's future earnings if those children care for them when they are elderly. Parents also gain from investing in their children's education whenever having flourishing, happy children increases their own happiness. Many families, however, would invest less than would be most beneficial for the larger community because they would not take into account the benefits that would accrue to others. Such under-investments are likely to be largest for low income families, for whom

the public benefits, including the creation of conditions for the democratic participation of the citizenry, of educating their children could be large relative to their perceptions of the private benefits. In addition, they may have less information as well as fewer resources to invest currently for future returns.

Governments have the potential to overcome some of these under-investments. They can raise taxes and make schooling less expensive to individuals so that these individuals invest more to account for the externalities of education, the benefits that others get from an individual's schooling. Governments can also give loans to make it easier for families to invest, and they can require attendance.

Given the benefits of education, almost all societies invest in an education system and the vast majority provides free education to young children. In so doing, each government needs to make many choices about how to fund and govern schools. These decisions have implications for the quality of schools and the educational opportunities available to children. Informing these choices well requires a clear definition of education quality and an understanding of how to measure it.

In this paper, we explore the complexity of defining and measuring education quality in a way that can help public decision making. We discuss common approaches to measuring education quality and explore the advantages and disadvantages of each approach in terms of accuracy and reliability. We then turn to a discussion of the distribution of education quality in relation to the normative standards of equal quality schooling, equal educational opportunity, and adequacy, and we highlight the merits of the different approaches to measuring school quality with respect to each equity standard.

Section 2: What is education quality?

In its simplest form education quality can be conceptualized as the investment and consumption value of the education. The investment portion captures benefits in the form of higher earnings, better health, contributions to the arts, effective participation in the democratic process, and other outcomes that education enhances. The consumption portion of education quality captures the benefits to children and their families of having safe, supportive, and happy environments. Taken from the perspective of the community, the quality of an education system refers not only to the sum of the investment and consumption benefits, but also to how they are distributed across individuals. The value of any particular pattern is likely to differ across societies. For example, highly unequal patterns of educational consumption across individuals may be unacceptable in some societies while in others it may be more acceptable provided all children receive a minimum floor of consumption. As another example, some communities may look for equality in the investment benefits of education while others may desire to provide greater investment benefits to students whose families are less able to provide them, thereby compensating these children for low family resources. The distribution of both consumption and investment benefits may also affect the robustness of the democratic process and the degree of societal cohesion, both of which are valued in a democratic society.

Of course, it is unlikely that all members of a community will agree on the value of different components of education quality or on how they should be distributed. Children and their parents often differ on what is a high-quality day at school. Families also disagree

on what is high quality education – with some valuing investment returns of one type and other families another type. For example, some parents value the development of art and music skills and appreciation for its own sake while others value the arts primarily for their ability to motivate students to learn more math or develop better reading skills. In addition, families disagree with respect to the values that children should be learning in school. Educators also often differ with children, with parents, and among themselves about what are worthwhile outcomes as well as what are worthwhile types of education consumption.

Communities and their government representatives have to decide on how to balance these differing perspectives, much the way they need to balance differing perspectives in other areas of public decision making. Education may be particularly sensitive because it touches on parenting, rights and values. It is further complicated in federalist systems in which many different communities each have a say in public education. Certainly in the United States, local, state and federal governments each play an important role in the public education system and often disagree on the best approach to schooling. While many papers could be and are written about how best to govern and finance education in a federalist system, those issues are not the subject of this paper. We have a more limited, but nonetheless challenging, goal of better understanding what education quality is and how to measure it.

In keeping with this discussion, ideally we would measure the quality of an education system by the investment and consumption benefits it provides. Measuring education in this way, however, is not an easy task in part because some of the benefits of education are difficult to quantify and in part because investment benefits do not emerge

immediately. We do not know, for example, how much income a first grade child will earn two or three decades in the future. Moreover, even if we could look at data 30 years after the child attended school, we would be learning about the quality of the education system 30 years earlier, which is not of much use for current decision making. As such, we need proxies for education quality. Not surprisingly, none of the available proxies is perfect.

In what follows we discuss the relative merits of commonly used proxies for education quality.

One set of proxies aims to capture the inputs to schools that are the building blocks or ingredients for producing a high quality education system. Measures of resources – either in the form of spending per pupil or specific school inputs such as the number of teachers per pupil – are the most concrete proxies for school quality and are the ones most commonly used. Although resources may be necessary for a high-quality school or district, they may not be sufficient given that some schools are likely to use their resources more effectively than others. Thus, direct measures of school processes, as observed by external evaluators, can serve as alternative measures of school quality. For a number of reasons, including the complexity of schooling and the difficulty of standardizing evaluators' ratings, even measures of school processes are flawed as a measure of education quality. A third type of quality measure uses proximal student outcomes such as test scores and educational attainment. Although these measures do not fully capture the investment outcomes of interest, they are often justified on the ground that they predict such outcomes. One of the challenges in using outcome-based proxies for quality is determining what student outcomes would have been in the absence of the schooling system so as to uncover the contribution of schools to the outcomes in question.

