

**NBER WORKING PAPER SERIES**

**SCHOOLING, LABOR FORCE  
QUALITY, AND ECONOMIC GROWTH**

**Eric A. Hanushek  
Dongwook Kim**

**Working Paper 5399**

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

**We have benefitted from comments and discussions with Mark Bills, Stanley Engerman, Michael Wolkoff, and participants in the Applied Economics Workshop at the University of Rochester. This is a revised version of RCER Working Paper No. 411, September 1995. This paper is part of NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.**

**© 1995 by Eric A. Hanushek and Dongwook Kim. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.**

SCHOOLING, LABOR FORCE  
QUALITY, AND ECONOMIC GROWTH

**ABSTRACT**

Human capital is almost always identified as a crucial ingredient for growing economies, but empirical investigations of cross-national growth have done little to clarify the dimensions of relevant human capital or any implications for policy. This paper concentrates on the importance of labor force quality, measured by cognitive skills in mathematics and science. By linking international test scores across countries, a direct measure of quality is developed, and this proves to have a strong and robust influence on growth. One standard deviation in measured cognitive skills translates into one percent difference in average annual real growth rates—an effect much stronger than changes in average years of schooling, the more standard quantity measure of labor force skills. Further, the estimated growth effects of improved labor force quality are very robust to the precise specification of the regressions. The use of measures of quality significantly improves the predictions of growth rates, particularly at the high and low ends of the distribution. The importance of quality implies a policy dilemma, because production function estimates indicate that simple resource approaches to improving cognitive skills appear generally ineffective.

Eric A. Hanushek  
Wallis Institute of Political Economy  
University of Rochester  
Rochester, NY 14627-0158  
and NBER

Dongwook Kim  
Wallis Institute of Political Economy  
University of Rochester  
Rochester, NY 14627-0158

# Schooling, Labor Force Quality, and Economic Growth

by Eric A. Hanushek and Dongwook Kim\*

## 1. Introduction

Recent analyses of international differences in growth rates has focused attention on the role of human capital in development. Most cross-country empirical studies of long run economic growth now include some proxy for human capital, regardless of the specific theoretical model motivating the analysis. Data limitations have, however, forced severe compromises in the ensuing analyses. Perhaps as an indication of the importance of human capital, almost any of the diverse set of proxies will appear to be significantly related to performance of national economies. Nonetheless, if the estimates are to be taken seriously from a policy perspective, it is important to understand better which of the potential explanations is key. The analysis in this paper suggests that cognitive skills are an important component of relevant variations in human capital, reinforcing the attention to school quality found in many countries today.

Previous investigations of growth have concentrated on a limited set of proxies for current schooling activities. The most frequently employed measure is either the primary- or secondary-school enrollment rate, used, for instance, in Barro (1991), Romer (1990b), Mankiw et al.(1992). Levine and Renelt (1992) and Levine and Zervos (1993), in their influential sensitivity studies, enumerated initial secondary-school enrollment rate as part of

---

\*We have benefitted from comments and discussion with Mark Bils, Stanley Engerman, Michael Wolkoff, and participants in the Applied Economics Workshop at the University of Rochester. This is a revised version of RCER Working Paper No. 411, September 1995.

the set of core variables always included in the extreme-bounds analysis of cross-country growth regressions.

One empirical problem with using school enrollment rates is that these schooling flow variables may not accurately represent changes in the relevant stock of human capital of the labor force, particularly during periods of educational transition. Moreover, simultaneity resulting from the possibility that these human capital investment flows could be caused by high growth rates, confuse their interpretation in simple cross-country growth models. Finally, more recent endogenous growth models emphasize the stock of human capital as the relevant input into invention and technological change. To deal with these problems, Barro and Lee (1993) pioneered an alternative approach. They calculated average years of schooling for 129 countries quintennially from 1960 to 1985 by filling in missing values to observed data points from various census/survey data, thus enabling a closer alignment of empirical analysis with theory.

A challenging problem of this new alternative, however, comes from the lack of adjustment for schooling quality. Few people, for example, would believe that a year of secondary schooling in the United States was equivalent to a year at the same grade level in Egypt. Indeed, Barro (1991) explored the inclusion of differences in real school resources as a crude measure of quality differences across countries in his growth regressions. He found that student-teacher ratio in primary schools in 1960 had a negative relationship to economic growth, while that in secondary was statistically insignificant with positive sign—giving little faith that these adequately captured any quality differences in schools.

Measuring schooling quality has been difficult and controversial. As discussed below, standard measures of schooling quality such as pupil-teacher ratio, class size, and teacher characteristics do not effectively explain student cognitive achievement. Their linkage to future labor market performance has also been questioned. This paper addresses the measurement problem of labor force quality directly. Rather than concentrating on conventional measures of schooling inputs, we construct new measures of quality based on student cognitive performance in various international tests of academic achievement. It can be regarded as a complement to Barro and Lee (1993), adding a series of worker quality to their schooling quantity measures.

These new measures are incorporated into the analysis of average growth rates of per capita real GDP in the framework of cross-country growth regression, prototyped by Kormendi and Meguire (1985) and popularized by Barro (1991). When added to initial GDP and Barro and Lee (1993)'s schooling quantity, the quality measures are found to be significant and quite robust in the sense of Levine and Renelt (1992). Both schooling quantity and quality, along with the initial income, explain about 40 percent of the variation in growth rates.

Direct observations of quality are available for thirty-nine countries which participated in international assessments of student achievement at least once. These quality measures are extended to other countries in the Summers and Heston (1991) data set by filling in missing values from international test score regressions. In the extended data set covering non-participating countries as well as participants in international comparison studies, the same qualitative conclusion holds: quality is significant with the correct sign even when quantity

tends to lose significance. The main results are used to find which countries are faster and slower growers than would be expected given their initial income and human capital, deepening understanding of the different growth experiences across countries.

## **2. The measurement of labor force quality**

Treating all workers as identical is a compromise frequently forced on researchers by data shortcomings. Initial empirical growth modeling (e.g., Solow(1957)) relied on either number of workers or man-hours to characterize labor input. Early refinements in the U.S. growth accounting literature, however, quite quickly moved to adjusting labor to account for quality differences. Initial efforts characterized the qualitative differences among workers in terms of years of schooling and length of the school year and estimated the contribution of schooling to productivity growth through separate estimates of how much schooling contributed to individual earnings (Denison (1962, 1972, 1993), Abramovitz (1962)). Similar approaches from the viewpoint of decomposition of residual growth were introduced by Jorgenson and Griliches (1967, 1972). This early growth accounting methodology was also applied to international comparisons of growth, and again the contribution of labor was modified to incorporate education changes (Denison(1967), Engerman(1971)).

These developments in modeling growth complemented and built upon studies of labor markets which highlighted the importance for individual productivity and wages of human capital, and particularly that related to formal schooling (Schultz (1961), Becker(1993), Mincer(1970, 1974)). Standard wage determination models invariably include quantity of

schooling as a key characteristic distinguishing workers with different productivity.<sup>1</sup> Perhaps the most complete extension of this work with application to aggregate growth determination is Jorgenson and Fraumeni (1992). They calculate the human capital stock from analysis of lifetime earnings (by education, age, and sex categories) and, going even further, adjust for the value of nonmarket activities.

Both the U.S. and the international analyses nonetheless have recognized that pure quantity of schooling, while important, is a crude measure of skill differences. Various strategies, largely originating in analyses of individual wage determination, have been employed to improve the measure of labor beyond simple quantity. Two fundamental approaches are available: addition of measures of schooling inputs such as expenditure or teacher salaries that are presumed related to school quality and direct measures of cognitive skills of individuals, often interpreted as a measure of schooling outcomes. The differences between these approaches, and especially the differences in their conclusions, have been the subject of considerable recent discussion within the context of U.S. labor markets and schooling policy, and the results of that discussion are relevant to these considerations.

The motivation behind inclusion of input differences in schooling is simple. Variations in school quality imply that a year of schooling gained at a different time, a different country, or even a different school may not be equivalent to a year in different circumstances. School systems are known to vary widely across countries in terms of organization, resources, and the preparation of students coming to schools. Spending per

---

<sup>1</sup>On-the-job training (OJT), the other major subject of early human capital analyses, is ignored here since its contribution to growth would come from intertemporal changes in the experience structure of the work force. Even though OJT is clearly important, the changes in the overall demographic (and training) profile would be presumed small and nonsystematic.

student varies across the world by a factor of 100. Behind this, schools range from outdoor meeting places under a shade tree with a teacher who herself has only primary school training to modern, well-equipped facilities with teachers possessing master's and doctor's degrees. Typical school years likewise vary from under school 100 days to over 200 days. Given these variations, it would be surprising to find that average years of schooling completed was a good proxy for the amount of human capital of the labor force. To account for such differences, the direct approach is to incorporate variations in spending or real resource flows such as the pupil-teacher ratio into models of earnings determination.

A simple model of the underlying structure helps clarify the interpretation. The common starting point is a model of wages as skill rents such as:

$$(1) \quad w = w(H) + u$$

where  $w$  is wages,  $H$  is human capital, and  $u$  is unmeasured other differences. Human capital, in a somewhat more general form than typical, can be written as:

$$(2) \quad H = h(S, Q, F, A) + v$$



where  $S$  is quantity of schooling,  $Q$  is school quality,  $F$  is family and other exogenous factors influencing skills,  $A$  is individual-specific fixed factors (“ability”), and  $v$  is other unmeasured influences on human capital.<sup>2</sup> Finally, school quality is given by

$$(3) \quad Q = q(R, O) + e$$

where  $R$  is resources available to schools,  $O$  is the organizational structure of schools, and  $e$  is unmeasured factors affecting school quality.

The simplest version of wage models, which form the starting point for most analyses, estimate (log) wages as a linear function of years of schooling:

$$(4) \quad w = r_0 + r_1 S + u'$$

The error term in (4) includes the various exogenous influences on individual skills, as found in (2) and (3). The desired interpretation is that  $r_1$  represents the rate of return to a year of schooling, but, even if the errors in equations 1-3 are i.i.d.,  $S$  and  $u'$  will in general be correlated. Thus, OLS estimation of (4) will yield biased estimates of  $r_1$ .

Some variant of this structure motivates corrective procedures that have been applied. Perhaps the simplest procedure but one not closely related to any data is that of Denison(1967). To allow for family inputs and aptitude differences, he simply lowered the

---

<sup>2</sup>This formulation takes the production function perspective, but generalizes from that commonly found in investment considerations. See Ben-Porath(1970), Hanushek(1979, 1986).

cross-sectionally estimated value for  $r_1$  by 40 percent.<sup>3</sup> This procedure was applied uniformly to United States and to European growth accounting.

An alternative approach found in wage determination models modifies (4) by adding estimates of the school resources,  $R$ , devoted to workers. This approach, found in Johnson and Stafford(1973) and Wachtel(1976), was focused directly on the issues of biases in  $r_1$  in the innovative international analysis of Behrman and Birdsall(1983). The attention to wages and school resources has emanated most recently from the expanded U.S. analysis in Card and Krueger(1992).<sup>4</sup> These analyses as a group have suggested that individual wages are directly associated with quantity of schooling and with the resources devoted to schools, thus indicating an approach for introducing directly quality differences in workers human capital.

