The Income Gap in Cognitive Skills in Rural Pakistan*

Harold Alderman
World Bank

Jere R. Behrman
University of Pennsylvania

Shahrukh Khan
Sustainable Development Policy Institute, Pakistan

David R. Ross
Bryn Mawr College

Richard Sabot
Williams College

I. Introduction

The schooling of today's children has important effects on the productivity, income, and health of tomorrow's adults.[1] The distribution of schooling among households today shapes the distribution of income and of other indicators of well-being among tomorrow's households. Basic schooling is heavily subsidized in most societies. This does not mean, however, that the distribution of schooling—or, more important, such products of schooling as cognitive achievement—is unaffected by the distribution of real income among households. To the contrary, in most societies household income and the cognitive achievement of children are positively associated. To the extent that real household income is correlated with the quantity and quality of schooling services, raises the household's demand for schooling, and enhances the capacity of children to convert educational opportunities into cognitive skills, schooling may actually perpetuate or even worsen welfare differences between members of low- and high-income households.

Evidence regarding the influence of household income on schooling

0013-0079/98/4601-0004$02.00

attainment and cognitive achievement in developing countries is quite limited. Most studies simply relate a measure of the quantity of schooling (e.g., age-specific schooling progress or schooling attainment) to household income and report significantly positive coefficient estimates. These studies have several limitations: (1) They tend to report only the statistical significance of the income coefficient estimates without communicating whether such effects are substantial. (2) They tend to use current income, which may bias estimated income coefficients toward zero if more permanent income constraints are more relevant.[2] And (3) they focus exclusively on one input into the production of cognitive achievement—duration in schooling—and ignore other inputs into cognitive achievement such as school quality, household environment, and community characteristics that may change with household income and affect cognitive achievement independent of the duration of schooling. Failure to control for such factors may bias substantially the estimated income effects.[3]

For this study, we generated data in rural Pakistan to investigate, for the first time, the impact of real household permanent income on cognitive achievement. The investigation has two components. First, we observe that respondents born to households in the top quartile of permanent income for our sample scored higher on tests of cognitive skills than did respondents from households in the bottom quartile. We refer to this difference between the mean cognitive skills test scores for the top income quartile and for the bottom income quartile as ‘‘the income gap in cognitive skills.’’ This gap is large: 40.4% of the average score for high-income households for the 20–24 age cohort, and 53.1% for the 30–44 age cohort.

Second, we decompose this income gap in cognitive skills to investigate the extent to which the gap can be explained by income-associated differences in (*a*) the availability of schools, (*b*) the probability that children start school (conditional on schools being available), (*c*) schooling attainment (conditional on starting school), and (*d*) school quality, household environment, or community characteristics (conditional on expected schooling attainment). In this decomposition, we distinguish two effects: the ‘‘direct income effects’’ that would result from changing real permanent income for low- and high-income households to that of average households and the ‘‘income-associated effects’’ that would result from changing other household characteristics from the levels of low- and high-income households to average levels for all households. This distinction is important in assessing the effectiveness of alternative policies designed to close the income gap in cognitive skills. Our results suggest that direct income effects are less important than previously thought and that most of the direct effects are through creating a home environment that is conducive to learning rather than through stimulating the demand for school.

Section II presents our conceptual framework for analyzing and decomposing the cognitive achievement gaps associated with income. Section III introduces the data that we use for our analysis. Section IV gives our estimates of the underlying critical relations and our decompositions of the cognitive achievement gaps associated with income. Section V provides conclusions.

To put our results in context, it is useful to look at the income distribution in Pakistan in general and in rural areas in particular. Among 36 low- and middle-income economies, Pakistan has the tenth-largest share of income received by the bottom quintile and the seventh-smallest share of income received by the top decile.[4] If the comparison is limited to the 15 low-income countries, in order to have a group of countries with more similar levels of development, Pakistan is at the median (eighth) with regard to the size of the share of the bottom quintile and has the fifth-smallest share for the top decile. Thus, these data suggest that income is distributed relatively equally in Pakistan, at least with regard to the extent of concentration in the tails of the distribution (with relatively high shares at the bottom and relatively low shares at the top), in comparison with all other developing countries, as well as somewhat more equally than in other low-income countries.

However, estimates for Pakistan indicate a relatively low concentration of households below the poverty line in rural areas, although there are substantial comparability problems in establishing the poverty levels across countries.[5] The estimates also indicate that the rural poverty rates in Pakistan, while not declining as rapidly as in China, Indonesia, and Korea, declined in recent years more rapidly than in the majority of Asian countries for which such comparisons can be made. Both the extent of overall income inequality and the share of the rural population below the poverty line in Pakistan (and the recent change in the latter) are at or below the median for developing countries in general and comparable countries in Asia in particular.

These comparisons are based on measures of income for 1 year, because, in general, those are the only available data. In much of rural agriculture in developing countries, however, transitory income fluctuations may be considerable because of weather shocks and limited ability to buffer them. If longer-run income measures are more likely to constrain longer-run investments in schooling, the extent of income variance in the annual data may be a misleading predictor of the influence of permanent income on cognitive achievement.[6] Moreover, if current income is influenced by the productivity of respondents who have completed schooling, then it will be a positively biased proxy for income at the time schooling decisions were made.

Therefore, for use in our analysis, we have calculated measures of permanent income at the time the respondent was of school age. To do so, we first regressed current household income on parental characteris-

TABLE 1

Household Permanent Income at Age of Schooling (Thousand 1988 Rupees)

	Age Cohort	
	20–24 Years	30–44 Years
75th percentile	29.018	26.527
Median	23.378	22.060
25th percentile	21.011	20.773

tics, including education, employment, and acreage farmed, if any. We then used the parameters of this equation, together with measures of the corresponding variables for each respondent's parents, to obtain a measure of the parents' permanent income (in 1989 rupees).[7] Among the 846 households in our sample for which permanent income data could be calculated in this procedure, average household income was just over 23,000 rupees with a standard deviation of 13,500 rupees. Thus, there remains substantial variation in our permanent income measure with which to conduct our cognitive skills decomposition exercise. Table 1 summarizes these data by presenting permanent incomes for households at the 75th percentile (households with the lowest incomes in the highest quartile) and at the 25th percentile (households with the highest incomes in the lowest quartile) for the two age cohorts that we consider below. The household permanent income at the 25th percentile is 72% and 78% of that at the 75th percentile.