Section 3: Proxies for school quality

The three most common proxies for school quality are measures of resources, of internal processes and practices, and of student outcomes. In this section, we discuss each of these measures and evaluate their strengths and weaknesses in light of the framework just presented.

Resources

Spending per pupil is an intuitively appealing rubric for measuring education quality. Setting aside for the moment the fact that price levels may differ across places, we all have a sense of the scale of a dollar, what it can and cannot buy. Such a measure can be interpreted as a weighted average of the various inputs used by a school, with the weights being the prices of each input. According to such a measure, a school or district with more teachers who are experienced (and hence have higher salaries) would be spending more than a school or district with fewer experienced teachers, all other factors held constant. Thus, to the extent that the differential salaries paid to teacher reflect their true quality, a measure of spending per pupil appears to be a reasonable way to capture both the quantity and quality of the resources available to a school.

An advantage of spending per pupil as a measure of school quality is that it is not based on any specific assumptions about the best, or preferred, way for schools to allocate their total resources among specific inputs. For example, the same amount of (per pupil) spending in two schools could be used for smaller classes with less experienced teachers in one school

and larger classes with more experienced teachers in the other. In the absence of evidence that certain configurations of resources are preferred to others in all schools regardless of their context, it would be inappropriate to attribute higher quality to one school than another. Finally, this single-dimensional measure allows for straightforward comparisons across schools or districts with statements of the form: district A spends 40 percent more than district B, with an implied comparable statement about the relative quality of the two districts.

In the United States in 2007, the average current per-pupil expenditures for public elementary and secondary education equaled \$9,683 (U.S. Department of Education 2007). This average masks great variation in spending both across states and across districts within states. On average, for example, schools in Utah spent \$5,706 per pupil, while those in New York State spent an average of \$15,546 per pupil. By one estimate, about 70 percent of the variation in per-pupil spending across U.S. school districts is attributable to variation across states and about 30 percent to variation across districts within states (Corcoran and Evans 2008: Table 19.2). Largely as a result of the school finance and property tax reform efforts in many states that reduced both the within-state and across-state variation, spending inequality across school districts throughout the country declined substantially between 1972 and 2000 but then rose slightly in subsequent years.¹

There is far less information on the distribution of dollars across schools within districts. Because of the single salary schedule and the associated well documented propensity for the more experienced — and hence, more costly — teachers to leave schools with high proportions of low-achieving, low income and black students, one might expect spending per pupil to be lower in these schools than in more advantaged schools. Working

in the other direction, these schools may receive more funds from state and federal governments targeted towards needy students of these types. As one example, New York City public schools spent an average of almost \$12,800 per pupil during the 2003-2004 school year, with school-level spending ranging from approximately \$3,500 per pupil to \$24,500 per pupil. In this district, schools with higher proportions of poor students, low achieving students, and especially special education students, spent more per pupil on average than other schools though much of the variation across schools is not easily explained by student characteristics (Schwartz, Rubenstein and Stiefel 2009).

Despite the intuitive appeal of per-pupil spending as a measure of school quality, it suffers from serious drawbacks even as a measure of a school's resources. First and most important, the costs of any given quality-adjusted input often differ significantly across districts and may differ as well across schools within a district. As a consequence, at a minimum, spending would need to be adjusted for the costs of inputs to be used as a measure of a school's resources. Costs of inputs differ for a number of reasons. The costs of facilities, or the annual debt service needed to finance them, are likely to be higher in large cities where land prices are higher than in smaller cities or rural areas with lower land prices. Probably most importantly, the fact that college educated workers earn different wages in different parts of the country means that districts in high-wage areas typically have to pay higher salaries to attract teachers than districts in low-wage areas. As an example, Taylor and Fowler (2006) find that in 1999-2000 starting teacher salaries were 27 percent higher in California than in Kansas (\$32,190 versus \$25,252). Because most of this difference was attributable to the higher wages for college educated workers in

California, the cost-adjusted teacher salaries in the two states were almost the same (\$29,481 and \$29,528, respectively).

Cost differences for teachers also arise across districts or across schools within a district because teachers prefer to teach in some neighborhoods and with some types of students than in and with others (Boyd et al. 2010; Jackson 2009; Clotfelter, Ladd and Vigdor, 2011). In particular, schools serving educationally advantaged or high achieving students may be able to recruit higher quality teachers at any given salary level than schools serving less advantaged or lower achieving students. Disadvantaged schools are likely to end up with lower quality teachers unless they are able to override the single salary schedule and raise the salaries of teachers in their schools. One recent study shows that the additional salary required to retain high quality teachers in disadvantaged schools at the same rate as in more advantaged schools could exceed 50 percent, with the required salary differential depending on the extent of school segregation (Clotfelter, Ladd and Vigdor 2011).