In large part the attention to the analyses of school resources has come from its differing conclusions from the direct analysis of education production functions. Of the nearly 400 separate estimates of models paralleling (3), there is little evidence that school resources are consistently associated with student performance (Hanushek(1986, 1989), Hanushek, Rivkin, and Taylor(1995)). The leading candidate to explain the lack of relationship is the importance of organization and the implied incentives, as captured in Eq.

---

<sup>3</sup>His actual procedure was to create indices of schooling from average wage differentials by race, sex, and age. These differentials were then reduced by 40 percent. The support he cites for this procedure does not involve direct estimation of the effects of omitted factors; nor does it justify the same adjustment across countries (Denison(1967)).

<sup>4</sup>These analyses do differ in their specification of how school resources enter. For example, Stafford and Johnson and Wachtel employ measures of expenditure while Card and Krueger and Behrman and Birdsall use measures of specific resources such as pupil-teacher ratios, teacher education, and teacher salaries. They also differ in the precise functional form and in whether or not they include measures of postsecondary schools or not. These distinctions do not appear crucial for the overall interpretation.

3, leading to significant inefficiency in the use of resources. The lack of incentives to improve student with unmeasured variation in organization (O) implies that there is no consistent effect from varying expenditure or resources (Hanushek with others(1994)). The conflicting evidence between the two types of studies has also generated new studies that shed some light on the issues. At least for the United States in more recent periods, any relationship between school resources and wages appears tenuous (see Betts(1994, forthcoming), Grogger(forthcoming), Johnson and Stafford(1995)). Moreover, the Card and Krueger results are very sensitive to sample and specification (Heckman, Layne-Farrar, and Todd(1994), Speakman and Welch(1995)). Finally, the interaction of specification and aggregation resulting from the common approach of employing measures of school resources at the state level may result in increased bias that could explain the apparent significance of estimated resource effects (Hanushek, Rivkin, and Taylor(1995)).

These arguments are not conclusive, however, when put in an international context. First, much of the variation in expenditure from the U.S. studies comes from inclusion of workers educated in the 1920s and 1930s, when real expenditure was considerably less than today (Hanushek and Rivkin(1994)). The Behrman and Birdsall (1983) results also reflect the lower level and wider variation in school resources found in Brazil in the 1970. If the apparent effectiveness of school resources is driven by such variation, measures of school resources may capture cross-national variations in quality even if currently not true in the United States. Second, some direct analysis of international achievement differences supports the importance of variations in resources (Heyneman and Loxley(1983)), even though this finding is far from consistent (Harbison and Hanushek(1992), Hanushek(1995)).

Note, however, that even if school expenditure and resources were related to wages, their interpretation in wage and in growth determination is ambiguous. From the production model (3), school quality is only one determinant of human capital differences. Family and other nonschool inputs affect individual learning and skill acquisition, as does the ability or motivation of individuals. Since resources are likely to be correlated with these omitted factors, any observed correlation of wages and resources is likely to be overstated from a causal perspective.<sup>5</sup> This, as discussed later, is particularly important in analyses of cross-country growth differences, because these analyses frequently enter directly into policy discussions (see, for example, World Bank(1993)).

The alternative approach to measure labor force quality has been the inclusion of measures of cognitive skills to capture systematic individual differences. The early investigations of this were aimed at identifying the bias in the estimated return to quantity ( $r_1$  in Eq.4) that came from ability differences. In other words, family (F) and school quality (Q) were ignored and test scores entered into estimation of wage equations were interpreted as exogenously determined ability (e.g., Griliches(1974)). A more reasonable interpretation is that the test scores measure significant variations in H that are in addition to dimensions captured by pure school quantity.

An important issue, however, is whether or not test scores provide good measures of qualitative differences among workers or of the important dimensions of human capital for

---

<sup>5</sup>These considerations may differ between standard wage determination models and growth models. Wage determination models estimated with individual data seldom include attributes of families, so that school resources in part proxy other inputs and their effects are unidentified. In growth models, when aggregate quantity of schooling is included, the aggregate education of parents is implicitly included—thus lessening problems of omitted family factors. The interpretation of the aggregate quantity of schooling then includes its effects on the education of subsequent generations.

growth. Much of the discussion of the role of human capital in technological change and growth stresses cognitive skills of workers. But, it is at the same time true that common achievement tests are not developed to measure either innovative talents or labor market skills. Early labor market studies found relatively little earnings impact of variations in tests. The most recent direct investigations of cognitive achievement, however, have suggested generally larger labor market returns to individual differences. For example, Bishop(1989, 1991), O'Neill(1990), Ferguson (1993), Grogger and Eide (1993), and Murnane, Willett, and Levy(1994) each find that the earnings advantages to higher achievement on standardized tests are quite substantial. These results are derived from quite different approaches. Bishop(1989) emphasizes the measurement errors that are inherent in most testing situation and demonstrates that careful treatment of that problem has a dramatic effect on the estimated importance of test differences. O'Neill (1990), Ferguson(1993), Grogger and Eide (1993), and Bishop(1991) on the other hand, simply rely upon more recent labor market data along with more representative sampling. They suggest that the earnings advantage to measured skill differences is larger than that found in earlier time periods and in earlier studies even without correcting for test reliability—perhaps reflecting increased demand for highly skilled workers. Murnane, Willett, and Levy (1994), considering a comparison over time, demonstrate that the results of increased returns to measured skills hold regardless of the methodology (i.e., whether simple analysis or error-corrected estimation). Bishop(1989) is particularly relevant because of its direct investigation of the importance of cognitive skills for productivity and growth.

The NAS/NRC study on employment tests (Hartigan and Wigdor 1989) also supports the view of a significant relationship of tests and employment outcomes, although the strength of the relationship appears somewhat less strong than that in the direct earnings investigations. It considers the relationship between the General Aptitude Test Battery (GATB), the standard employment test of the Department of Labor, and job performance. Their synthesis of a wide number of studies suggests a systematic but somewhat modest relationship with correlations to performance on the order of .2 to .4. The analysis also finds that the validity of these tests in predicting performance has gone down over time. These results, being at somewhat odds with the recent studies, may simply reflect the specialized nature of GATB.<sup>6</sup>

While past analyses tend to conclude that direct measures of cognitive skills are superior to measures of schooling inputs, ambiguities remain. Our approach is to develop a consistent series for cognitive achievement and then to examine the empirical relevance of this and of school input measures in explaining variations in growth rates across countries. Further, even if test scores are accepted as a measure of human capital, important policy questions remain. Specifically, if the relationship between resources and school quality is not well understood, any policies to improve human capital through schools and quality changes are more uncertain. This issue is discussed below.

---

<sup>6</sup>The GATB is a very old test that may not reflect changes in the economy. It also suffers from some psychometric problems (see Hartigan and Wigdor 1989). The central purpose of the study was assessment of the Department of Labor practice of providing test information normed to racial groups.

### **3. International comparisons of cognitive achievement**

The comparison of cognitive achievement across countries relies on a series of tests conducted over the past three decades. While a varying set of countries has participated over time, sufficient data exist to permit splicing the performance series together into a measure of worker quality. The first part of this section describes the history of international testing, while the second details the construction of the summary quality measures used in modeling growth.

#### *A. History and Extent of International Testing*

The most important testing has been six international tests to assess student achievement in the fields of mathematics and science. Four were administered by the International Association for the Evaluation of Educational Achievement (IEA) and two by International Assessment of Educational Progress (IAEP). The IEA, since its establishment in 1959, has a long and unique role of specializing in comparative education research for almost all aspects of primary and secondary education. On the other hand, the IAEP, starting in 1988, builds on the statistical techniques and procedures developed in the United States for the National Assessment of Educational Progress (NAEP). (The NAEP has been the main national testing instrument in the United States since 1969). While the IAEP is geared to the U.S. curriculum, the IEA has an international focus not associated with the curriculum in any particular country.

The first IEA mathematics study was conducted in 11 countries during 1963 and 1964.<sup>7</sup> More than 133,000 students, 18,500 teachers, and 5,450 schools participated in the study. Two age groups took the tests: age 13 and the age appropriate for the 12th grade. Twelfth grade students were also divided into two groups of students, one taking mathematics and the other not taking any mathematics at the time of tests. One defect of the study is that the details of sampling procedures and their implementation are sparse, leading to some questions about the quality of these data. The more ambitious second IEA mathematics study covered 20 educational systems in 17 countries during 1980-82.<sup>8</sup> Thirteen-year-old students were tested in five subfields of mathematics, and students in the last year of secondary schools were tested in six subfields. Substantial improvement was made in terms of sampling quality by the second IEA test.

The first IEA science study was conducted from 1966-73 as a part of a broader research project, the so-called Six Subject Survey.<sup>9</sup> Among the seventeen participating countries were four developing countries, an innovation that is important in considering varying growth experiences. Tests were given to students aged 10, aged 14, and students in

---

<sup>7</sup>For details, see Húsen (1967). The participating countries are: Australia, Belgium, Finland, France, W. Germany, Israel, Japan, Netherlands, Sweden, United Kingdom, and United States. On this and subsequent testing, each country may have several participating educational systems, mainly based on regional or lingual basis. The multiple observations characterize data from Canada, Belgium, and United Kingdom and are the subject of special consideration below.

<sup>8</sup>See Robitaille and Garden (1989). Participating countries were Belgium, Canada, Finland, France, Hong Kong, Hungary, Israel, Japan, Luxembourg, Netherlands, New Zealand, Nigeria, Swaziland, Sweden, Thailand, United Kingdom, and United States.

<sup>9</sup>See Comber and Keeves (1973). The planned assessments included science, literature, reading comprehension, English and French as foreign languages, and civic education. Participating countries included Australia, Belgium, Finland, France, W. Germany, Hungary, Italy, Japan, Netherlands, New Zealand, Sweden, United Kingdom, United States, Chile, India, Iran, and Thailand.



the last year of their secondary education program. Similarly, the second IEA science study in 1983-86 included the same three age groups in 23 countries.<sup>10</sup>

The IAEP was an independent research project. NAEP assessment procedures and technology were used to save cost and time and ensure standardized implementation of test procedure across countries. The first IAEP study was experimental, covering twelve school systems in six countries and administered in 1988.<sup>11</sup> A total of 24,000 students, age 13, were assessed in mathematics and science. The second IAEP study was administered in 1991 for two age groups, age 9 and age 13, in thirty-four school systems across nineteen countries.<sup>12</sup> As in the first IAEP study, students were tested for mathematics and science.

Table 1 shows the breakdown of countries by frequency of participation in both IEA and IAEP studies. Out of total 39 countries that have ever participated in international science and mathematics achievement studies, nineteen countries participated only once, and five participated twice. On the other end, nine countries were involved in studies four times and more. The United States and the United Kingdom are the only countries taking all international tests. Finland and Sweden participated in all IEA tests; Canada, Ireland, Korea, and Spain took all IAEP tests. The number of separate observations is actually greater, since

---

<sup>10</sup>See Postlethwaite and Wiley (1992). Among the 23 countries, published data are consistently available for 17 countries : Australia, Canada, China, Hong Kong, Hungary, Italy, Japan, Korea, Netherlands, Norway, Philippines, Poland, Singapore, Sweden, Thailand, United Kingdom, and United States.