The gap in mean household income between the top and bottom income quartiles is 51.1% for the 20–24 age cohort and 47.1% for the 30–44 age cohort.[8] The corresponding cognitive skills gaps, noted above, are 40.4% and 53.1%, respectively. Thus, income inequality in our sample is associated with substantial inequality in an important dimension of human capital.

II. Conceptual Framework

Cognitive skills will vary with income to the extent that income varies with the availability of reasonably accessible schools, the household's demand for schooling, and the quality of instruction in schools as supplemented by out-of-school investments by the household. We consider these in turn.

School Availability

In rural Pakistan, primary school (in general, kindergarten plus grades 1–5) is followed by middle school (grades 6–8). Students continuing in school usually complete two more grades (9 and 10) before sitting for the matric exam, two more grades (11 and 12) before taking the FA or

FSc exam, and two more grades (13 and 14) before taking the BA or BSc exam. Preparation for these exams occurs at secondary schools and postsecondary institutions that, for many sample households, are located in towns some distance from home. Few respondents in our sample have taken these exams. Therefore, we focus on the availability of primary and middle school.

Although in rural Pakistan public primary school establishment decisions are made beyond the village-government level, they plausibly may be in responses to local characteristics such as income or political power.[9] The possibility of schools being available in response to income, of course, is important from the point of view of this study. The local availability of social services provided by higher governmental units may be in response to political pressures, in which case higher-income villages might be expected to have higher probability of having schools available. Or availability may reflect equity concerns, in which case higher-income villages might be expected to have lower probability of having schools available.[10]

Estimation of governmental preferences for the allocation of school resources across communities in Brazil indicates that combined (i.e., federal, state, and municipal) governmental allocations of teachers at differing schooling levels are consistent with the combined governments making some trade-off between productivity and equity in such allocations. But they do not appear to weigh equity enough to make investments compensatingly larger in poorer than in richer areas. Similar estimates for Mexico in 1980 and 1990 suggest that the federal government allocates such resources with concern about both equity and productivity and that the concern about equity probably increased over the decade. Finally, standard cross-national estimates of school enrollment decisions assume that schooling importantly reflects income and price demand determinants that manifest themselves through the political process to affect school availability and school quality.

To examine the importance of these effects in rural Pakistan, we estimate relations for the determination of primary school availability over time in different sample villages that are dependent, inter alia, on average village sample income. The data used for these estimates are discussed in Subsection III.A, and the estimates themselves are presented and discussed in Subsection IV.A.

Household Schooling Demands Conditional on Local School Availability

Conditional on availability, whether a household demands schooling depends on the balance between the costs of schooling and the expected benefits. Schooling costs depend on prices for out-of-pocket expenditures, the opportunity cost of time spent in school, and perhaps other household characteristics, including income and parental schooling. The expected benefits are assumed to be primarily the increments in expected marginal product, which is posited to be mainly a function of cognitive achievement. Cognitive achievement, in turn, depends on individual attributes of the child (gender, age, and potential to learn), family background variables that determine the home learning environment (income and parental education), and the quality of the school system.

Income can affect the demand for schooling through several channels, though, a priori, the total effects may be either positive or negative. If schooling in part is a normal consumption good, ceteris paribus, higher income would increase the demand for schooling. If such consumption is an important part of the motive for schooling, and this consumption is viewed as a luxury (plausible for very-low-income households), this effect could be large. If schooling is viewed, in part, as an investment and capital markets are imperfect, then, ceteris paribus, income would positively affect the demand for schooling. Higher income would permit financing of more time in school, a household environment that is more conducive to learning, and better information regarding the expected returns to school.[11] Higher-income households can afford to purchase better information about the costs and benefits of schooling.[12] On the other hand, given the high contribution of opportunity cost to the total cost of schooling, if there are high transaction costs for children working away from the family farm, and if income is associated with land assets, then, ceteris paribus, the relationship between household income and the demand for schooling would be negative.

In addition, there are a number of reasons why schooling may be associated with income in a cross-section of households, even if these reasons do not imply a causal relationship between income and the demand for schooling. That is, in standard analysis, income may be a proxy for other variables, so that the estimated effect of schooling also includes the correlated effects of these other variables. Four examples, all holding other factors constant, are (1) tastes for consumption and work intensity may vary with income and the household may invest in schooling to increase the future command over income; (2) there may be abilities correlated across generations that affect both success in school and adult economic success; (3) better-schooled parents may more effectively process signals regarding the expected returns to schooling (particularly given rapidly changing markets and technologies), and the schooling of parents may be associated with household income; and (4) family connections in the labor market may affect the expected rate of return to schooling, and family connections may be correlated with income.[13]

Separating the causal effects from the correlative effects of income on household schooling demands is crucial to assessing what would happen if income were to change. Although in our data set (and even more so in most other data sets) this is difficult to do, we control for some of the most important correlations. For example, we control directly for parental schooling. We also control, at least partially, for ability, a variable that usually is unobserved.

We estimate a household demand function for schooling with household income included among the possible explanatory variables. Because such a large proportion of our sample has no schooling at all, we model the demand for schooling in two steps. First, we estimate as a probit, conditional on the availability of a primary school, the probability that a respondent attends at least one year of primary school. We assume there is some unobserved index of the net benefit of starting school, which in turn is a linear function of the determinants of the demand for schooling and a normally distributed error term. If this index is greater than zero, we then observe the respondent attending at least one year of school. The probit technique estimates coefficients of the net benefit function to maximize the likelihood that the fitted value of the index exceeds zero if a respondent reports at least one year of schooling.