Given the imperfections of per-pupil spending as a measure of school resources, it may be tempting to measure resources directly. For example, one could look at the number of teachers per pupil in a school, access to computers, the length of the school day or year, and/or the availability of after-school programs. In U.S. public elementary and secondary schools in 2007 there were 15.7 students per teacher on average, down from 17.9 students per teacher 20 years earlier in 1987 (U.S. Department of Education 2009). As with other resources, the number of teachers per student varies substantially across and within schools and districts. The main benefit of using specific school resources as a measure of

school quality is that by quantifying resources instead the dollars, one avoids the difficulty of having to adjust for cost differences.

As we have already suggested, when using resources as a measure of school quality, one must measure the quality as well as the quantity of resources. Clearly, not all classroom books or computers are the same. Similarly, not all teachers are the same. Two schools might employ the same number of teachers per pupil but these teachers may come with very different skills and knowledge that affect their teaching ability. Unfortunately, just as it is difficult to measure quality of the overall education system, it is difficult to measure the quality of teachers for generating educational benefits. Instead, until recently, most studies of the distribution of teachers across schools used proxies such as teachers' experience, certification, and academic ability (as typically measured by licensure test scores) when assessing differences in the quality of teachers serving different students. While these proxies are not ideal measures, they are likely to capture differences in the appeal of teaching in different places and thus, at least in part, adjust for differences in quality. Studies using such proxies in North Carolina and New York found that regardless of the proxy used, schools serving the high poverty student populations had far higher proportions of teachers with weak credentials than did the schools serving more affluent students (Clotfelter, Ladd and Vigdor 2007; Lankford, Loeb and Wyckoff 2002).

Although using school inputs to measure education quality has some advantages relative to per-pupil spending, it too has drawbacks. First, as discussed above, there is no consensus on the correct measure of quality for many education resources, particularly for the key human resources in schools – teachers and other staff. Second, the best configuration of resources probably differs across schools and, again, there is no consensus

on these optimal configurations. Third, measures of school resources either focus on only one of the relevant inputs or, if they are intended to represent a bundle of inputs, the question arises of how to weight the inputs within the bundle. That brings us back to some form of spending measure, with the elements of the bundle weighted by their prices.

As a measure of education quality, however, cost-adjusted spending still suffers from two additional problems. First, it takes no account of differences in the effectiveness or efficiency with which dollars are spent. At the extreme, some school leaders simply may be corrupt and not use the dollars to benefit students. Even without corruption, more knowledgeable and effective leaders can achieve greater education quality for the same cost-adjusted spending level by implementing school processes and practices directed toward the valued outcomes.

Second, cost-adjusted spending levels take no account of the fact that some schools need more resources than other schools to offer equivalent schooling. For example, consider two schools, one of which has a far higher proportion than the other of special needs students whose legally mandated individual education plans require that they be taught in small classes. Compared to the second school, the first school would need more teachers per student on average to provide an equivalent quality of schooling to the non-special needs students within the school. For a more general example, consider two students, A and B, who are similar except that they attend different schools. If the other students in the school attended by A are more likely to be disruptive or the variation in the achievement levels of students in that school is much greater than in the school attended by B, the school serving A may need more resources, either in the form of more teachers or

more teachers with special qualifications, to provide student A with an education that is equivalent to that received by student B in the other school.

Arguments similar to these have been used by the Dutch to justify their system of weighted student funding. Ever since 1917 with the extension of public funding to religious schools, the Dutch objective has been to provide equal quality schools, regardless of whether such schools were operated by Roman Catholics, Protestants or the government. Historically, that meant providing equal funding for each school. With the influx of low skilled immigrants in the 1960s and 1970s and the resulting concentrations of immigrant children in some schools, however, the Dutch realized that equal resources did not translate into equal quality. Thus, starting in 1985, they implemented a system in which immigrant children whose parents had low education would bring with them almost twice as much funding to the school they attended as would native Dutch with well-educated parents (Ladd and Fiske 2009).

Importantly, this argument for the need for more resources in one school relative to another is based on the goal of providing *equivalent* quality education to each student, regardless of the school he or she attends. It is not grounded in the view that educationally disadvantaged students may require higher quality education in order to have equal – or, possibly, adequate – educational opportunity. Questions of how best to distribute education quality across students raise a different set of issues than those of interest in this discussion of how to define and measure school quality. We address the arguments for providing different quality of education to different students in the final section of the paper. For now, we are still working on how to measure quality.

Observational Measures of Internal Processes and Practices

As suggested by some of the examples in the previous sections, even schools – or broader education systems – with similar cost-adjusted levels of resources may differ in terms of quality because of the way they use those resources. An alternative approach to measuring education quality is to observe what goes on within classrooms, schools, districts or broader education systems. Such measures typically focus on internal processes and practices. The best quality measures of this type would be generated by trained observers and would be based on formal rubrics or protocols designed to produce consistent measures across units.

Examples of process measures of education quality include evaluations of teacher quality based on observations of their practices in the classroom (Grossman et al. 2010; Kane et al. 2010) and inspectorate reports on individual schools. Such reports are typically based on visits by external review teams to individual schools on a periodic basis. Although multiple countries around the world use such an approach, it is less common in the U.S. Only recently have some states and cities begun experimenting with school or district inspections systems of this type. The New York City Department of Education, for example, currently sends review teams to individual schools as part of its larger effort to promote school quality, and Massachusetts has had various permutations of a statewide system for evaluating both districts and individual schools for several years. Interest in charter schools has induced some states, including Massachusetts, to send review teams to all their charter schools to evaluate the quality of the school, both early in the life of the school and just prior to the reauthorization decision.