<sup>11</sup>See Lapointe, Mead, and Phillips (1989). Participating countries were Canada, Ireland, Korea, Spain, United Kingdom, and United States. Multiple school systems (by geography and language) were included for Canada and the United Kingdom.

<sup>12</sup>See Department of Education(1994). Participating countries included Brazil, Canada, Chile, France, Hungary, Ireland, Israel, Italy, Jordan, Korea, Mozambique, Portugal, Slovenia, Spain, Soviet Union, Switzerland, Taiwan, United Kingdom, and United States.

Table 1  
 Frequency of Test Participation  
 (39 countries)

Frequency of participation	Number of countries
1	19
2	5
3	6
4	7
5	-
6	2

Table 2  
 Number of Observations by Age/Subtest/Year  
 (39 countries)

Number of Individual Observations	Number of countries
1	1
2	5
3	3
4	4
5	6
6-10	5
11-15	8
16-20	4
21-26	3

all but one international study included either two or three age groups taking tests on several subfields in mathematics or science. Table 2 shows the breakdown of countries by the number of observation after considering test subjects and age groups.

The concentration on mathematics and science corresponds to the theoretical emphasis on the importance of research and development activities as the source of growth (e.g., Romer 1990a). Able students with a good understanding of mathematics and science are a pool of future engineers and scientists. Bishop (1992) provides separate confirmation of the importance of mathematics in determining individual productivity and income. Additionally, while some test information exists for other subjects, it cannot be compared readily with the mathematics and science scores.<sup>13</sup> Therefore, it is not used here.

### *B. Construction of Quality Measures*

The overall objective is to develop a single measure of quality for each country. To do this, we combine all of the available information on mathematics and science scores for each country. This approach clearly could lead to mismeasurement when school systems change significantly in quality over time. Nonetheless, since the ultimate goal is obtaining information about the quality of the labor force which itself is a mixture of workers receiving their schooling at different points in time, this averaging may be justified even with changing educational quality. Moreover, while data are hard to come by, the usual presumption is that

---

<sup>13</sup>Reading literacy assessments, for example, are available for 30 countries in 1991 (Department of Education(1993)). The significant differences in test domains makes construction of single measures impossible.

quality of schooling systems evolves slowly over time, in part related to the stationarity of the technology and to the slow turnover of teachers and other personnel.

The underlying test score data present problems of comparability.<sup>14</sup> Not only are different subtests available, but results are reported in a variety of forms: (1) number of items correct; (2) percent correct; and (3) scores in proficiency scales.<sup>15</sup> Because of availability across most of the test reports, we first transformed all data into a “percent correct” format.<sup>16</sup>

Two approaches were taken to adjusting the separate tests for differing degrees of difficulty. As shown in Table 2, twenty-six performance series—reflecting different age,

---

<sup>14</sup>The data set for this paper was constructed by collecting relevant data from various places. The main source of data is the compilation in Medrich and Griffith (1992), from which the results of the first and second IEA mathematics and the second IEA science studies are available. For the first IEA science study, Comber and Keeves (1973) was used directly to extend the number of countries beyond available in Medrich and Griffin. Lapointe, Mead, and Phillips (1989) provides scores for the first IAEP study. The results of the second IAEP study can be found in *Digest of Education Statistics (1992)* published by the U.S. Department of Education. Also the *Condition of Education (1994)* provided scores for the various NAEP studies in the United States..

<sup>15</sup>Proficiency scales were developed, using item response theory, to facilitate comparison of results across different tests and participants. The proficiency scores are obtained through applying complicated, nonlinear transformations to original scores in order to allow for differential difficulty of some items. Scores are then placed on an externally set scale, say one constructed to generate a mean of 500 and a standard deviation of 100 over the range of 0 to 1000 for the population. (In NAEP studies, scales have a mean of 250 over the range of 0 to 500). While having conceptual appeal for many purposes, they cannot be used as the basis of this analysis because it is not possible to transform all of the available tests into scale scores.

<sup>16</sup>For example, in the second IEA science study, data are published in the “number of items correct” format for 10-year-old and 14-year-old students and in the “percent correct” format for biology, chemistry, and physics tests which were taken by student in the last year in secondary education. These data are reproduced in the *Digest of Education* with the same inconsistent formats. In two IAEP studies, data are presented in proficiency scores. The U.S. Department of Education, however, has prepared data for the second IAEP study in the percent correct format (*Digest of Education (1992)*). Special care is needed to transform the standard deviation of performance as well as the mean (see below).

subtest scores, and years—are available (for varying subsets of countries), and these series have different mean percent correct for the countries participating on each. *Method 1* uses a multiplicative transformation to convert each performance series to a mean of 50. Thus, if the group mean percent correct for all taking test  $\tau$  is  $m_\tau$ , the standardized percent correct for each country is found by multiplying the nominal percent correct by  $50/m_\tau$ . (A similar transformation is employed for the standard error of each test,  $\sigma_\tau$ ). This transformation assumes an intertemporally constant world mean in science and mathematics performance and that the countries taking the tests are a random draw from the world distribution. Both of these assumptions are problematic. *Method 2* incorporates the additional information that is provided by the time series information from the National Assessment of Educational Progress (NAEP). At varying times from 1970 to the present, U.S. students aged 9, 13, and 17 have taken both mathematics and science tests. These tests, scored on a consistent basis, provide an absolute benchmark of performance to which the U.S. scores on international tests can be keyed. Thus, the mean for each international test series is allowed to drift in accordance to U.S. NAEP score drift and the mean U.S. performance on each international comparison.<sup>17</sup> The first IEA mathematics test in 1963 and 1964 does not have an NAEP comparison, because it pre-dated the NAEP. Therefore, two minor variants of this method were employed: using the first IEA test at the observed percent correct or simply discarding

---

<sup>17</sup>Since the U.S. took all tests in international comparison studies, the performance of U.S. students participating international tests can be transformed to mimic the performance of U.S. students in NAEP in the closest time and age groups. While NAEP data have been published in various (inconsistent) forms over time, we use data published in the *Condition of Education (1994)*, which presents data on a common scale. Scores of all other countries are adjusted proportionately on the basis of matched U.S. score adjustments.

these scores from the subsequent country calculations (while adjusting all other scores to the appropriate NAEP mean).

One difficulty arises when several education systems from an individual country independently participate in an international tests. To get a country-level mean score, we simply calculate an arithmetic average. There is no reliable way to calculate a standard error for the test in such situations, so we also use the arithmetic average for the standard error.<sup>18</sup>

The completed measures of schooling quality for each country are constructed by taking a weighted average over all available transformed test scores. The combining of tests exploits the information on score distribution for each test-country data point through weighting by the (normalized) inverse of the test standard error ( $\sigma$ ).<sup>19</sup> That is, the weight for the test  $\tau$  in country I,  $WT_{\tau i}$ , is :

$$WT_{\tau i} = (1/\sigma_{\tau i}) / \sum (1/\sigma_i).$$

This presumes that a high standard error conveys less accurate information about the country's performance.

Figure 1 gives a graphical depiction of the available test information over time. In this, scores for each age group and subtest are combined into a country score on each of the six underlying assessments (with the mean for each assessment set according to method 1 to

---

<sup>18</sup>In principle, the standard error should incorporate the covariances across the systems (which would generally be presumed to be positive). We had no way to estimate these, however, and thus ignored such considerations.

<sup>19</sup>When a standard error is missing, the group average over participating education systems is used for the country in question. Those countries are : France and Netherlands (12th grade non-math students, first IEA math study); Chile, India, Iran, and Thailand (first IEA science); Singapore (12th grade, second IEA science); Poland (physics test for 12th graders, second IEA science).

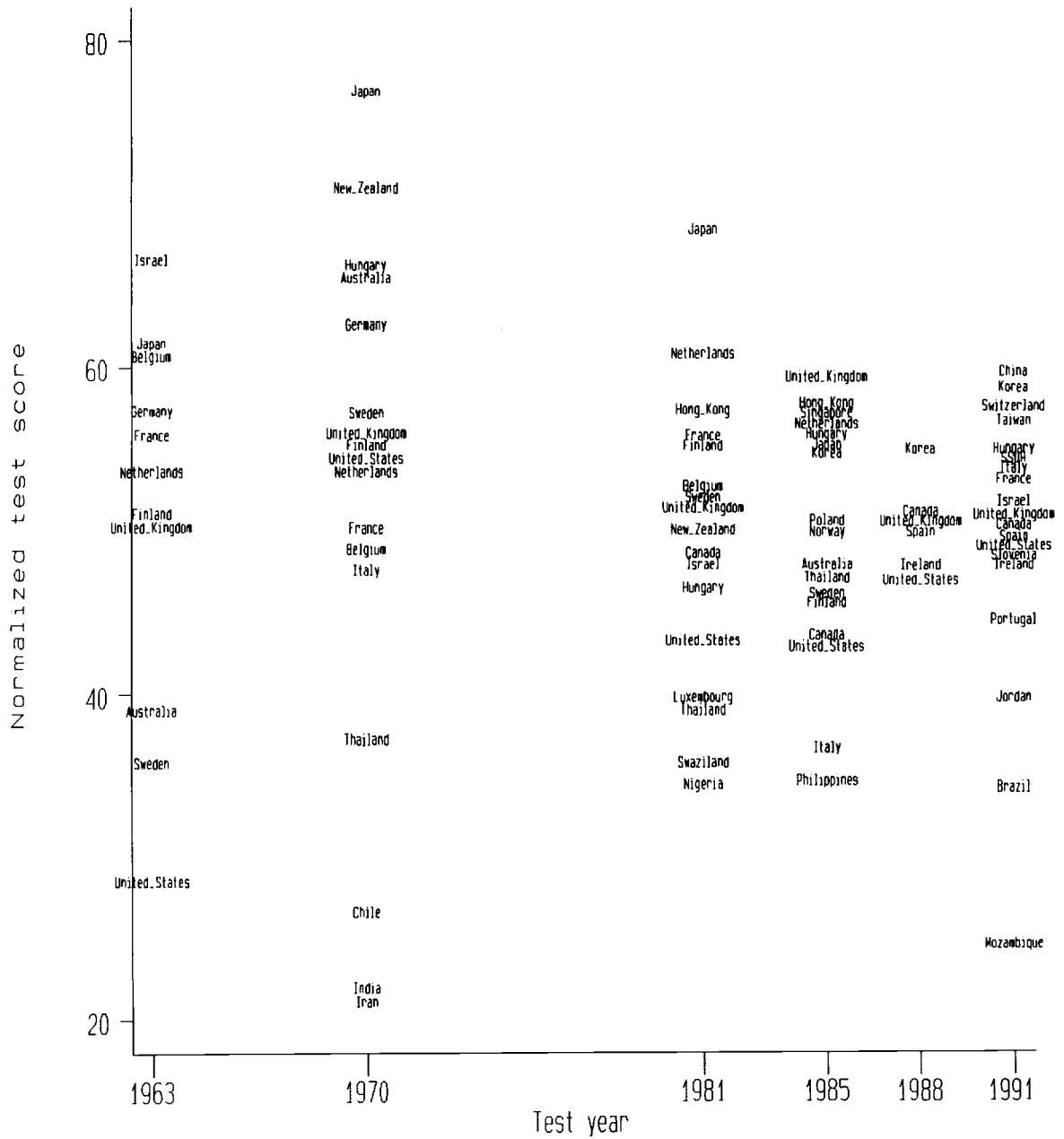


Figure 1. International Test Score Performance

50). The overall quality measures subsequently used for each country combine these scores across years to obtain a single measure.