We then model schooling attainment, conditional on starting school, as an ordered probit. We assume there is an unobserved index of the net benefit of schooling, which in turn is a linear function of the determinants of the demand for schooling and a normally distributed error term. As this index exceeds successively higher thresholds, the respondent completes additional years of schooling. The ordered probit technique estimates values for these thresholds and the coefficients of the net benefit function to maximize the likelihood that the fitted value of the index falls between the thresholds corresponding to the respondent's particular level of schooling.[14]

School Quality and Home Learning Environment

Cognitive skills are determined by a production function reflecting individual characteristics, years of schooling, school quality, and out-of-school family investments. We include income in our representation of the home learning environment, in addition to parental schooling, to represent dimensions of that environment that affect learning and that are purchased with income but are not directly observed in the data (e.g., reading material, radios). This representation may overstate the direct impact of home income on cognitive achievement because income, in part, may be representing unobserved income-associated household factors.

Preliminary estimates demonstrated that the coefficients of a linear cognitive skills production function are sensitive to the choice of explanatory variables. Because the level of schooling, school quality, the individual's potential to learn, and family inputs all interact, we use a multiplicative (Cobb-Douglas) functional form for the production function—with gender, regional, and school dummies serving as shifters for the constant. We control for the simultaneous determination of cognitive skills and schooling attainment by using predicted schooling attainment from the ordered probit model to instrument years of schooling in the cognitive skills production function.[15]

In Subsection III.B we discuss the data used for the estimation of the cognitive achievement production functions. In Subsection IV.B we present and discuss the estimates.

Decomposition of Income Effects on Cognitive Achievement

We use our analysis of the determinants of school availability, schooling attainment, and cognitive skills to predict expected cognitive skills of individuals at the sample means for respondents from low- and high-income households.[16] Details of the calculations appear in Subsection IV.3. From this we calculate the expected income gap: the difference between high-income cognitive skills and low-income cognitive skills as a percent of high-income cognitive skills. In our decomposition simulations we successively equate each of the component estimates to the full sample means and thereby measure the contribution of that component to the overall income gap in cognitive skills.

III. Data

Since 1986, the International Food Policy Research Institute (IFPRI), under the auspices of the Pakistan Ministry of Food and Agriculture, has been administering a multipurpose survey to a panel of over 800 rural households containing more than 7,000 individuals drawn from villages in two districts (Attock and Faisalabad) of the Punjab, one district (Dir) of the North West Frontier Province, and one district (Badin) of the Sind.[17] Human-capital modules, on which we draw heavily for this study, were administered in the spring of 1989, the tenth round of the survey. These modules contain inter alia school characteristics (including when schools became locally available) and household and individual characteristics (including household income, parental schooling, prices, schooling attainment, reasoning ability, and postschooling cognitive achievement).

A. Data for Village-Level School Availability Relations

Table 2 contains descriptive statistics for the 43 villages for which we could accumulate reasonably complete data. Of the 43 villages, 51.2% had primary schools in the village for boys age 10–15 at the time of the survey; 21% had primary schools for girls in the same age cohort.[18] Permanent village-level mean household income averaged 24,800 rupees per year, which is close to the average for all households in our sample. The mean effective distances to Mandi and to market were about 12 kilometers, to the Tehsil capital about 16 kilometers, and to the district capital more than 40 kilometers.[19] Almost half (47%) of the villages had a union council member, about a fifth had a government official, but only 7% had a district council member and only 2.3% (one village) had a

TABLE 2

Village-Level Descriptive Statistics (Means and Standard Deviations)

	Full Sample	District: Attock	Faisalabad	Dir	Badin
Village school availability (%):					
Primary for boys	51.2	37.5	66.7	90.0	31.6
Primary for girls	20.9	25.0	16.7	50.0	5.26
Effective distances (Km):					
Distance to Mandi	12.3 (7.7)	9.4 (3.4)	13.2 (10.6)	10.5 (7.2)	14.1 (8.3)
Distance to market	11.9 (6.5)	9.4 (3.4)	9.2 (4.3)	8.2 (3.1)	14.1 (8.3)
Distance to Tehsil capital	15.9 (11.0)	17.9 (15.5)	13.8 (10.6)	10.3 (5.0)	18.6 (10.7)
Distance to district capital	43.6 (30.5)	50.9 (33.5)	49.0 (13.6)	45.3 (48.4)	37.9 (21.2)
Indicators of political links (%):					
Union council member	46.5	62.5	100.0	50.0	21.1
District council member	7.0	12.5	0.0	10.0	5.3
Government official	20.9	12.5	50.0	30.0	10.5
Member of parliament	2.3	0	16.7	0	0
Income, land, and population:					
Mean household income (in rupees)	24.8 (5.8)	22.8 (1.7)	22.7 (2.3)	34.2 (3.0)	21.4 (2.7)
Land area (acres)	3,057.3 (3,355.3)	5,110.6 (6,922.5)	1,991.7 (932.8)	1,513.0 (1,448.2)	3,342.0 (1,636.8)
Population	1,866.1 (1,814.0)	1,775.0 (1,805.3)	4,416.7 (2,102.8)	1,591.1 (1,975.8)	1,243.7 (779.8)
Households	323.7 (225.0)	278.8 (152.1)	540.5 (275.3)	109.6 (60.8)	386.8 (199.9)
Number of villages	43	8	6	10	19

Note.—Data in parentheses are standard deviations.

member of parliament. Village land area averaged 3,057.3 acres, and village populations averaged 1,866 individuals in 324 households.

There is considerable intervillage variation in these characteristics, much of which is interdistrict. As is indicated in the last four columns of table 2, for example, sample mean household incomes tend to be relatively high in Dir; sample villages in Dir tend to have relatively great school availability and those in Badin tend to have relatively little school availability; sample villages in Badin tend to be relatively far from Mandi, market, and Tehsil capitals but close to the district capital; sample villages in Faisalabad tend to have more governmental connections except for district council members; and sample villages tend to be relatively large in land area in Attock and relatively large in population in Faisalabad.[20] But there also are considerable intradistrict variations across some of these characteristics, such as the distance to the district capital in Dir and the mean land area in Attock.

B. Data for Individual Schooling Demand and Cognitive Achievement Production Function Relations

For our analysis of the income gap in cognitive skills, we sought the youngest of the age cohorts in which all respondents had completed formal schooling and for which one could assume that school and community characteristics were reasonably unchanged.[21] For this 20–24 age cohort, table 3 presents means and standard deviations of key variables for the 330 respondents, for whom we have useable data.[22] For intertemporal comparison, we include data on the 30–44 age cohort—chosen to include comparable numbers of schooled respondents, despite lower enrollment rates in the past.