This observational approach is advantageous in that it can provide a more nuanced and comprehensive measure of quality than can resource-based or spending-based measures. Moreover, by highlighting areas of quality shortfalls, such information can be useful to the observed units in that it provides guidance on areas in need of improvement. For observational measures to provide valid measures of school quality, however, the observed measures must be closely linked to educational outcomes of value.

One area in which observational measures of quality are relatively well developed and well justified by research is in the area of early childhood education. In this context, research documents that how teachers interact with children in the classroom affects children's learning and development (Pianta et al. 2004, 2007). Classroom observations can therefore provide more valid information on the quality of a pre-school program than simple measures of resources, such as the number or educational achievement of teachers. This example illustrates not only the potential for observational measures to be useful measures of quality, but also suggests that the case for using direct observation is most compelling when it is difficult to get crucial information on quality in other ways.

Even with their potential benefits, observational measures are rarely used at a large scale at either the early education or K-12 levels in the United States. Both the early stage of development of most of the measures and cost of effectively implementing a system of assessment based on observational measures limits their current use. The experience of New Zealand is illustrative of the challenges involved in measuring school quality with the use of direct observation. An innovative part of the country's reform package of the early 1990s, which turned operating responsibility over to individual schools, was the establishment of an Education Review Office (ERO) designed to monitor school quality

through periodic school visits. The initial intent was to evaluate each school in terms of its own mission statement, but the vagueness of the mission statements made that approach unworkable. The ERO instead turned its attention to how well the school was complying with national guidelines for school policies and also to how well the school itself was using information on student performance to make internal policy. Despite ERO efforts to focus on learning outcomes, the reviews often became mechanistic, were heavily focused on management procedures and did not necessarily foster better educational outcomes. Starting in 2003, the country introduced a new planning and reporting framework for schools. Under this new system, the ERO now focuses on the process questions of the following type: How well is information on student achievement used, both formally and informally, to develop programs to meet the needs of individuals and groups of students; how well is available time used for learning purposes; how effective are the systems for identifying and meeting staff professional development needs; and how well does the school establish partnerships around learning with its community (Ladd 2010). The focus is not on learning outcomes themselves but rather on the robustness and coherence of the internal processes and practices that policy makers believe are associated with good outcomes.

The focus on how well schools make use of data on student achievement to allocate resources within schools and on the coherence of policies for supporting student learning emerge as central components of all the inspectorate systems of which we are aware. The logic underlying such evaluation systems is at one level quite compelling. Because schools differ so dramatically in terms of their students and, in many cases, in the resources

available to them, it makes intuitive sense to judge quality based on how well the school uses those resources to meet the needs of the students it serves.

At the same time there are some potential drawbacks to this approach. One is the danger of placing emphasis on processes that are in fact not ultimately related to valued educational outcomes, either insofar as the processes have not been validated in the literature as being linked to those outcomes or insofar as the validation is based on measures of quality that are themselves imperfect or narrow. For example, if the reviewers place a lot of emphasis on the use of data, and those data are all based on tests of basic skills in math and reading, then the observational measure may provide a narrow measure of school quality. Working in the other direction, the review protocol could be so broad that the reports are not useful for distinguishing between those processes that contribute significantly to quality and those that are less directly predictive of quality. Furthermore, because it is based on human judgment, the observational approach may be subject to variation in ratings that reflect differences across reviewers rather than differences in education quality. Unless the reviewers are well trained, and the reports tested for reliability across reviewers, the system could provide misleading information on school quality.

While the validity of many current observational measures can be questioned, the most common concern about the observational approach is its high cost. As the school-based approach has been implemented in practice, costs include the time of the reviewers and the costs of their training as well as the time of school officials for preparing for reviews and responding to them. These costs depend on the nature of the program including the size of the review team for each visit, the salaries of the reviewers, the

frequency of visits and the costs of training high quality reviewers and of assuring reliability. Offsetting some of these costs, however, are any school improvements that the process generates.

The well-developed English inspectorate system as it operated in the 1990s was very expensive, with visits to primary schools costing about \$20,000 and to secondary schools, about \$37,500 (Kogan and Maden 1999 – converted from pounds to dollars). The smaller scale system in Charlotte/Mecklenburg, North Carolina (CMS) that built on the English model but made greater use of internal personnel was far less expensive with an estimated cost of each school evaluation below \$9,000. The program of charter school evaluations in Massachusetts has an average cost per charter school visit of about \$3,500. The potentially high costs of the observational approach tend to make U.S. policy makers wary of this approach and skeptical of the potential for taking it to scale. At the same time, many other countries have historically made much greater use of the observational approach and have used it for all their schools. In some countries, most notably England, however, the school visits have become less extensive and comprehensive in recent years as emphasis has shifted to a greater reliance on outcome measures, and most prominently on test scores.