Table 3 shows summary statistics for the three constructed quality measures. The first series, QL1, is based on world average scores of 50 for all six tests. QL2 calculates overall means for each of the six assessments that vary with the United States NAEP performance (and the specific U.S. performance on each international assessment). QL3 is constructed in the same way as QL2 except that the first IEA mathematics study was deleted. (Note that no country is lost since every participant on the first IEA assessment also participated in at least one subsequent assessment). The correlations among them are very high (more than 0.9). The reason for the high correlations is easily seen by tracing the position of the U.S. in figure 1 from 1970 to 1991. The NAEP scores for the U.S. fell during the 1970s and rose in the 1980s, just the pattern that is seen in U.S. performance on international tests. Because of the similarity, we often look at only one of them in the empirical exercise. A complete list of quality measures are presented in Appendix C.

#### **4. Quality effects on growth**

The analysis begins with comparisons of cross-country growth models that incorporate and that neglect variations in cognitive skills. The initial estimation employs just the observed quality measures developed in the previous section. Subsequent analyses expand the analysis to include estimated quality for a wider sample of countries.



Table 3  
Summary Statistics for Constructed School Quality Measures

Quality Measure	median	mean	standard deviation	minimum	maximum	number of countries
QL1	48.76	46.43	10.67	20.79	60.65	39
QL2	54.52	51.05	13.22	18.26	72.13	39
QL3	55.61	52.57	13.97	18.26	72.13	39

Simple Correlations

	QL1	QL2	QL3
QL1	1.00	.914	.935
QL2		1.00	.980
QL3			1.00

### *A. Baseline Estimates*

Table 4 reports the results of our baseline cross-country regressions. The focus is what affects growth in average real per capita GDP between 1960 and 1990 (Summers and Heston 1991).<sup>20</sup> This estimation, in the tradition of Levine and Renelt (1992), begins with a small set of determinants of growth rates and investigates the magnitude and stability of any influence of labor force quality, along with other frequently assessed factors. The simplest models for the sample of 100 countries with complete data estimate growth as a function of initial income (Y60) and the Barro/Lee<sup>21</sup> measure of school quantity (S) [col. 1] and these explanatory variables plus the annual growth rate in population times 100 (GPOP) [col. 2]. These base estimates give rather common results and explain about 20 percent of the variation in economic performance. The corresponding estimates with the addition of our alternative measures of labor force quality for the countries with test data are found in col. 3-6. TEST equals 1 if a country participated in any international comparison of student achievement studies and equals zero otherwise. Thus the formulation permits an intercept difference and growth effects related directly to quality.<sup>22</sup> White's (1980) variance-

---

<sup>20</sup>Some countries do not have 1990 GDP data, so growth rates are calculated to the nearest year to 1990 when 1990 is unavailable.

<sup>21</sup>The actual measure averages the separate Barro/Lee quantity figures across their five year periods.

<sup>22</sup>The model implicitly imposes a restriction that the coefficients of both initial income and schooling quantity regressor are the same across non-participants and participants in international comparison studies. Direct testing of equality of Y60 and S across the two parts of the sample does not reject equality when test scores are not included;  $F(2,94)=0.91$ ,  $\text{prob}=0.40$ . When the test scores are included, this hypothesis on Y60 and S is rejected at the 5 percent level but not at the 1 percent level;  $F(2,93)=4.04$ . The results when GPOP is included are similar. We use the pooled model, as that is consistent with the subsequent estimation of test scores for a wider set of countries.

Table 4  
 Baseline Estimates of 1960-1990 Cross-country Growth Models with Labor Force Quality  
 Dependent variable: Average annual growth rate in real per capita GDP ( $\times 100$ )  
 (100 countries)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
constant	0.933 (3.98)	2.179 (3.38)	1.293 (5.29)	2.276 (3.79)	0.854 (3.22)	1.279 (5.20)	1.291 (5.12)
Initial per capita income (Y60) [\$1,000]	-.407 (-2.97)	-.433 (-3.43)	-0.49 (-4.38)	-0.511 (-4.68)	-0.37 (-3.44)	-0.49 (-4.28)	-0.51 (-4.39)
Quantity of schooling (S)	0.529 (4.40)	0.445 (3.37)	0.363 (3.16)	0.300 (2.64)	0.442 (3.86)	0.365 (3.13)	0.373 (3.18)
Annual population growth (GPOP)		-.414 (-2.18)		-.332 (-1.76)			
Assessment available (TEST)			-2.825 (-2.99)	-2.629 (-2.60)	-4.401 (-4.48)	-2.069 (-2.36)	-1.816 (-2.10)
Labor force quality (TEST·QL1)			0.0948 (4.40)	.0902 (3.92)	0.182 (7.36)		
Labor force quality x schooling (TEST·QL1·S)					-0.010 (-5.05)		
Labor force quality (TEST·QL2)						0.0711 (3.99)	
Labor force quality (TEST·QL3)							0.0644 (3.74)
R-squared (adjusted)	0.211	0.230	0.365	0.376	0.422	0.352	0.343

Note: Estimates are corrected for heteroscedasticity according to White (1980). t-statistics are in parentheses below coefficients.

covariance matrix estimator is used to allow for heteroskedasticity across countries in calculating consistent t-values.

The basic models with achievement measures are very stable and have the expected signs. Initial income negatively affects growth, supporting notions of conditional convergence in growth rates.<sup>23</sup> When population growth is included, a one percentage point higher rate of population growth translates into a .3-.4 percentage point lower economic growth rate. As with various influences that have been estimated in the past, however, the causality is a concern, since higher income may lead a country to reduce its birthrate.

Both cognitive skills (QL) as well as schooling quantity (S) positively contribute to explaining variations in per capita growth rates. Adding the quality measures boosts the adjusted  $R^2$  to .37, a substantial increase from the simpler models. An increase of one standard deviation of quality QL1 enhances the real per capita growth rate by about 1 percent per year ( $0.0948 \cdot 10.67 = 1.01$ ) for participants in international comparison studies. This is almost identical to the effect of a one standard deviation increase in the quantity of schooling ( $.363 \cdot 2.63 = 0.95$ ). From comparing either columns 1 and 3 or 2 and 4, the quantity effect is seen to fall by about 30 percent with inclusion of direct performance measures. (The Denison adjustment, based on no data but including more than just achievement, was 40 percent).

The basic model considers additive effects of quantity and quality, but the effect of each may differ according to the level of the other. This variant, found in column 5,

---

<sup>23</sup>Easterly and Rebelo (1993) point out that using World Bank data reduces the possibility of the negative coefficient on initial income arising from potential measurement error in initial income. Nevertheless, for consistency with other studies, we rely on Summers and Heston (1991) data for initial income and growth rate calculation.

indicates a negative interaction. The marginal effect of quality declines with an overall more highly schooled population and vice versa. This interaction may, for example, indicate how the United States has managed to overcome many of the effects of its low performance on cognitive assessments (see figure 1), although postsecondary schooling is also not considered. With sufficiently high scores, the marginal effect of years of schooling even appears negative—a highly unlikely effect.<sup>24</sup> This situation appears to us to be a reflection of the rigid relationship forced by the linear interaction model, but with the small number of observations experimentation with alternative functional forms seems inappropriate.<sup>25</sup>

The remaining columns in Table 4, (6) and (7), employ the alternative labor force quality measures. Because of high correlations (larger than 0.9) among quality measures, the results using other measures, QL2 and QL3, are virtually indistinguishable. We drop further consideration of QL3 and return below to the parallel consideration of QL2 in the investigation of the determinants of quality and to their policy implications.

Table 5 includes a series of other explanatory factors, consistent with various previous estimation. The added factors include the ratio of real government consumption expenditure net of spending on defense and education to real GDP (GCN), the ratio of private investment to GDP (PRIINV), and ratio of total trade to GDP (TRD). As has been pointed out previously, each is open to some interpretative question about direction of causation. Our main interest, however, is the sensitivity of the human capital estimates to inclusion of this

---

<sup>24</sup>The marginal impact of higher quantity (S) becomes negative for values of QL1 above 44.2. This characterizes the situation in roughly half of the countries in the assessment sample. For the marginal effect of test performance (QL1), no plausible or observed values are negative.

<sup>25</sup>The substitution of logarithmic models yielded virtually identical qualitative results.

Table 5  
 Estimates of Alternative 1960-1990 Cross-country Growth Models with Labor Force Quality  
 Dependent variable: Average annual growth rate in real per capita GDP (x100)

	(3)	(7)	(8)	(9)	(10)
constant	1.293 (5.29)	2.439 (6.14)	0.318 (0.76)	1.043 (1.83)	0.866 (2.51)
Initial per capita income (Y60) [\$1,000]	-0.49 (-4.38)	-0.58 (-6.48)	-0.52 (-5.99)	-0.529 (-5.82)	-0.40 (-3.32)
Quantity of schooling (S)	0.363 (3.16)	0.404 (4.25)	0.300 (3.18)	0.266 (2.79)	0.311 (2.65)
Assessment available (TEST)	-2.825 (-2.99)	-1.662 (-1.95)	-1.907 (-1.82)	-1.328 (-1.35)	-1.329 (-1.75)
Labor force quality (TEST·QL1)	0.095 (4.40)	0.065 (3.34)	0.067 (2.94)	.050 (2.27)	0.060 (3.15)
Government consumption net of defense and education/GDP (GCN)		-9.113 (-3.47)		-7.340 (-2.91)	
Private investment/GDP (PRIINV)			10.807 (2.80)	12.646 (3.30)	
Total trade/GDP (TRD)					0.681 (1.37)
Number of observations	100	87	90	84	94
R-squared	0.365	0.460	0.461	0.540	0.304

Note: Estimates are corrected for heteroscedasticity according to White (1980). t-statistics are in parentheses below coefficients.

wider set of factors (cf. Levine and Renelt (1992), Levine and Zervos (1993)). For convenience, column 3 is reproduced from the baseline model of Table 4. While the quality coefficient is reduced in magnitude, the coefficients for initial income, quantity and quality remain remarkably stable even with the reduced sample sizes. GCN and PRIINV are significant and of correct signs. On the other hand, TRD is not significantly different from zero. We take these alternative specifications as further support for the fundamental importance of variations in labor force quality.<sup>26</sup>

The results of this basic estimation provide strong support for the importance of quality difference as measured by cognitive achievement in mathematics and science. They do, nonetheless, suffer from the data limitations of such analysis. Before going into the implications of these results, we first attempt to improve further on the data about labor force quality differences.