Family income. We use predicted rather than observed household income for the three reasons discussed in the introduction: there are large transitory fluctuations; a priori, a permanent income measure seems more appropriate for analysis of long-run investments such as in schooling; and schooling attainment may bias current income as a measure of household income at the time the schooling decision was made.[23] Predicting income on the basis of parents' assets and other characteristics as is discussed above yields an unbiased measure of permanent household income. The substantial difference in income between the subsamples is the key to our efforts to measure the direct effect of income on cognitive skills.

Gender. Women account for a larger portion of the 30–44 age cohort. For the 20–24 age cohort, the representation of women is relatively higher in the high-income sample, while for the 30–44 age cohort this is reversed. We are unaware of any systematic explanation for this reversal.[24] The distribution of women in our sample tends to depress the observed income gap for the younger age cohort and accentuate it in the older one.

TABLE 3

INDIVIDUAL LEVEL DESCRIPTIVE STATISTICS (Means and Standard Deviations)

VARIABLE	20–24-YEAR COHORT			30–44-YEAR COHORT		
	Full Sample	Low Income	High Income	Full Sample	Low Income	High Income
Family income (rupees)	25.643 (8.78)	17.717 (2.80)	36.227 (7.26)	24.556 (7.95)	18.455 (2.08)	34.865 (7.38)
Female	.539	.539	.590	.558	.573	.528
Attock	.197	.174	.090	.195	.156	.101
Faisalabad	.197	.330	.080	.141	.124	.082
Dir	.303	.043	.770	.185	.013	.642
Age (years)	21.924 (1.24)	21.983 (1.18)	21.920 (1.28)	36.475 (4.43)	35.831 (4.26)	36.750 (7.38)
Reasoning ability	18.139 (7.01)	18.052 (7.06)	18.130 (6.80)	17.326 (6.69)	17.084 (6.87)	17.610 (6.77)
Reading score	4.309 (7.03)	3.197 (5.99)	5.214 (7.85)	2.026 (5.07)	2.002 (3.58)	3.878 (6.91)
Math score	4.245 (6.82)	2.967 (5.46)	5.133 (7.90)	1.831 (4.63)	1.467 (3.17)	3.525 (6.28)
Schooling attainment (years)	2.945 (4.44)	2.078 (3.55)	3.310 (4.83)	1.287 (3.36)	.564 (2.14)	2.522 (4.63)
Mother primary schooling or more	.015	.009	.020	.005	.0	.019
Father middle schooling or more	.124	.026	.200	.046	.022	.075
Primary school available	.679	.678	.650	.440	.404	.459
Middle school available	.630	.630	.604	.336	.310	.351
Distance to primary school (km)	17.233 (17.25)	17.332 (15.52)	17.980 (20.69)	10.785 (13.73)	10.278 (13.80)	9.890 (12.65)
Distance to middle school (km)	20.679 (19.33)	19.184 (17.14)	23.089 (22.39)	12.235 (17.12)	9.424 (15.66)	15.798 (19.67)
Primary school book cost proxy	35.760 (29.30)	38.883 (29.75)	30.044 (27.99)	25.244 (30.57)	23.935 (30.19)	21.791 (27.28)
Middle school book cost proxy	145.501 (105.04)	158.758 (117.48)	135.312 (85.12)	89.701 (112.56)	74.064 (111.21)	95.689 (103.93)
N	330	115	100	568	225	159

NOTE.—Data in parentheses are standard deviations.

Regional differences. There are substantial regional differences between the low- and high-income samples. Respondents from Dir represent a large portion of the high-income subsamples, while respondents from Badin (the residual category in table 3) comprise a large portion of the low-income subsample.[25] A substantial portion of the indirect effect of income on cognitive skills may reflect these regional differences.

School availability. Whether a school was available in the village at the time the respondent had attained school age was indicated by our school survey (Subsection III.A). If a respondent was from a village not included in the school survey, we proxied school availability by determining the earliest date a respondent from the village attended school in or near the village.[26] Only 44% of the older cohort had access to a primary school at the time respondents were of school age. Respondents from high-income families were more likely to have a primary and middle school available than were respondents from low-income families.[27] Fewer than 70% of the younger cohort had access to school. Children from low-income families had a slightly higher probability of having a primary and middle school available.

Schooling attainment. The average number of years of schooling completed is low for both cohorts. For the 20–24 age cohort, respondents from low-income households completed an average of 2 years of schooling—more than a full year less than respondents from high-income households. The gap between subsamples was a full 2 years for the older cohort, with respondents from low-income households reporting an average of 0.5 years of schooling.

Cognitive achievement. Our measure of cognitive skills was generated by administering (in the regional language) tests of literacy and numeracy specially designed by the Educational Testing Service to every person in our sample older than 10 years of age and with at at least 4 years of schooling.[28] Among those who took the cognitive-skills test, the distribution of the scores was not truncated; it exhibits substantial variance and appears to be normal. The means indicate substantial income and cohort gaps in cognitive achievement. For the 30–44 age cohort, the income gap in cognitive skills is 53.1%.[29] Cognitive skills for both subsamples have risen for the 20–24 age cohort. The income gap has fallen to 40.4%, still a substantial level.

Reasoning ability. To obtain a measure of reasoning ability, we administered Raven's Coloured Progressive Matrices (CPM), a test that involves the matching of patterns, to everybody in the sample older than 9 years of age.[30] The test is designed so that formal schooling does not influence performance, although performance may reflect early childhood environment as well as innate capacity. The distribution of the CPM test scores is not truncated at either tail; it exhibits substantial variance and appears to be normally distributed. The disaggregated distributions for Dir, the Punjab, and Badin are very similar. Since schooling levels differ substantially across regions, this similarity is consistent with the presumption that schooling attainment does not influence performance on Raven's CPM test. This test has been used to control for ability in estimates of the determination of cognitive achievement in other economic and noneconomic analyses.[31] In our analysis, we treat the Raven's score as predetermined. Because of some controversy over what this test score means, however, we also have undertaken parallel estimates in which we exclude the Raven's score as an explanatory variable.[32] The magnitudes of the key estimates presented in Section IV change little.