The processes approach to measuring school quality, while potentially better at measuring the quality of schooling than purely resource-based measures, requires an understanding of *how* the benefits of education are produced. For example, in order to measure teacher quality or the quality of school leadership, observers need to know what good teaching and good school leadership look like. Although school systems have made substantial recent progress in measuring processes, most education processes have no

protocols for observation, much less well-validated protocols. To the extent that schools continue to diversify — for example through multiple pathways to high school completion, home schooling, or distance learning – the job of directly measuring education practice becomes more challenging. Given the complexities of measuring quality through the inputs (either resources or processes), it is appealing to consider whether it is feasible to measure the outcomes of education directly.

Student Outcomes

As described above, education quality encompasses the consumption and investment benefits of schooling. These benefits could include general satisfaction, earnings, non-pecuniary job benefits, health, low crime, and range of other outcomes. Most of the returns to education accrue long after students have completed school and certainly too far in the future to provide meaningful measures of the current state of education quality or to serve as useful measures for education decision makers.

More immediate measures of student outcomes such as achievement and educational attainment may proxy for these later outcomes. Ample evidence confirms that student test performance is associated with higher educational attainment and with earnings (Johnson and Neal 1998). The contribution of cognitive skills, as measured by test scores, to future earnings also appears to have risen over time (Murnane, Willett, and Levy 1995). Given this link between test performance and subsequent earnings, one might use measures of the student achievement gains attributable to schooling as a proxy for school quality. This approach to measuring school quality is currently popular in the United

States, arising from state-level test-based accountability programs introduced in the 1990s and currently represented by the No Child Left Behind Act at the national level.

The benefits of using outcome-based measures of school quality are clear. Directly measuring outcomes negates the need to know how education is or should be produced. If we measure quality by how much a student learns, we do not need careful analyses of how to adjust spending for costs across regions or across schools serving different populations of students, or careful observational measures of good teaching. If it were easy to measure education quality directly through observation, this advantage of outcome-based measures would not be important, but measuring education quality directly is difficult and costly.

Not surprisingly, the use of student outcomes as a measure of education quality brings with it meaningful drawbacks. First, measures of student outcomes used in current education systems do not capture the breadth of student outcomes that individuals and society value and that therefore are central to the concept of school quality. By requiring that all states test all students in grades 3-8 annually in math and English language arts, No Child Left Behind (NCLB) has dramatically increased the availability of student test performance in those two subjects, and has required that those test scores serve as the basis of measures of school quality. Those two subjects, however, represent only a portion of the content areas that may generate future education benefits. They ignore completely other skills and dispositions that could, at least in principle, be measured at the time of schooling and that are likely to be important for future outcomes – such as the ability to work effectively in groups and to empathize, as well as achievement in science, history and the arts (see Rothstein, Jacobson, and Wilder 2008 for a fuller discussion). Moreover, current tests do not even attempt to measure the consumption benefits of education, such

as student happiness and health. Even within the domains of math and English language arts, tests focus on skills that are more easily tested and thus emphasize the importance of some content over other content, creating outcome weights that do not necessarily align with society's values. While the tests may be highly reliable measures of student achievement in the tested content, they are unlikely to be valid measures of the full set of education goals.

The focus of current testing programs on math and English language arts stems from a variety of sources. It is impractical to measure all outcomes of interest for all students each year, even if the outcomes are limited to current (in contrast to future) knowledge and skills. The time necessary to measure each domain reliably is simply too great. To get a sense of the success of an overall education system, one could measure different outcomes for different students or in different years, but this approach is more complex and sacrifices some comparability across students and over time. In addition, while reliable measures of student outcomes span areas broader than math and English language arts, many important outcomes are not as easy to measure with currently available instruments. Whether a broader approach to testing would better capture education outcomes (both consumption and investment benefits) of interest is an empirical question that has yet to be answered. Logically, tests covering broader content should capture a broader set of valued outcomes, and should also reduce the tendency of narrow tests to lead to narrow curriculum. At the same time, though seemingly broader, the tests may simply pick up skills similar to those measured by the current tests (using different questions), whether those be test-taking skills or skills important for future success. Broader tests may sacrifice both time and reliability for little gain if they do not improve

our ability to measure the consumption and investment benefits that characterize high quality education. Although we cannot directly measure the long-run outcomes of education, we can judge a testing regime that aims to measure education quality on the extent to which it appears to measure the outcomes of ultimate value.