### *B. Determinants of schooling quality*

The estimates of how labor force quality affect growth rates are based just on the countries that have participated in international assessments of mathematics and science performance. Two immediate questions arise from these. First, can we use information on the schooling systems in these countries to resolve the policy ambiguity resulting from the lack of information about how cognitive skills can be increased? Second, since less than half of the countries had sufficient data to be included in growth analyses, can we project labor

---

<sup>26</sup>The same sensitivity analysis was performed on the models with quantity and quality interaction (column 5 in Table 3), and the estimation again appears robust.

force quality to a wider set of countries in order to understand their growth? This section considers the production of achievement; the next describes methods of projecting labor force quality.

The strategy is to estimate Equation 2, the human capital production function, in terms of measured test scores and aggregate characteristics of the schools and populations across countries. The approach follows the microlevel studies of school performance (Hanushek (1979, 1986)) except we assume that the international level of average ability of students does not vary across countries (or at least is exogenous to the other determinants considered here). This assumption then leads us to concentrate on standard, readily available resource measures and aggregate population characteristics.<sup>27</sup> To estimate the school production relationships, we consider normalized individual country test scores on the six separate tests (i.e., the individual data points plotted in figure 1) and relate them to the family characteristics and school resources of prior years. (The precise matching of years is listed in Appendix B). This approach matches test performance with the relevant inputs for the students taking the tests and ensures that the inputs are exogenous to the achievement outcomes. Complete data are available for 67 country-test observations.

---

<sup>27</sup>The sources of variables, as explained in the Data Appendix, are mainly from Barro and Lee (1994)'s data set and Lockheed and Verspoor (1991). For characteristics of education system, we consider: pupil/teacher ratio in primary schools (PT-pri) and in secondary schools (PT-sec), pupils/school ratios, teaching materials and teacher salary in primary schools, repeaters in primary schools, percentage of primary school cohort reaching the last grade, public recurrent expenditure per student in primary school relative to GNP, ratio of recurring nominal government expenditure on education to nominal GDP (RECUR), and ratio of total nominal government expenditure on education to nominal GDP (EXPEND). For socio-economic variables, we consider: general education level (schooling quantity - S), per capita income at 1960 (Y60) and average income, demographic variables (fertility rate, growth rate of population - GPOP, infant mortality rate, life expectancy at birth).



Table 6 displays several variants of the production models. The overall story that comes through is that variations in school resources do not have strong effects on test performance. The estimated effects of resources are either statistically insignificant or, more frequently, statistically significant but with the wrong sign. This finding holds regardless of the specific measure of school resources, whether pupil-teacher ratios, recurrent expenditure, total expenditure, or a variety of other measures.<sup>28</sup> The education of parents, proxied by the quantity of schooling of the population, tends to be positive and significant at conventional levels. Also, countries with higher population growth rates tend to have lower achievement, consistent with standard arguments about the trade-off between quantity and quality of children and the impact of larger family sizes (Becker and Lewis 1973; Willis 1973; Hanushek 1992). The lack of strong relationship between resources and school performance internationally mirrors findings within individual developed and developing countries (Hanushek 1989, 1995). In many ways, this result is even less surprising in the international context. The organization and incentive structures of schools, important elements of equation 3, vary widely across nations, would be expected to influence performance strongly, and are unmeasured in this work. (On the other hand, resources vary more widely across countries than within the countries, suggesting that any resource effects

---

<sup>28</sup>The pupil-teacher in primary schools is statistically significant but positive. The estimated lack of importance of school resources is somewhat surprising given the impact of aggregation on the estimated effects of school resources (Hanushek, Rivkin, and Taylor 1995). The situation here appears to be similar to the conditions for upward bias in school effects, since aggregation of the school analysis to the national level is accompanied missing information about the organization and policies across countries.

The previous footnote describes a variety of other factors that were investigated in the course of this analysis. Most, however, are available for an even more restricted subset of countries than those found here.

Table 6  
 Production of Mathematics and Science Achievement:  
 Country Scores Separated by Specific Test  
 (67 observations)

	Normalized test performance, year t					
	(11)	(12)	(13)	(14)	(15)	(16)
constant	40.22 (8.69)	52.27 (10.59)	40.00 (6.17)			
IEA—math1				50.15 (8.94)	37.90 (4.82)	50.33 (9.02)
IEA—science1				54.30 (9.91)	41.45 (6.24)	54.84 (10.28)
IEA—math2				55.49 (9.82)	41.20 (5.96)	56.42 (10.29)
IEA—science2				51.66 (8.39)	37.50 (4.94)	52.85 (8.81)
IAEP—math and science				52.45 (10.69)	38.71 (5.86)	53.69 (11.28)
Adult schooling ( $S_{t,i}$ )	2.04 (2.50)	1.62 (2.14)	1.08 (1.80)	1.75 (2.39)	1.16 (1.90)	2.00 (2.81)
Pupil-teacher ratio in primary schools ( $PT-pri_{t,i}$ )			0.343 (2.59)		0.35 (2.58)	
Recurring expenditure on education/GDP ( $RECUR_{t,i}$ )						-286.29 (-2.99)
Total expenditure on education/GDP ( $EXPEND_{t,i}$ )	-96.24 (-1.01)	-165.9 (-1.83)		-189.8 (-2.14)		
Annual population growth ( $GPOP_{t,i}$ )		-4.60 (-3.38)	-5.38 (-3.70)	-4.98 (-3.51)	-5.60 (-3.86)	-5.28 (-3.71)
R-squared (adjusted)	0.11	0.19	0.22	0.17	0.20	0.21

The sample consists of one observation per country for each of the six test years that each country participated in the international testing. Scores are normalized to 50 in each test year. Estimates are corrected for heteroscedasticity according to White (1980). t-statistics are in parentheses below coefficients.

may be easier to observe). The impotence of pure resource policies, while complicating governmental actions, primarily underscores the need for alternative incentive programs (see Hanushek with others, 1994).

The lack of systematic income and expenditure effects on labor force quality actually strengthens this estimation of growth models. A significant concern with previous analyses of cross-country growth relationships has been the likelihood of simultaneity—fast growing countries are likely to invest in more schooling, more plant and equipment, and the like. Thus, it is unclear whether the single equation estimates of the effects of these factors on growth truly reflect causal influences. On the other hand, if common policy directions to improve quality through added resources are not systematically related to actual performance, the problems of simultaneity will be less important.

### *C. Projecting Labor Force Quality*

In order to project labor force quality, we return to the sample of country values for labor force quality, QL1 and QL2. These measures, aggregated across all available test information, are related to average population and schooling characteristics for the entire 1960-90 period. Our primary purpose is developing projection equations, and we employ expanded models that go beyond the previous production function estimates. The estimates for the two measures of quality are shown in Table 7.

Three basic factors are generally important in determining variations in school quality. Primary school enrollment rates strongly influence performance, probably through indicating the overall importance that each country puts on education. Higher population

Table 7  
Prediction Models for Labor Force Quality

	QL1			QL2		
	(17)	(18)	(19)	(20)	(21)	(22)
constant	-35.59 (-2.40)	-37.19 (-1.94)	-36.19 (-1.72)	-13.10 (-0.59)	-28.40 (-1.20)	-29.27 (-1.13)
Primary school enrollment rate (ENROLL-pri)	81.50 (5.72)	85.78 (4.30)	86.23 (4.06)	62.36 (2.96)	73.28 (2.90)	75.32 (2.81)
Quantity of schooling (S)	0.714 (1.56)	0.038 (0.07)	0.320 (0.58)	1.651 (2.47)	0.97 (0.96)	1.181 (1.21)
Pupil-teacher ratio in primary schools (PT-pri)	0.142 (2.60)			-0.043 (-0.37)		
Recurring expenditure on education/GDP (RECUR)			55.49 (0.54)			133.23 (0.88)
Total expenditure on education/GDP (EXPEND)		121.07 (1.29)			170.37 (1.18)	
Annual population growth (GPOP)	-3.11 (-2.05)	-3.08 (-1.82)	-2.80 (-1.59)	-4.105 (-2.02)	-4.17 (-1.84)	-3.85 (-1.72)
Asia (=1)	5.123 (1.67)	7.52 (3.18)	6.33 (2.41)	12.25 (2.18)	13.77 (3.38)	12.76 (2.97)
Latin American (=1)	-4.004 (-2.01)	-3.87 (-1.54)	-4.31 (-1.49)	-1.24 (-0.39)	0.203 (0.06)	0.432 (0.11)
Africa (=1)	3.170 (1.55)	2.94 (1.48)	3.30 (1.44)	11.97 (2.43)	8.71 (2.96)	9.16 (2.88)
Number of countries	31	30	30	31	30	30
R-squared (adjusted)	0.68	0.63	0.62	0.59	0.56	0.54

Note: Estimates are corrected for heteroscedasticity according to White (1980). t-statistics are in parentheses below coefficients.

growth rates are associated with lower labor force quality. Finally, distinct regional differences influence performance with Asian countries doing best (given the other characteristics).<sup>29</sup>

The average quantity of schooling is generally positively related to performance, although it is typically insignificant at conventional levels. School resources again are not strongly related to quality. The incorrect positive sign for pupil-teacher ratio appears whether or not there is a dummy variable for the Asian region, a region with traditionally high pupil-teacher ratios and high student performance. The expenditure measures, while positive, are statistically insignificant.

The primary purpose of this modeling is developing a way to project quality to the labor forces of other countries for which direct test information is not available. For these purpose, we utilize equations 18 and 21 (for the two basic measures of labor force quality). From these we construct expanded quality measures,  $QL1^*$  and  $QL2^*$ , which combined observed quality with predicted quality for all countries without observed test data but with data on the right hand side measures. A primary concern, however, involves the range of observation. While there are some developing countries represented in the testing, there are not many. Moreover, relatively few countries appear in the low-achieving end of the distribution. When we project scores, we obtain a number of estimates not only outside the range of test observation (that is, less than 20) but also even less than zero. Because of the limitations on our understanding of the structural determinants of school performance and

---

<sup>29</sup>The limited number of countries makes the regional differences suspect. While there are six Asian countries, there are only two African (Mozambique and Swaziland) and one Latin American (Brazil) country.

because of the thinness of observations at the lower end (see Figure 1), we eliminate consideration of any country with predicted achievement below 20, the lower bound on actual observations. This cut-off lies more than  $2\frac{1}{2}$  standard deviations below the country means. This criteria eliminates nine countries which meet overall data availability criteria.<sup>30</sup>

#### *D. Expanded Estimates of Quality Effects*

Table 8 displays the estimated growth models relying on the expanded quality measures (*QL1\** and *QL2\**). These estimates replicate the basic models in Table 4 but now cover a much wider set of countries. In practice, the two measures of labor force quality are very highly correlated (above .95) so there is little qualitative or quantitative difference between the alternative versions of quality.

The expanded estimates are extremely consistent with the baseline estimates in Table 4. Column 23 indicates significant conditional convergence with higher initial income translating into lower growth rates. While quantity of schooling is positive, it is statistically insignificant, and its estimated effect on growth is one-third what was observed previously. Most important, labor force quality—measured by *QL1\**—is statistically significant and almost precisely the same magnitude as the estimates for the restricted set of countries in the baseline. Again, a one standard deviation change in quality translates into a slightly greater than one percent difference in annual real growth rates. The effect of quality improvements now appears much more important than increases in quantity, where the point estimate for a

---

<sup>30</sup>Note that the criteria has slightly different sample implications for the two separate quality measures. There are 77 common observations. For *QL1\**, Iran is added; for *QL2\**, the Central African Republic, Papua New Guinea, and Ghana are added. These sample differences have minor impacts on the final estimation.