Parental education. Parental schooling is substantially greater for the high-income subsamples. For the 20–24 age cohort, 20% of respondents from high-income families had fathers who had completed middle school, while fewer than 3% had fathers who had achieved that level of education in the low-income subsample. Two percent of respondents from high-income families had mothers with primary schooling, while fewer than 1% from low-income families had educated mothers. Parental schooling for the 30–44 age cohort is substantially lower. Note that no respondent from a low-income family in the older cohort had an educated mother. The substantial differences in parental schooling by income subsamples mean that control for parental schooling may be critical in our estimates in order to avoid attributing to direct income effects what in fact may be parental schooling effects.

Price of schooling. To calculate distance (time) to the nearest available school, we used the average travel time for children currently in school in each village as a proxy for the travel time of all respondents. Where there were fewer than five children in school for a village at the time of the survey, we used as a proxy the mean travel time conditional on district and whether the nearest school is located in the village or a nearby village or town.[33]

Expenditures on books and school supplies are dependent not only on the school system but also on the household's preferences and income, and, consequently, they have endogenous components. To obtain a proxy for the exogenous cost component, we estimated educational expenditure functions, including a vector of household characteristics and dummy variables for district, level of schooling, gender, and whether the school was located in the village or a nearby town. The household variables were then held constant to predict exogenous costs, i.e., prices.

IV. Estimates and Income Gap Decomposition

To decompose the income gap in cognitive skills, we require estimates of the direct income effects and income-associated effects on school availability, the probability of starting school (conditional on school availability), expected years of schooling (conditional on starting school), and expected cognitive skills (for given schooling attainment). In Subsection IV.A, we provide evidence that there are no direct income effects on school availability in our sample. Subsection IV.2 presents the estimates of the schooling demand and cognitive skills production relations.

A. Village-Level School Availability

Is village-level school availability in rural Pakistan associated with average village-level income? Table 4 presents probit estimates relating, by gender, primary school availability to village-level characteristics.[34] The probit in column 1 contains mean village-level household income, distances from political and urban centers, indicators of political influence, and village-size measures.[35] The Wald statistic indicates a significant relationship between these variables taken together and the probability of finding a school in that village. In this relation, for boys (though not for girls) mean village-level household income appears to have a significant positive effect on school availability. If district dummy variables are added (col. 2), however, the village-level characteristics become insignificant.[36] The implication is that interdistrict, not intradistrict, variations in the village characteristics affect school availability so that the school availability–income correlation observed in column 1 results is spurious.

Column 3 repeats the specification in column 1 but drops villages in Dir; in this reduced sample this specification has no explanatory power. A possible explanation is that there is greater, unobserved interest in schooling in Dir than in the other districts. This may reflect the fact that since Mogul times the Pathan regions have relied on out-migration (first as soldiers, then as urban laborers) to supplement the meager agriculture of the area, which likely adds to the perceived returns to schooling. In any case, these estimates suggest that, once there is control for region, village-level characteristics—including mean household income—do not significantly affect the probability that schools are available. Thus, the allocation of schools across villages appears to favor neither higher-income villages (which would seem to have more political power) nor poorer ones (which might be favored if equity considerations were weighed heavily), i.e., there are no direct income effects on school availability.

This has two implications for the remainder of our analysis. First, in the estimates in Subsection IV.B we consider the availability of locally available schools to be given from the point of view of individual households. Second, the sample means by income level for school availability from table 3 measure the income-associated effects for the decomposition in Subsection IV.C.

B. Estimates of Household School Demands and Cognitive Achievement Production Functions

Household school attainment demand estimates. Table 5 gives estimates of the probit regressions for the probability of starting school con-

TABLE 4

Probit Estimates of Impact of Village Characteristics on Primary School Availability

	Boys			Girls		
	1	2	3	1	2	3
Mean household income (rupees)	.103† (2.49)	−.148 (1.31)	−.034 (.34)	.076 (1.57)	.194 (1.48)	.127 (.79)
Distance to Mandi (Km)	.002 (.07)	.005 (.13)	−.134 (.34)	.093 (1.63)	.146 (1.61)	.138 (1.28)
Union council member	.972* (2.08)	.993 (1.58)	.891 (1.64)	.935 (1.52)	.945 (1.22)	1.988 (1.60)
Households	.001 (.89)	.002 (1.64)	.002 (1.62)	−.004 (1.58)	−.006 (1.67)	−.007 (1.35)
Attock	.345 (.50)	...	...	.934 (.92)	...	...
Faisalabad	.155 (.20)	...	...	.426 (.35)	...	...
Dir	4.71* (2.22)	...	...	−1.21 (.67)	...	...
Constant	−3.342	1.351	−.533	−3.412	−6.623	−4.647
N	43	43	33	43	43	33
Prediction ratio‡	17/22	16/22	7/13	4/9	3/9	1/4
Wald (df)	12.13* (7)	20.09† (4)	8.03 (4)	13.24* (7)	15.71* (4)	7.41 (4)

Note.—Data in cols. 1–3 explained in Subsection IV.A, Village-Level School Availability. Data in parentheses are standard deviations.
* Significant at 5% significance level.
† Significant at 1% significance level.
‡ Of villages with schools, the proportion correctly predicted by the model.