A second difficulty of using student outcomes as a proxy of school quality is that, even if current student outcomes such as test scores well represent the long-run outcomes of interest, it is often difficult to correctly attribute the portion of a student's outcome that is due to schools. We have defined school quality as the benefits that students and society get from their schooling. The benefits of interest arise, however, not only from the schooling itself but also in part from other parts of their lives, particularly from their families but also from their communities and other experiences. If we simply use student test performance as the measure of school quality, we would falsely attribute to schools' differences in performance due to ability and to family background. A first step is to look at gains in performance of students instead of levels, so as to separate the differences at the start of schooling from the changes during schooling. In fact, however, the evidence shows that ability and family background affect gains as well as levels. Despite the development of sophisticated approaches to estimating the effects of schools on student outcomes, even the most well considered does not cleanly isolate school effects from other effects. For example, most empirical models of student outcomes adjust statistically for differences in achievement gains by family background, but these adjustments inappropriately eliminate any components of school quality that are systematically associated with family background. That is, some of the difference in the achievement gains of students in schools with a higher proportion of students in poverty is due to fewer educational opportunities

outside of school, but some of the difference is probably due to lower quality schools. Statistical adjustments eliminate both of these sources of lower achievement from the measure of school quality when, ideally, the differences in achievement due to lower quality schools would remain in the measure of school quality. The models could be run without adjustments, but then we would inappropriately attribute differences in educational opportunities outside of schools to the schools. The bottom line is that it is very difficult to separate the contributions of differences in average quality of schools across groups from those of differences in other inputs to the variation in student outcomes across groups. As we discuss further below, if the goal is to understand the quality of educational opportunities available to students overall (both inside and outside of school), then this difficulty of attribution is not important. However, if the goal is to assess the quality of schools themselves, then attribution is both important and difficult.

Section 4: Education quality in the context of education equity

Thus far we have introduced the concept of education quality and different approaches to measurement: resources, observational measures of internal processes and practice, and proximal outcomes. Each of these measures has advantages and disadvantages. In the context of equity discussions, the relative salience of these advantages and disadvantages depends on how one conceptualizes educational equity. Setting aside the voluminous literature by philosophers, lawyers, public finance economists, and education researchers, we focus here on only three concepts: equity as

access to equal quality schools, equity as equal educational opportunity, and equity as adequacy.

Equity as access to equal quality schools

One plausible equity goal is that all schools should be of equal, or equivalent, quality. By this we mean that student A would do equally well, or badly, by going to school X, Y, or Z. Average student outcomes would probably still differ across schools but those differences would be attributable to differences in the abilities, motivations, and outside-of-school supports of the students in each school and not to differences in school quality. This definition of equal guarantees that a student would receive equal quality schooling regardless of which school he or she attends; it does not guarantee that all students receive the same quality education since education quality can vary between students within a school. One consequence of this equity standard is that the systematic sorting of students across schools (for example, as a result of residential segregation by income or race) would not increase achievement gaps between groups beyond those associated with the background characteristics of the students themselves. Nor, however, would schools serve to narrow those gaps.

Our goal here is to consider the relative merits of the different approaches to measuring school quality in the context of this equity standard. As discussed above, resource measures can take the form of cost-adjusted spending or direct measures of resources. There is very little merit to spending measures that do not include cost adjustments. On the one hand, although appropriate cost adjustments can be difficult to estimate, cost-adjusted spending measures may be preferred to direct resource measures

because they allow for differences in the allocation of funds to meet the needs of students, needs that may vary widely across contexts. On the other hand, cost-adjusted spending measures will not capture differences in school quality due to inefficiency (including corruption) in the use of dollars. Schools may appear to be of equal quality on cost-adjusted spending but may not be on direct measures of resources if they differ in how effectively they are able to use their money to buy resources. Direct measures of resources (such as the number of equal quality computers or equal quality teachers) would uncover some of the differences in quality due to ineffective use of funds, as long as the measures of resources appropriately identify the quality, as well as the quantity, of resources.

The quality of the school is a function not only of the quality of the specific resources but also of how those resources are utilized together; thus, direct measures of resources will not identify all quality differences across schools even if they are a better measure than cost-adjusted spending for assessing the quality of the individual resources. Observational measures of processes in schools aim to carefully measure quality differences not only in each specific resource but in how the resources are used. In theory, these observational measures would be a productive way to determine whether the quality of education provided is equal across schools. However, the technology for measuring the quality of processes is just emerging, and it is currently impractical to measure the quality of all of the key schooling processes. Although instruments are in development, it is likely to be many years before comprehensive measurement of schooling is feasible. Finally, the third measure, proximal student outcomes, may at first appear to have some advantages. As discussed further below, however, the difficulty in attributing outcomes to schools limits its usefulness for assessing whether the quality of education is the same across schools.

In summary, if the goal is to assess the extent to which schools are of equal quality, then some combination of the resources measures and processes measures would constitute the best approach in most situations.

Equity as equal opportunity

A second standard, equity as equal opportunity, goes beyond the “do no harm” principle and calls for schools to compensate for, or redress, background disadvantages that children bring with them to school so that all children have the opportunity to participate equally in the political and economic life of the community. According to this more demanding equity standard, social disadvantage would not be an excuse for differential outcomes. Because students come to school with different family backgrounds and different capacities to learn, the provision of equal quality schools would not meet this equal opportunity standard. Instead, some schools would have to offer higher quality schooling than others to compensate for the differences that society deemed as unacceptable contributors to unequal outcomes.