Table 8  
 1960-1990 Cross-country Growth Models with Labor Force Quality  
 Dependent variable: Average annual growth rate in real per capita GDP (x100)  
 (Expanded quality samples)

	(23)	(24)	(25)	(26)	(27)	(28)
constant	-1.606 (-2.20)	-1.18 (-0.99)	-0.475 (-0.46)	-1.483 (-2.61)	-0.869 (-0.91)	-0.657 (-0.86)
Initial per capita income (Y60) [\$1,000]	-0.382 (-4.87)	-0.390 (-5.11)	-0.453 (-6.02)	-0.370 (-4.49)	-0.384 (-4.86)	-0.442 (-5.65)
Quantity of schooling (S)	0.127 (1.48)	0.117 (1.31)	0.112 (1.25)	0.120 (1.29)	0.103 (1.07)	0.112 (1.12)
Annual population growth (GPOP)		-0.097 (-0.47)			-0.161 (-0.80)	
Labor force quality ( <i>QL1*</i> )	0.108 (5.34)	0.104 (4.69)	0.076 (2.90)			
Labor force quality ( <i>QL2*</i> )				0.094 (6.17)	0.090 (5.69)	0.072 (3.61)
Assessment available (TEST)			-1.392 (-1.00)			-0.628 (-0.46)
Observed labor force quality (TEST· <i>QL1*</i> )			0.054 (1.74)			
Observed labor force quality (TEST· <i>QL2*</i> )						0.034 (1.26)
Number of Countries	78	78	78	80	80	80
Adjusted R-squared	0.39	0.38	0.46	0.39	0.38	0.44

Note: Estimates are corrected for heteroscedasticity according to White (1980). t-statistics are in parentheses below coefficients.

one standard deviation change in school quantity has now fallen to .32 percentage points in average growth.

The remainder of the table demonstrates stability of the estimated effects of quality. In column 24, the addition of population growth rates has no impact on the estimated quality effect. Column 25 provides information on the projection of labor force quality by adding terms that indicate countries with direct observations as distinct from projections. The point estimates on marginal test score effects ( $TEST \cdot QL1^*$ ) indicate somewhat stronger test influences in countries where there are direct observations, although the coefficient is not significant at the five percent level.<sup>31</sup> Similarly, the results from the alternative quality measure ( $QL2^*$ ) demonstrate stability and again confirm the appropriateness of projection to the expanded set of countries. Importantly, the estimates of the effects of a one standard deviation change in  $QL2^*$  are virtually identical to those for  $QL1^*$ . When other common measures of economies used in Table 5 are added (not shown), the importance of quality differences remains both in terms of consistent strength and statistical significance.<sup>32</sup>

## 5. The Importance of Quality Measurement

The story of cross-country growth regressions has been one of many variations in modeling without a firm sense of the implications or importance of alternative influences on growth (see, for example, Levine and Renelt 1992). One facet of the quality measures

---

<sup>31</sup>The F-test on the addition of both the intercept and slope dummies does reject the hypothesis that both parameters are jointly zero ( $F[2,72] = 5.59$ ), but slope equality seems most important.

<sup>32</sup>The alternative forms of growth models in Table 5 were estimated with  $QL1^*$ . While school quantity ( $S$ ) was insignificant, the consistent result was that  $QL1^*$  remained statistically significant, although at times slightly smaller in magnitude than in equation 23.



developed here is the consistency of estimated effects across alternative specification and across inclusion of a wide variety of other factors which might influence growth. Two other attributes of the modeling are important. First, the direct quality measures based on international test scores are far superior to input measures of schooling differences. Second, the inclusion of quality measures resolves a number of anomalies of countries growing much faster or much slower than expected based on other objective information.

Table 9 provides direct comparisons of estimated models with quality measures versus no quality measures or schooling input measures. Without quality measures (columns 29-31), school quantity is consistently significant. Further, primary school pupil-teacher ratio appears close to statistically significant, but the magnitude is small. Cutting average pupil-teacher ratios in half (to 18 from its current mean of 36), an extraordinarily expensive policy, would increase growth rates by .7 percent per year. Direct measures of labor force quality, however, double the proportion of variance explained, reaching almost 40 percent. Quality measures also replace quantity in demonstrating the importance of human capital for national growth.

The importance of adequate measures of quality becomes clear in Table 10 which provides comparisons of the root mean squared error (RMSE) in explaining growth rates across a consistent set of countries. Regardless of which quality measure is employed, the root mean squared error falls by 0.2 percentage points (in annual growth) when compared with the other models in Table 9. Moreover, the quality measures are especially powerful in explaining anomalies in growth rates. The final three columns in Table 10 divide the countries into slow growers (average annual growth less than 1 percent), fast growers

Table 9  
Comparison of Alternative Measures of School Quality and Growth

	(29)	(30)	(31)	(23)	(26)
constant	0.933 (3.98)	2.140 (2.06)	0.558 (1.37)	-1.606 (-2.20)	-1.483 (-2.61)
Initial per capita income (Y60) [\$1,000]	-0.407 (-2.96)	-0.455 (-4.10)	-0.408 (-3.05)	-0.382 (-4.87)	-0.370 (-4.49)
Quantity of schooling (S)	0.529 (4.40)	0.481 (3.97)	0.486 (3.73)	0.127 (1.48)	0.120 (1.29)
Pupil-teacher ratio in primary schools (PT-pri)		-0.040 (-1.72)			
Pupil-teacher ratio in secondary schools (PT-sec)		0.027 (0.67)			
Total expenditure on education/GDP (EXPEND)			14.42 (1.00)		
Labor force quality ( <i>QL1*</i> )				0.108 (5.34)	
Labor force quality ( <i>QL2*</i> )					0.094 (6.17)
Number of Countries	100	96	96	78	80
Adjusted R-squared	0.21	0.23	0.20	0.39	0.39

Note: Estimates are corrected for heteroscedasticity according to White (1980). t-statistics are in parentheses below coefficients.

Table 10  
Root Mean Squared Error for Alternative Growth Models

Model	All countries	growth < 1%	growth 1-3.5%	growth > 3.5%
<i>A. QL1* quality measure</i>				
(29) C, Y60, S	1.50	2.16	0.87	2.16
(30) C, Y60, S, PT-pri, PT-sec	1.50	2.16	0.87	2.17
(31) C, Y60, S, EXPEND	1.50	2.14	0.86	2.18
(23) C, Y60, S, <i>QL1*</i>	1.28	1.95	0.79	1.71
Number of countries	76	13	47	16
<i>B. QL2* quality measure</i>				
(29) C, Y60, S	1.52	2.08	0.89	2.21
(30) C, Y60, S, PT-pri, PT-sec	1.52	2.07	0.89	2.22
(31) C, Y60, S, EXPEND	1.52	2.07	0.88	2.23
(26) C, Y60, S, <i>QL2*</i>	1.31	1.90	0.87	1.67
Number of countries	78	15	47	16

(average annual growth greater than 3.5 percent), and the remainder. The reduction in root mean squared error comes at the extremes of the growth distribution, and especially from better explaining rapid growth. The RMSE for fast growers falls from about 2.2 percent to 1.7 percent when *QL1\** is used to explain growth differences. While the improvement at the bottom of the distribution is less, it is still noticeable at about 0.2 percent. The results for *QL2\** are similar.

The way in which explicit quality measures changes the picture of growth is best seen in Table 11. This table lists the countries that are at the extremes of the distribution according to different criteria. The first column simply identifies the ten fastest and ten slowest growing countries in our sample over the period 1960-1990. The remaining four columns identify countries that do particularly well or poorly, after considering the observed characteristics of the explanatory factors in Table 10; i.e., which have the largest and the smallest unexplained residuals.<sup>33</sup> The first three models from Table 10, which did not include any quality measure, provide precisely the same listing of the unexpectedly poor-performing countries and provide the same listing for nine out of the top ten unexpectedly good-performing countries. Once quality is taken into account, however, four new countries do unexpectedly well and four do unexpectedly poorly. The conditional top performers (italicized in Table 11) now include Egypt, Cyprus, Syria, and the United States. At the other extreme, the newly identified poor performers include Jamaica, Fiji, Zimbabwe, and Sri Lanka.

---

<sup>33</sup>Table 11 is based just on models with *QL1\**, but the picture with *QL2\** is very similar. Because the samples differ, there is a slight change, but the comparison with estimation not including labor force quality is qualitatively similar.

Table 11  
 Top and Bottom Performing Economies Conditional on Observed Human Capital:  
 Average annual growth rate in real per capita GDP, 1960-1990

Observed growth	Conditional performance based on expectations of equation:			
	(29)	(30)	(31)	(23)
<i>A. Top ten performers</i>				
Korea	Botswana	Botswana	Singapore	Botswana
Hong Kong	Singapore	Singapore	Botswana	Singapore
Singapore	Hong Kong	Hong Kong	Hong Kong	Hong Kong
Botswana	Korea	Taiwan	Korea	Korea
Taiwan	Taiwan	Korea	Taiwan	<i>Egypt</i>
Malta	Portugal	Portugal	Portugal	<i>Cyprus</i>
Japan	Malta	Malta	Malta	<i>Syria</i>
Cyprus	Japan	Japan	Spain	Thailand
Thailand	Tunisia	Malaysia	Thailand	Taiwan
Malaysia	Thailand	Thailand	Japan	<i>United States</i>
<i>B. Bottom ten performers</i>				
Zambia	Guyana	Guyana	Guyana	Guyana
Guyana	Zambia	Zambia	Zambia	Zambia
Zaire	Argentina	Argentina	Argentina	Argentina
Iraq	Peru	Peru	Peru	Uruguay
Venezuela	Zaire	Zaire	Zaire	Zaire
Nicaragua	Nicaragua	Nicaragua	Nicaragua	Peru
Peru	Philippines	Uruguay	Philippines	<i>Jamaica</i>
Argentina	Uruguay	Philippines	Uruguay	<i>Fiji</i>
Uruguay	New Zealand	New Zealand	New Zealand	<i>Zimbabwe</i>
Iran	Bolivia	Bolivia	Bolivia	<i>Sri Lanka</i>

It is interesting to put this in the framework of the often noted growth performance of East Asia (see World Bank 1993). The final column, which includes measures of labor force quality, still identifies Singapore, Hong Kong, Korea, Thailand, and Taiwan as unusually fast growers, but this identification already takes into account the growth that is expected based on the quantity and quality of schooling. In other words, the East Asian miracle includes a significant component that is over and above any emphasis and performance in the development of human capital. (As shown in Figure 1, these countries additionally tend to do well in developing human capital quality). Thus, without minimizing the importance of human capital development, it is important to note that a significant component of growth falls outside of human capital development.

## **6. Conclusions**

International comparative studies seldom yield conclusive findings. The limited number of countries combined with severe data shortcomings imply that generally the best that can be hoped for is suggestive conclusions which focus attention on general areas of consideration. Comparative growth analyses—relying on varying samples, differing analytical foci, and imperfect data—have led to some general findings along with many suggestive answers of questionable reliability.