TABLE 5

PROBABILITY OF ATTENDING SCHOOL PROBITS BY GENDER AND SCHOOLING ATTAINMENT ORDERED PROBIT

VARIABLES	PROBABILITY OF ATTENDING SCHOOL		SCHOOL ATTAINMENT ORDERED PROBIT
	Men 10–24	Women 10–24	
Income	.031 (4.11)	.031 (1.04)	.017 (4.11)
Female	. . .	. . .	−.385 (3.97)
Reasoning ability	.057 (1.12)	−.037 (.25)	.035 (1.45)
Reasoning ability^2	.0006 (.47)	.003 (.77)	.0006 (1.04)
Age	.287 (4.09)	.276 (.99)	.128 (14.17)
Age^2	−.010 (4.81)	−.004 (.57)	−.002 (13.45)
Father primary schooling or more	.119 (.56)	−.513 (1.06)	.042 (.44)
Father middle schooling or more	.113 (.35)	1.323 (2.21)	.182 (1.49)
Middle school available	.321 (.45)	3.919 (.03)	2.880 (9.93)
Distance to primary school	−.042 (4.72)	−.405 (5.52)	.017 (5.09)
Distance to middle school	.024 (2.58)	.373 (4.42)	−.002 (.77)
Primary school book cost	−.060 (9.48)	−.051 (1.41)	−.005 (1.88)
Middle school book cost	−.0099 (.37)	−.034 (.001)	−.010 (8.82)
Attock	1.379 (1.73)	7.472 (.03)	.354 (1.26)
Faisalabad	1.051 (3.29)	.475 (.27)	1.022 (3.54)
Dir	−.996 (2.09)	4.218 (.07)	−.102 (.45)
Village 1	.450 (1.31)	10.735 (.005)	−.605 (2.03)
Village 2	.573 (1.61)	10.274 (.005)	−.283 (1.00)
Village 4	.132 (.35)	10.879 (.005)	.075 (.26)
Village 5	1.331 (3.16)	9.933 (.005)	−.208 (.76)
Village 6	.552 (1.43)	9.921 (.005)	−.246 (.88)
Village 8	−.152 (.17)	4.871 (.002)	.361 (1.25)
Village 9	−.716 (.75)	3.063 (.001)	.256 (.84)
Village 10	−.948 (.98)	3.821 (.002)	.221 (.74)
Village 11	0.155 (.16)	−3.282 (.42)	−.153 (.49)
Village 12	.898 (.87)	2.625 (.001)	−.275 (.81)
Village 13	−.379 (.46)	4.629 (.002)	−.248 (.84)
Village 14	−.026 (.02)	1.772 (.001)	.017 (.04)
Village 23	−.469 (.12)	−2.963 (1.32)	.215 (.76)
Village 26	−.267 (.46)	. . .	.426 (1.24)
Village 27	.289 (.55)	9.401 (.001)	.920 (3.68)
Village 29	.637 (1.41)	−.105 (.02)	.089 (.33)
Village 41	−.184 (.37)	2.682 (1.86)	−.309 (1.46)
Village 44	1.234 (2.02)	. . .	−.278 (1.03)
Village 45	.560 (.74)	2.891 (1.90)	.033 (.13)
Village 46	7.023 (.009)	. . .	.886 (.32)
Village 47	.709 (1.13)	−.660 (.32)	.267 (.99)
Village 48	6.924 (.01)	1.419 (.47)	.222 (.80)
Village 49	3.863 (.005)	−.780 (.55)	.188 (.59)
Village 50	.292 (.37)	3.523 (2.62)	−.130 (.49)
Grade 5 threshold	. . .	. . .	.520
Grade 6 threshold	. . .	. . .	1.40
Grade 7 threshold	. . .	. . .	1.66
Middle (8) threshold	. . .	. . .	1.92
Matric (10) threshold	. . .	. . .	2.56
FA/FSc (12) threshold	. . .	. . .	3.31
BA/BSc (14) threshold	. . .	. . .	3.96
Log-likelihood	−446.8	−263.6	−1828.0
N	745	384	1,071

NOTE.—*t*-Statistics in parentheses.

ditional on schools being available and the ordered probit for schooling attainment conditional on starting school.[37] In the latter, the thresholds indicate the value of the schooling index needed to move the respondent to progressively higher levels of schooling. To ensure that the low direct income effects we found were not artifacts of imprecise estimates, we expanded the sample for table 5 beyond the 20–24-year age cohort. In particular, we use, separately, males and females age 10–25 years with a primary school available for the probit estimates and all respondents who started school for the ordered probit.

Of primary interest is the estimated impact of household permanent income. Higher household income significantly increases the probability of starting school for males but not for females.[38] We do not have an explanation for the gender difference in effect of income on the probability of starting school, although it is not inconsistent with the discussion in Section II above. Conditional on starting school, higher household permanent income also significantly increases schooling attainment. However, for a respondent in the 20–24-year age cohort, we estimate that increasing household income from the low- to high-income household average raises the schooling index by only .31 points, enough to raise schooling attainment by one grade for middle school students but representing less than a grade step increase for higher levels.

A number of the other variables that enter into our model are associated with income (as is discussed in Sec. II) and enter into our income-associated calculations below in Subsection IV.C. Some of them have significant effects on starting school and on schooling attainment. If the father had at least primary schooling, for example, it will significantly affect the probability of starting school for females. Therefore, without control for paternal schooling, the direct income estimates probably would be upward biased because in part they would be representing paternal schooling.[39]

Cognitive achievement production function estimates. Table 6 gives estimates separately for math and reading for respondents age 10–24. As noted in Section II, we use a variant of the Cobb-Douglas form in which cognitive skills are the product of schooling attainment, reasoning ability, an index of the home learning environment, and a shift term that varies by gender and village. More explicitly, we regress the math or reading score on $(\alpha_0 + \alpha_1\text{Female} + \alpha_2\text{Attock} + \alpha_3\text{Faisalabad} + \alpha_4\text{Dir})^{\alpha 5} \times (\text{Preschool Ability})^{\beta 1} \times (\text{Schooling Attainment})^{\beta 2} \times (\text{Home Learning Environment})^{\beta 3} + \epsilon$, where Home Learning Environment $= \gamma_0\text{Income} + \gamma_1(\text{Father Primary or More}) + \gamma_2(\text{Mother Primary or More})$; the α's, β's, and γ's are parameters to be estimated, and ϵ is an error term. There are more coefficients than variables, so one normalization is necessary; we set $\gamma_0 = 1$. An increase in income raises the value of the home learning environment index, and hence directly increases cognitive skills.