Because this standard focuses on educational outcomes, the relative merits of the three approaches to measuring educational quality change. In particular, the difficulties of attributing educational outcomes to schools become less problematic. Consider first an extreme — and admittedly unrealistic — example in which the community defines educational equity as equal outcomes for all students. In this case, the cause or attribution of the outcomes is irrelevant. All that matters is whether the outcomes differ, not why they differ. Of course, any limitations associated with the use of a set of outcome measures that do not reflect the full range of investment outcomes of actual interest remain a problem.

Defining educational equity in this extreme way of equal outcomes for all students may in fact, be an undesirable conception of equity, for the reasons emphasized by Amy Gutmann (1987). Children differ in their aptitudes and interests, and requiring them all to reach the same level on each outcome of interest is unrealistic. Moreover, the equalization of outcomes at the level of the individual would undoubtedly require a level of government intervention into family life, and perhaps into the gene pool, that most people would deem inappropriate

A somewhat more realistic conception of this equity goal would require equality of average outcomes across groups of students defined by their demographic characteristics (Roemer 1998). For instance, equity could require similar outcomes on average for males and females, blacks and whites, southerners and northerners, or children from low income and more affluent households. This interpretation of the equity goal may require too much of schools alone to be fully achievable in practice, but it, at least, provides guidance about background differences for which schools would need to provide higher quality education and thus probably require greater inputs. Once again, though, because this equity concept is based on outcomes, the problem of attributing outcomes to schools is less salient than when the equity standard is equal quality schools.

An equity concept that is based on educational outcomes, rather than on school inputs, is appealing because it aligns well with an ultimate goal of an equitable distribution of outcomes such as income and health. Such an equity concept, however, is not by itself very useful to educational policy makers. Because many factors in addition to schooling contribute to educational outcomes, policy makers need information on the schooling component if they are to make wise policy decisions that balance schooling, income

redistribution, housing programs, individual incentives, and other potential approaches to equalizing individual or group outcomes. As such, some direct measures of education quality – in the form of cost-adjusted spending measures and observational protocols – are warranted in combination with outcomes-based measures.

Equity as adequacy

Yet a third standard of educational equity shifts the focus away from equality to the sufficiency or adequacy of the education system. According to an adequacy standard, every child should receive a level of education quality sufficient to achieve some specified goal or goals. Once that standard were met, it would be acceptable, from a normative perspective, for some children to receive a far higher quality than others. Thus, adequacy need not require equality.

The concept of adequacy can be applied either directly to educational inputs or, as more common in practice, to educational outcomes. As applied to inputs, the standard might require schools to have qualified teachers and manageable class sizes, and to provide safe and healthy learning environments. As one part of the first phase of the New York Campaign for Fiscal Equity (CFE) case, for example, the court specified that children are “entitled to minimally adequate physical facilities and classrooms,” “reasonably current textbooks,” an adequate number of qualified teachers, and schools in which “reasonably up-to-date curricula such as reading, writing, mathematics, science, and social studies” are being taught according to “minimally adequate” standards (Campaign for Fiscal Equity v. State of New York, 295 A. D. 2d 1 2002: 317). Measuring educational quality for the purposes of assessing compliance with this equity standard would clearly require

information on the specific input measures of interest, including information on teacher quality, and might well require some attention to school processes. The main focus would be only on whether schools did or did not meet the threshold. The outcome approach to measuring educational quality would be neither useful nor necessary.

The CFE case also defined adequacy in terms of outcomes for students, as has been done in many other court cases. Interpreted in terms of outcomes, the adequacy standard raises the central question of, “adequate for what?” One answer might lie in the Rawlsian concept of primary goods and the notion that every student attains a minimum set of educational outcomes connected to his or her long-term life chances (Rawls 2001: 57-61). Another might draw on philosopher Amy Gutmann’s concept of a democratic threshold. In her view, the primary role of education is to promote a democratic society, characterized by deliberative and collective decision-making, and hence the threshold is that level at which a person has the ability to participate effectively in the political process (Gutmann 1987; see also discussion in Ladd and Hansen 1999: 102-06). Combining these two views, an adequate education may be conceived of as one that is sufficient for someone to participate fully in both the economic and political life of the country (see also Allen, forthcoming, for such a combined view).

In general the definition of educational adequacy would allow for disparities above the adequate threshold. To the extent that education is viewed as a “positional good,” however, adequacy defined in terms of outcomes becomes more complicated and, in fact, may require that educational outcomes be equalized. A positional good is one in which one’s position in the queue matters for one’s outcome. In other words, “the absolute value of the good one holds, to the extent it is positional, can only be determined by referring to

one's standing in the distribution of that good." (Koski and Reich 2008: 45; Reich this volume). Hence, if education is viewed as a positional good, the only way to assure that everyone gets an adequate education is to make sure that educational outcomes are similar.