Throughout consideration of varying national growth rates, perhaps the most robust empirical finding and the most readily accepted conclusion is that a nation's human capital is central to the economic performance it can expect. This conclusion has, nonetheless, come from models with very different specifications of human capital, and virtually all have relied on an unbelievable although implicit assumption: that any variations in quality of human

capital are small relative to the importance and variation in pure quantity of human capital as proxied by years of schooling.

The analysis here explicitly considers the quality of the labor forces as measured by comparative tests of mathematics and scientific skills for school children. With some 39 countries participating in various international testing programs over the past three decades, it is possible to construct comparative performance measures. These performance measures are then projected to countries not participating in the testing programs. A single conclusion emerges from the various analytical specifications of how quality of the labor force affects growth: Quality has a consistent, stable, and strong influence on economic growth. Across a wide variety of specifications which include many of the measures employed in past work, the impact of quality indicates that one standard deviation in mathematics and science skills translates into one percentage point in average annual real growth. This growth effect is larger than would be obtained from over eight years in average schooling.

While other analyses have highlighted the importance of school quality within countries (Hanushek 1995), this is the first to identify its importance in explaining differential growth. Here too we arrive at the dilemma of determining how school performance can be improved. Just as was the case within countries, little evidence suggests that simple resource policies are likely to improve student (and national) performance. While beyond the scope of this paper, other work (Hanushek with others 1994) points to the importance of school organization and incentives. The current general absence of performance incentives almost certainly contributes to the lack of relationship between resources and student performance internationally.

## References

- Abramovitz, Moses. "Economic growth in the United States: A review article." *American Economic Review* 52, no. 4 (September 1962): 762-82.
- Barro, Robert J. "Economic growth in a cross section of countries." *Quarterly Journal of Economics* 106, no. 2 (May 1991): 407-43.
- Barro, Robert J., and Jong-wha Lee. "International comparisons of educational attainment." *Journal of Monetary Economics* 32, no. 3 (December 1993): 363-94.
- . *Data set for a panel of 138 countries*. (1994).
- Barro, Robert J., and Xavier Sala-I-Martin. "Convergence." *Journal of Political Economy* 100, no. 2 (April 1992): 223-51.
- Barro, Robert J., and Xavier Sala-I-Martin. *Economic Growth*. New York: McGraw-Hill, Inc., 1995.
- Becker, Gary S. *Human capital: A theoretical and empirical analysis, with special reference to education*. 3rd ed. Chicago: University of Chicago Press, 1993.
- Becker, Gary S., and H. Gregg Lewis. "On the interaction between the quantity and quality of children." *Journal of Political Economy* 81, no. pt. II (March/April 1973): S279-88.
- Behrman, Jere R., and Nancy Birdsall. "The quality of schooling: quantity alone is misleading." *American Economic Review* 73, no. 5 (December 1983): 928-46.
- Ben-Porath, Yoram. "The Production of Human Capital Over Time." In *Education, Income and Human Capital*, edited by W. Lee Hansen, 129-47. New York: National Bureau of Economic Research, 1970.
- Betts, Julian R. "Is there a link between school inputs and earnings? Fresh scrutiny of an old literature". University of California, San Diego (mimeo), November 1994.
- . "Does school quality matter? Evidence from the National Longitudinal Survey of Youth." *Review of Economics and Statistics* (forthcoming).
- Bishop, John. "Is the test score decline responsible for the productivity growth decline?" *American Economic Review* 79, no. 1 (1989): 178-97.
- . "Achievement, test scores, and relative wages." In *Workers and their wages*, edited by Marvin H. Kosters, 146-86. Washington, DC: The AEI Press, 1991.



---. "The impact of academic competencies of wages, unemployment, and job performance." *Carnegie-Rochester Conference Series on Public Policy* 37 (December 1992): 127-94.

Card, David, and Alan B. Krueger. "Does school quality matter? Returns to education and the characteristics of public schools in the United States." *Journal of Political Economy* 100, no. 1 (February 1992): 1-40.

Comber, L. C., and John P. Keeves. *Science education in nineteen countries: An empirical study*. Stockholm: Almquist and Wiksell, 1973.

Denison, Edward F. *The sources of economic growth in the United States and the alternatives before us*. Vol. Supplementary Paper No. 13. New York: Committee for Economic Development, 1962.

---. "Measuring the contribution of education (and the residual) to economic growth." In *The residual factor and economic growth*, edited by OECD, 13-55. Paris: Organization for Economic Co-operation and Development, 1964.

---. *Why growth rates differ: Postwar experience in nine western countries*. Washington, DC: Brookings, 1967.

---. "Final comments." *Survey of Current Business* 52, no. 5, Pt. II (May 1972): 95-110.

---. "The growth accounting tradition and proximate sources of growth." In *Explaining economic growth*, edited by Adam Szirmai, Bart van Ark, and Dirk Pilat, 37-64. New York: Elsevier Science Publishers, 1993.

Easterly, William, and Sergio Rebelo. "Fiscal policy and economic growth: An empirical investigation." *Journal of Monetary Economics* 32, no. 3 (December 1993): 417-58.

Engerman, Stanley L. "Human capital, education, and economic growth." In *The reinterpretation of American economic history*, edited by Robert W. Fogel and Stanley L. Engerman, 241-56. New York: Harper & Row, 1971.

Ferguson, Ronald F. "New evidence on the growing value of skill and consequences for racial disparity and returns to schooling". Malcolm Wiener Center for Social Policy, H-93-10, Harvard University, September 1993.

Griliches, Zvi. "Errors in variables and other unobservables." *Econometrica* 42, no. 6 (November 1974): 971-98.

Grogger, Jeffrey T. "Does school quality explain the recent black/white wage trend?" *Journal of Labor Economics* (forthcoming).

Grogger, Jeffrey T., and Eric Eide. "Changes in college skills and the rise in the college wage premium." *Journal of Human Resources* 30, no. 2 (Spring 1993): 280-310.

Hanushek, Eric A. "Conceptual and empirical issues in the estimation of educational production functions." *Journal of Human Resources* 14, no. 3 (Summer 1979): 351-88.

---. "The economics of schooling: Production and efficiency in public schools." *Journal of Economic Literature* 24, no. 3 (September 1986): 1141-77.

---. "The impact of differential expenditures on school performance." *Educational Researcher* 18, no. 4 (May 1989): 45-51.

---. "The trade-off between child quantity and quality." *Journal of Political Economy* 100, no. 1 (February 1992): 84-117.

---. "School Resources and Student Performance". paper prepared for Brookings Institution conference, "Do School Resources Matter?", June 1994.

---. "Interpreting recent research on schooling in developing countries." *World Bank Research Observer* 10, no. 2 (August 1995): 227-46.

Hanushek, Eric A., Steven G. Rivkin, and Lori L. Taylor. "Aggregation and the Estimated Effects of School Resources". Working Paper No. 397, Rochester Center for Economic Research, University of Rochester, February 1995.

Harbison, Ralph W., and Eric A. Hanushek. *Educational performance of the poor: lessons from rural northeast Brazil*. New York: Oxford University Press, 1992.

Hartigan, John A., and Alexandra K. Wigdor, eds. *Fairness in employment testing: Validity generalization, minority issues, and the general aptitude test battery*. Washington, DC: National Academy Press, 1989.

Heckman, James J., Anne S. Layne-Farrar, and Petra E. Todd. "Does Measured School Quality Really Matter?". University of Chicago (mimeo), September 1994.

Heyneman, Stephen P., and William Loxley. "The effect of primary school quality on academic achievement across twenty-nine high and low income countries." *American Journal of Sociology* 88 (May 1983): 1162-94.

Husén, Torsten, ed. *International study of achievement in mathematics*. Vol. II. New York: John Wiley and Sons, 1967.

Johnson, George E., and Frank P. Stafford. "Social returns to quantity and quality of schooling." *Journal of Human Resources* 8, no. 2 (Spring 1973): 139-55.

---. "On the Intensive Rate of Return to Schooling". Department of Economics, University of Michigan (mimeo), April 1995.

Jorgenson, Dale W., and Barbara M. Fraumeni. "Investment in education and U.S. economic growth." *Scandinavian Journal of Economics* 94, no. Supplement (1992): 51-70.

Jorgenson, Dale W., and Zvi Griliches. "The explanation of productivity change." *Review of Economic Studies* 34, no. 3 (July 1967): 249-82.

---. "Issues in growth accounting: A reply to Edward F. Denison." *Survey of Current Business* 52, no. 5, Pt. II (May 1972): 65-94.

King, Robert G., and Ross Levine. "Finance, entrepreneurship, and growth: Theory and evidence." *Journal of Monetary Economics* 32, no. 3 (December 1993): 513-42.

Kormendi, Roger C., and Philip G. Meguire. "Macroeconomic determinants of growth: Cross-country evidence." *Journal of Monetary Economics*, 16 (1985): 141-63.

Lapointe, Archie E., Nancy A. Mead, and W. Phillips Gary. *A world of difference: An international assessment of mathematics and science*. Princeton, NJ: Educational Testing Service, 1989.

Levine, Ross, and David Renelt. "A sensitivity analysis of cross-country growth regressions." *American Economic Review* 82, no. 4 (September 1992): 942-63.

Levine, Ross, and Sara J. Zervos. "What we have learned about policy and growth from cross-country regressions?" *American Economic Review* 83, no. 2 (May 1993): 426-30.

Lockheed, Marlaine E., and Adriaan Verspoor. *Improving primary education in developing countries*. New York: Oxford University Press, 1991.

Mankiw, N. Gregory, David Romer, and David Weil. "A contribution to the empirics of economic growth." *Quarterly Journal of Economics* (May 1992): 407-37.

Medrich, Elliott A., and Jeanne E. Griffith. *International mathematics and science assessment: what have we learned?* Washington, D.C.: Office of Educational Research and Improvement, U.S. Department of Education, 1992.

Mincer, Jacob. "The distribution of labor incomes: a survey with special reference to the human capital approach." *Journal of Economic Literature* 8, no. 1 (March 1970): 1-26.

---. *Schooling Experience and Earnings*. New York: NBER, 1974.

Murnane, Richard J., John B. Willett, and Frank Levy. "The growing importance of cognitive skills in wage determination". Harvard Graduate School of Education (mimeo), March 1994.

O'Neill, June. "The role of human capital in earnings differences between black and white men." *Journal of Economic Perspectives* 4, no. 4 (Fall 1990): 25-46.

Postlethwaite, T. Neville, and David E. Wiley. *The IEA study of science II: Science achievement in twenty-three countries*. Pergamon Press: Oxford, 1992.

Robitaille, David, and Robert Garden, eds. *The International Association for the Evaluation of Education Achievement (IEA) study of mathematics II: Contexts and outcomes of school mathematics*. Vol. II. Pergamon: Oxford, 1989.

Romer, Paul. "Endogenous technological change." *Journal of Political Economy* 99, no. 5, pt. II (1990): S71-S102. (a)

---. "Human capital and growth: Theory and evidence." *Carnegie-Rochester Conference Series on Public Policy* 32 (1990): 251-86. (b)

Schultz, Theodore W. "Investment in human capital." *American Economic Review* 51, no. 1 (March 1961): 1-17.