TABLE 6

Nonlinear Least Squares Cognitive Skills Production Functions for Math and Reading, Respondents 10–25 with ≥4 Years of School

Variables	Math	Reading
Home Learning Environment	.114 (1.40)	.376 (3.73)
Income	1.0	1.0
Father primary schooling or more	35.649 (.75)	−.984 (.24)
Mother primary schooling or more	−19.819 (7.30)	1.616 (.21)
Female	−0.363 (0.97)	.119 (1.12)
Attock	−.136 (.83)	.099 (.84)
Faisalabad	.490 (1.83)	.276 (1.58)
Dir	.387 (1.68)	−.223 (1.57)
Reasoning ability	.502 (3.62)	.402 (2.89)
Schooling attainment*	.249 (2.17)	.256 (2.11)
Boys, village 1	.109 (.62)	−.306 (1.59)
Boys, village 2	−.403 (1.48)	−.371 (1.57)
Boys, village 3	−.405 (1.50)	−.544 (1.82)
Boys, village4	.028 (.11)	−.449 (1.69)
Boys, village 5	−.286 (1.38)	−.066 (.51)
Boys, village 6	−.129 (.70)	−.294 (1.59)
Boys, village 7	−.216 (1.02)	−.125 (.84)
Boys, village 8	.299 (1.28)	.146 (.96)
Boys, village 9	−.208 (1.07)	−.182 (1.21)
Boys, village 10	−.003 (.02)	−.150 (1.10)
Boys, village 11	−.096 (.53)	−.086 (.63)
Boys, village 13	−.223 (.99)	−.197 (1.10)
Boys, village 14	−.051 (.20)	.070 (.39)
Boys, village 23	.048 (.19)	.080 (.46)
Boys, village 27	.078 (.46)	.091 (.68)
Boys, village 41	−.160 (.83)	−.051 (.46)
Boys, village 42	−.418 (1.35)	−.113 (.71)
Boys, village 44	.363 (1.34)	−.090 (.53)
Boys, village 47	.138 (.73)	.174 (1.23)
Boys, village 48	−.008 (.03)	−.118 (.81)
Boys, village 50	−.316 (1.29)	−.065 (.51)
Boys, village 51	.088 (.50)	.252 (1.49)
Girls, village 2	−.276 (1.03)	−.307 (1.32)
Girls, village 8	−.095 (.33)	.197 (.92)
Girls, village 9	−.536 (1.20)	−.329 (1.10)
Girls, village 10	−.174 (.71)	.101 (.59)
Girls, village 12	.298 (.98)	.247 (1.08)
Constant	.890	.663
R^2	0.32	.30
N	317	317

Note.—*t*-Statistics in parentheses.
* Instrumental variable from table 5.

The estimates indicate that this effect is significantly positive for reading, though not for mathematics.[40] An increase in income also increases cognitive skills in both reading and mathematics through the significant effect of schooling attainment (based on the estimates in table 5). In addition, there are income-associated effects both through the other inputs in table 6 and through the schooling-attainment estimates from table 5.

C. *Decomposition of the Income Gap in Cognitive Achievement*

To decompose the income gap in cognitive skills, we begin by predicting the expected cognitive skills of individuals at the sample means for respondents from low- and high-income households. This calculation requires estimates of the following components:

- *A*1 the probability that a primary school is available,
- *A*2 the probability that a middle school is available,
- *B*1 the probability that the individual starts school,
- *B*2 conditional on starting school, expected schooling attainment,
- *C*1 conditional on expected schooling attainment, predicted cognitive skills, and
- *C*2 expected cognitive skills.

For the base case, components *A*1 and *A*2 are the averages for the low- and high-income samples. To calculate *B*1, we first substitute the sample means for the variables in the schooling probit, except that we use the availability of middle school conditional on the availability of a primary school. This yields an estimate of the conditional probability that respondents start school. Multiplying this by *A*1 yields *B*1. To calculate *B*2, we substitute the sample means into the ordered probit (again adjusting middle school availability to reflect the availability of a primary school). This yields a predicted value for the net benefit of schooling index. By comparing this with the thresholds in the ordered probit in the last column of table 5, we derive the appropriate predicted years of schooling. Substituting this value into the cognitive skills production function, along with other variables at the sample means, yields *C*1. Multiplying *C*1 by *B*1 yields the calculation of the expected cognitive skills for an individual at the sample means, which in turn allows us to calculate the income gap in cognitive achievement.

The direct income effect is obtained by equating the value of the income variables to the full sample mean. The additional income-associated effect is obtained by equating all the remaining variables to their sample means. This income-associated effect captures the impact of all right-side variables that we use in our estimates in tables 5 and 6 (except for household income) that are correlated with income.

What does the decomposition in table 7 suggest about the importance of the major underlying components in explaining the income gap in cognitive achievement?

First, differences in school availability do not explain the income gap in cognitive skills. As the means in table 3 suggest, there is virtually

TABLE 7

Decomposing the Income Gap in Cognitive Skills for 20–24-Year Age Cohort

Component	Percentage
Base case	49.4
Equating school availability	51.5
Equating probability of starting school:	
Direct income effect	44.1
Direct and income-associated effects	16.8
Equating expected years of schooling:	
Direct income effect	13.8
Direct and income-associated effects	12.8
Equating expected cognitive skills:	
Direct income effect	−3.3
Direct and income-associated effects	.0

no difference in school availability by income for the 20–24-year age cohort, and schools are somewhat closer for lower-income households.[41]

Second, direct income effects account for only about half (54%) of the total income gap in cognitive achievement.[42] The most important component of the direct income effect is through changing the home environment in a manner that works directly through the cognitive achievement production function, conditional on school attainment. This accounts for about three-fifths of the total direct income effect, but it may be an overestimate of the direct effect because income may be representing in part income-associated household characteristics. Second in importance in the total direct income effect is the probability of starting school (accounting for 28% of the total direct income effect).

Third, income-associated effects are about as important as the direct income effects in the total income gaps in cognitive achievement in our estimates, accounting for 55% (and may be larger in reality, since, as noted above, some of their effects may be captured in our direct measures).[43] Virtually all of this total income-associated effect works through the probability of starting school.

V. Conclusions

In this study, we use a specially generated data set from rural Pakistan to decompose the substantial income gap in cognitive achievement for a cohort of recent graduates of the school system. To our knowledge, this is the first such study for any society. Our results confirm the positive association between household income and the schooling of children but suggest that, over time, the income gap in cognitive achievement due to schooling is narrowing in rural Pakistan. For the 20–24 age cohort, a 51.1% income gap between high- and low-income households is associated with a 40.4% gap in cognitive skills. For those in the 30–44 age cohort, a 47.1% income gap is associated with a 53.1% gap in cognitive skills.