Regardless of whether education is or is not viewed as a positional good, judging the adequacy, defined in terms of outcomes, of an education system raises most of the same issues for the measurement of education quality as does the equal opportunity standard. Specifically, because the focus is outcomes, one need not necessarily isolate the contribution of schools to student outcomes to measure the quality of the system. Thus, while equity as equal opportunity and equity as adequacy differ conceptually, the use of proximal outcome measures of quality make them quite similar in terms of the relative merits of different approaches to measuring quality. While U.S. policy makers have not pursued equality of educational outcomes, broadly defined, as a serious policy goal, both the courts and the federal government have embraced outcome adequacy with various degrees of ambition. Some state courts, including the Kentucky court in a 1989 adequacy case, conceived of the outcome goals very broadly.² In practice, however, adequacy is typically defined more narrowly. In the New York Campaign for Fiscal Equity case (see Smith this volume), for example, the issue was whether an eighth grade education was adequate (though this position lost and a much higher standard was adopted), and under the federal No Child Left Behind Act adequacy is defined narrowly in terms of proficiency on math and reading tests.

As is the case for the equal opportunity standard, while proximal outcome measures may be necessary and useful for assessing the adequacy of an education system, that

approach alone is not sufficient for educational policy makers because it sheds no light on the contributions of schools relative to other sources to the generation of the outcomes of interest. Even with an adequacy standard, therefore, some combination of the three approaches to assessing school quality would be needed.

Section 5: Conclusion

Education quality has been and continues to be the focus of policy debates, as well as academic discussions within a range of disciplines from political theory to economics and sociology. Yet, the definition of education quality in these deliberations is often hazy, relying on examples of spending patterns or patterns of specific resources available across schools, observations in schools, or student test performance. None of these approaches to measuring education quality is perfect; each brings with it advantages and disadvantages. In this chapter, we begin with a definition of education quality as the benefits of education to students and other members of the community – benefits that can be described as consumption and investment benefits, intrinsic and extrinsic benefits, private and public benefits, or in a variety of other ways. Higher quality schools provide more benefits than do low quality schools.

None of the available measures of education quality perfectly capture these benefits of schooling, at least in part because most education benefits accrue long after students leave school. School spending adjusted for cost differences is an appealing measure of quality because it is easy to understand and, as a ratio scale, it allows for quantitative comparisons such as one school spending 25 percent more than another school.

Unfortunately, because of meaningful efficiency differences as well as unknown cost differences across schools, school spending poorly approximates school quality. Direct measures of education practices can account for efficiency and cost differences in ways spending measures cannot. Unfortunately, validated observational measures are available for only a very limited number of school processes which is especially problematic in the quickly changing and diversifying education sector. Using student test performance as a measure of school quality eliminates the need to understand or agree on the best way to teach or to run schools. It also is not affected by rapidly changing approaches to schooling. Thus, measuring education quality by student test performance has benefits. However, while current test performance is predictive of future job market opportunities, it does not capture the full range of benefits of value. In addition, attributing the portion of test performance that is due to schooling in contrast to educational opportunities outside of school is not easy.

Given that none of the available measures of school quality fully (or sufficiently) capture actual quality, which measure is most useful? The answer to this question depends on the reason for measuring education quality. If the goal is equal or adequate education outcomes for students, as would be consistent with the normative standards of equal educational opportunity or adequacy, measuring quality in terms of student outcomes may be appealing. The reason is that determining whether students achieve specified outcomes is more important than correctly attributing quality to particular schools. Even in the context of this relatively compelling case for using an outcomes-based measure, however, the limited range of measurable contemporaneous outcomes for students makes it useful to supplement the outcome measure of quality with information on school resources and

processes. In addition, resource and process measures can help policy makers choose where to target investments and interventions. Thus, while student outcome measures are clearly useful for understanding the extent to which equal opportunity or adequacy goals are achieved, they are best supplemented with other quality measures.

If the normative goal is equal quality schools, so that a student (and society) receives the same level of benefits regardless of which school that student attends, then attribution of outcomes to specific schools is of great importance. Consequently, the advantages of resource and process measures of education quality increase relative to measures based on student outcomes. The point is that all three approaches to measuring quality are imprecise and each has its own strengths and weaknesses. All three are potentially useful but how much emphasis to put on one approach relative to another differs depending on the normative standard of interest.

1. The size of the decline between 1972 and 2000 depends on the measure of variation. The gini coefficient declined by 20 percent and the coefficient of variation (defined as the standard deviation across districts divided by the mean) declined by 24 percent (Corcoran and Evans, 2008: Table 19.2).

2. *Rose v. Council for Better Education* defined adequacy in terms of seven learning goals: (1) sufficient oral and written communication skills to enable students to function in a complex and rapidly changing civilization; (2) sufficient knowledge of economic, social, and political systems to enable the student to make informed choices; (3) sufficient understanding of governmental processes to enable the student to understand the issues that affect his or her community, state, and nation; (4) sufficient self-knowledge and knowledge of his or her mental and physical wellness; (5) sufficient grounding in the arts to enable each student to appreciate his or her cultural and historical heritage; (6) sufficient training or preparation for advanced training in either academic or vocational fields so as to enable each child to choose and pursue life work intelligently; and (7) sufficient levels of academic or vocational skills to enable public school students to compete favorably with their counterparts in surrounding states, in academics or in the job market. (http://www.schoolfunding.info/states/ky/lit_ky.php3).