Solow, Robert M. "Technical change and the aggregate production function." *Review of Economics and Statistics* 39, no. 3 (August 1957): 312-20.

Speakman, Robert, and Finis Welch. "Does school quality matter? --A Reassessment". Texas A&M University (mimeo), January 1995.

Summers, Robert, and Alan Heston. "The Penn World Table (Mark 5): An expanded set of international comparisons, 1950-1988." *Quarterly Journal of Economics* 106, no. 2 (May 1991): 327-68.

U.S. Department of Education. *The Condition Of Education 1993*. Washington, D.C.: Education, National Center for Education Statistics, 1993.

U.S. Department of Education, National Center for Education Statistics. *The condition of education, 1994*. Washington, DC: U.S. Government Printing Office, 1994.

Wachtel, Paul. "The effect on earnings of school and college investment expenditures." *Review of Economics and Statistics* 58, no. 3 (August 1976): 326-31.

Wharton, Clifton R. Jr. "Education and agricultural growth: The role of education in early-stage agriculture." In *Education and economic development*, edited by C. Arnold Anderson and Mary Jean Bowman, 202-28. Chicago: Aldine, 1965.

White, Halbert. "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." *Econometrica* 48, no. 4 (May 1980): 817-38.

Willis, Robert J. "A new approach to the economic theory of fertility behavior." *Journal of Political Economy* 81 (April/March, Pt. II 1973): S14-64.

World Bank. *The East Asian miracle*. New York: Oxford University Press, 1993.

## Appendix A: List of Variables and Sources

### *SOURCES OF DATA*

- (BL) : Barro and Lee (1994)
- (ER) : Easterly & Rebelo (1993)
- (KL) : King and Levine (1993)
- (SH) : Summers and Heston (1991)

### *VARIABLES in ANALYSIS*

Asia, Latin America, and Africa : regional dummies for Asian, Latin American, and Sub-Saharan African countries (BL: ASIAE, LAAM, and SAFRICA).

GCN : ratio of real government consumption expenditure net of spending on defense and on education to real GDP (ER: HSGVXDxE).

ENROLL-pri : average of total gross enrollment rate for primary education (BL: Pxx)

EXPEND: average ratio of nominal government expenditure on education to nominal GDP (BL: GEETOT).

GPOP : average of population growth rate (BL: GPOP).

GR : annual average growth rate of per capita real GDP for the period of 1960-1990. If data are not available for both end points, sub-periods are used for which data are available. Real GDP per capita is SH: RGDPCH in 1985 international prices.

PRIINV : average ratio of private investment relative to GDP. (BL: INVWB - INV PUB).

PT-pri: average pupil/teacher ratio in primary schools (BL: TEAPRI).

QL1, QL2, QL3: measures of schooling quality constructed as described in text; see below.

QL1\*, QL2\* : quality measures expanded to non-participants and participants in international comparison studies for student achievement with predicted values for non-participants; see text and below.

RECUR: ratio of recurring nominal government expenditure on education to nominal GDP. (BL:GEEREC)

S : arithmetic average quantities for 1960, 65, 70, 75, 80, and 85 (BL:HUMAN).

TEST: dummy for participants in international comparison studies of student achievement; see text.

TRD : ratio of total trade to GDP (KL)

Y60 : initial income, 1960 [\$1,000] (SH: RGDPCH)

## Appendix B: Data for Production Function Estimates

The production function estimates (Table 6) employ a sample combining different test years in which the dependent variable is the average normalized score for each country and test. This table describes the dating of exogenous variables for the production function models, based on the specific test year for the performance measure.

	IEA math 1964-66	IEA science 1966-73	IEA math 1980-82	IEA science 1983-86	IAEP 1988	IAEP 1991
$S_{t-1}$	1960	1960-65	1965-75	1970-80	1975-85	1975-85
$EXPEND_{t-1}$	1960-64	1960-69	1965-74	1970-79	1975-84	1975-84
$PT-pri_{t-1}$	1950-60	1955-65	1960-70	1965-75	1970-80	1970-80
$RECUR_{t-1}$	1960-64	1960-69	1965-74	1970-79	1975-84	1975-84
$GPOP_{t-1}$	1960-64	1960-69	1965-74	1970-79	1975-84	1975-84

## Appendix C: School Quality Data

The following data are generated from the methods described in the text. TEST=1 if performance is observed, at which time QL1\*, QL2\*, and QL3 are the observed test data. If TEST=0, QL1\* and QL2\* are projected scores. If blank, the data necessary for projecting scores or for estimating growth models were missing or the projected scores were less than 20 (see text).

COUNTRY	TEST	QL1*	QL2*	QL3
1 ALGERIA	0	28.28	28.06	
2 ANGOLA	0			
3 ARGENTINA	0	42.99	48.50	
4 AUSTRALIA	1	48.13	59.04	64.9
5 AUSTRIA	0	53.20	56.61	
6 BAHAMAS	0			
7 BAHRAIN	0	26.03	23.19	
8 BANGLADESH	0			
9 BARBADOS	0	50.41	59.80	
10 BELGIUM	1	53.25	57.08	63.49
11 BELIZE	0			
12 BENIN	0			
13 BHUTAN	0			
14 BOLIVIA	0	22.10	27.47	
15 BOTSWANA	0	25.05	31.71	
16 BRAZIL	1	33.91	36.60	36.6
17 BULGARIA	0			
18 BURKINA FASO	0			
19 BURUNDI	0			
20 CAMEROON	0	39.00	42.36	
21 CANADA	1	47.57	54.58	54.58
22 CAPE VERDE IS.	0			
23 CENTRAL AFR.R.	0		24.77	
24 CHAD	0			
25 CHILE	1	26.30	24.74	24.74
26 CHINA	1	59.28	64.42	64.42
27 COLOMBIA	0	34.78	37.87	
28 COMOROS	0			
29 CONGO	0	46.04	50.90	
30 COSTA RICA	0	42.15	46.15	
31 CYPRUS	0	42.24	46.24	
32 CZECHOSLOVAKIA	0			
33 DENMARK	0	53.48	61.76	
34 DJIBOUTI	0			
35 DOMINICA	0			
36 DOMINICAN REP.	0	37.41	39.34	
37 ECUADOR	0	35.78	38.99	
38 EGYPT	0	26.01	26.43	
39 EL SALVADOR	0	21.81	26.21	



40	ETHIOPIA	0			
41	FIJI	0	50.02	58.10	
42	FINLAND	1	48.76	59.55	62.81
43	FRANCE	1	54.15	56.00	61.65
44	GABON	0			
45	GAMBIA	0			
46	GERMANY, WEST	1	59.03	48.68	60.84
47	GHANA	0		25.58	
48	GREECE	0	49.11	50.88	
49	GRENADA	0			
50	GUATEMALA	0			
51	GUINEA-BISS	0			
52	GUINEA	0			
53	GUYANA	0	45.71	51.49	
54	HAITI	0			
55	HONDURAS	0	26.43	28.59	
56	HONG KONG	1	56.93	71.85	71.85
57	HUNGARY	1	53.85	61.23	61.23
58	ICELAND	0	48.13	51.20	
59	INDIA	1	21.63	20.80	20.8
60	INDONESIA	0	37.98	42.99	
61	IRAN	1	20.79	18.26	18.26
62	IRAQ	0	29.34	27.50	
63	IRELAND	1	47.59	50.20	50.2
64	ISRAEL	1	51.29	54.46	55.61
65	ITALY	1	44.59	49.41	49.41
66	IVORY COAST	0			
67	JAMAICA	0	44.19	48.62	
68	JAPAN	1	60.65	65.50	69.31
69	JORDAN	1	39.38	42.28	42.28
70	KENYA	0	24.43	29.73	
71	KOREA, REP.	1	56.21	58.55	58.55
72	KUWAIT	0	28.36	22.50	
73	LAOS	0			
74	LESOTHO	0	46.14	51.95	
75	LIBERIA	0			
76	LUXEMBOURG	1	39.45	44.49	44.49
77	MADAGASCAR	0			
78	MALAWI	0			
79	MALAYSIA	0	47.89	54.29	
80	MALI	0			
81	MALTA	0	53.16	57.14	
82	MAURITANIA	0			
83	MAURITIUS	0	49.53	54.95	
84	MEXICO	0	35.06	37.24	
85	MONGOLIA	0			
86	MOROCCO	0			
87	MOZAMBIQUE	1	24.26	27.94	27.94
88	MYANMAR	0			
89	NAMIBIA	0			
90	NEPAL	0			
91	NETHERLANDS	1	56.84	54.52	61.95
92	NEW ZEALAND	1	52.44	67.06	67.06
93	NICARAGUA	0	24.19	27.30	

94	NIGERIA	1	34.15	38.90	38.9
95	NIGER	0			
96	NORWAY	1	49.60	64.56	64.56
97	OMAN	0			
98	PAKISTAN	0			
99	PANAMA	0	42.02	46.78	
100	PAPUA N.GUINEA	0		22.58	
101	PARAGUAY	0	37.99	39.96	
102	PERU	0	37.83	41.18	
103	PHILIPPINES	1	34.35	33.54	33.54
104	POLAND	1	50.28	64.37	64.37
105	PORTUGAL	1	44.09	44.22	44.22
106	PUERTO RICO	0			
107	QATAR	0			
108	REUNION	0			
109	ROMANIA	0			
110	RWANDA	0			
111	SAUDI ARABIA	0			
112	SENEGAL	0			
113	SEYCHELLES	0			
114	SIERRA LEONE	0			
115	SINGAPORE	1	56.51	72.13	72.13
116	SOLOMON IS.	0			
117	SOMALIA	0			
118	SOUTH AFRICA	0	45.25	51.30	
119	SPAIN	1	49.40	51.92	51.92
120	SRI LANKA	0	41.54	42.57	
121	ST.LUCIA	0			
122	ST.VINCENT	0			
123	SUDAN	0			
124	SURINAME	0			
125	SWAZILAND	1	35.46	40.26	40.26
126	SWEDEN	1	47.41	57.43	62.54
127	SWITZERLAND	1	57.17	61.37	61.37
128	SYRIA	0	31.66	30.23	
129	TAIWAN	1	56.28	56.31	56.31
130	TANZANIA	0			
131	THAILAND	1	39.83	46.26	46.26
132	TOGO	0	28.08	32.69	
133	TONGA	0			
134	TRINIDAD&TOBAG	0	40.57	46.43	
135	TUNISIA	0	41.79	40.50	
136	TURKEY	0	41.52	39.72	
137	UGANDA	0			
138	URUGUAY	0	46.33	52.27	
139	U.ARAB E.	0			
140	U.K.	1	53.98	62.52	65.61
141	U.S.A.	1	43.43	46.77	52.29
142	U.S.S.R.	1	53.89	54.65	54.65
143	VAUATU	0			
144	VENEZUELA	0	36.78	39.08	
145	W.SAMOA	0			
146	YEMEN	0			
147	YUGOSLAVIA	0	50.91	53.97	

148	ZAIRE	0	30.03	33.53
149	ZAMBIA	0	30.54	36.61
150	ZIMBABWE	0	35.97	39.64