Our estimates suggest that school availability is influenced by income differences and therefore is not a factor in explaining the income gap in cognitive skills. Direct income effects and income-associated effects account for roughly equal shares of the income gap in cognitive skills in our estimates, and the former may be representing part of the latter. Our estimates have two important implications. First, the substantial income gap in cognitive achievement is not reflective of direct income effects alone. Simple associations of cognitive achievement with income would overstate substantially, at least by a factor of two, the probable impact of direct income increases on cognitive achievement. This is so because the income-associated determinants would not change just because income changed and because even our estimated income effects may be representing in part preference heterogeneity and other factors that are discussed above in Section II. High-income households are more likely to enroll their children in school not because they have more income but because their characteristics (not income determined) make schooling more attractive or feasible.

Second, although previous studies have focused on the direct income impact on starting school or on school attainment, our simulations suggest that most of the estimated direct income effect is through creating a home or school environment that is conducive to learning, given schooling attainment, not by affecting the probability of starting school or schooling attainment.[44] Policies designed to narrow the disparity in school availability or school attainment across income groups will be far less effective in closing the income gap in cognitive skills than simple cross-tabulations among these indicators would suggest. Research should focus on how school quality and the home learning environment enters the cognitive skills production function.

Notes

* We are grateful to the World Bank and the U.S. Agency for International Development (USAID) (through a contract with the International Food Policy Research Institute [IFPRI]) for financial support; to Mary Bailey, Emily Mellott, John Parsons, and Amy Whritenour for able research assistance; and to three anonymous referees for useful comments on an earlier draft. The views presented here are ours and should not be interpreted as reflecting the views of IFPRI, USAID, or the World Bank.

1. A sampling of recent studies and surveys that emphasize the importance of schooling in developing countries includes Robert J. Barro, ‘‘Economic Growth in a Cross-Section of Countries,’’ *Quarterly Journal of Economics* 106 (May 1991): 407–43; Jere R. Behrman, *Human Resource Led Development?* (New Delhi: Asian Regional Training and Employment Programme/International Labour Organization, 1990); Nancy Birdsall and Richard H. Sabot, ‘‘Virtuous Circles: Human Capital, Growth, and Equity in East Asia’’ (World Bank,

Washington, D.C., 1994, mimieographed); Elizabeth M. King and M. Anne Hill, eds, *Women's Education in Developing Countries: Barriers, Benefits, and Policies* (Baltimore and London: Johns Hopkins University Press for the World Bank, 1993); John B. Knight and Richard H. Sabot, ''Educational Policy and Labor Productivity: An Output Accounting Exercise,'' *Economic Journal* 385 (1987): 199–214; Robert E. Lucas, ''On the Mechanics of Economic Development,'' *Journal of Monetary Economics* 21 (1988): 3–42; George Psacharopoulos, ''Returns to Investment in Education: A Global Update,'' *World Development* 22 (September 1994): 1325–44; T. Paul Schultz, ''Education Investments and Returns,'' *Handbook of Development Economics,* ed. Hollis Chenery and T. N. Srinivasan (Amsterdam: North-Holland, 1988); Lawrence H. Summers, ''Investing in *All* the People,'' Quaid-i-Azam Lecture at the Eighth Annual General Meeting of the Pakistan Society of Development Economists, Islamabad, Pakistan, *Pakistan Development Review, Papers and Proceeding* 31, pt. 1 (Winter 1992): 376–93; United Nations Development Program, *Human Development Report 1990* (New York: UNDP, 1990); and World Bank, *World Development Report, 1990* (Oxford: Oxford University Press for the World Bank, 1990).

2. For older respondents, the estimated income coefficients may be biased away from zero, to the extent that current income is a function of the enhanced earning power conferred by higher levels of schooling.

3. See, e.g., Nancy Birdsall, ''Public Inputs and Child Schooling in Brazil,'' *Journal of Development Economics* 18, no. 1 (May–June 1985): 67–86; and Jere R. Behrman, Masako Ii, and David Murillo, ''How Family and Individual Characteristics Affect Schooling Demands in Urban Bolivia: Multiple Schooling Indicators, Unobserved Community Effects, Nonlinearities and Interactions'' (UDAPE/Grupo Social, La Paz, 1995, mimeographed).

4. These are the countries for which data are presented in World Bank, *World Development Report, 1993* (Oxford: Oxford University Press for the World Bank, 1993). A number of problems in making income comparisons across countries and over time given existing data are discussed in T. N. Srinivasan, ''Data Base for Development Analysis: An Overview,'' *Journal of Development Economics* 44 (June 1994): 3–26, and other articles in the same issues of that journal. These articles do not address explicitly the problems in comparing income distributions across countries, but the nature of these problems is similar, to the extent that the problems affect high- versus low-income households differentially across countries or over time.

5. These comments on what is known about rural poverty head counts and on recent changes in those head counts in Asian countries are based on Jere R. Behrman, ''Rural Poverty in Asia,'' *Asian Development Outlook, 1992* (Oxford: Oxford University Press for the Asian Development Bank, 1992).

6. Raghav Gaiha and Anil B. Deolalikar, ''Persistent, Expected and Innate Poverty: Estimates for Semi-Arid Rural South India, 1975–84'' *Cambridge Journal of Economics* 17 (December 1993): 409–21, e.g., present estimates from a 10-year panel for rural south India that the incidence of poverty is considerably less if transitory fluctuations across years are averaged out than if only 1 year of data is used.

7. This approach assumes that the relative contribution of parental assets to household income remains constant over time.

8. The gap measures the difference in average income for high- and low-income households as a percent of the high-income average. See table 3. For the 20–24-year age cohort, the calculation is $(36.2 - 17.7)/36.2$.

9. There are very few private schools in rural Pakistan, and none of our sample respondents attended private schools. The absence of private schools

may reflect policies that subsidize public, but not private, schools. See Jere R. Behrman, Shahrukh Khan, David Ross, and Richard Sabot, ''Low Schooling and Large Gender Gaps in Pakistan: Market Failure? Policy Failure?'' (Bryn Mawr College, Bryn Mawr, Pa., 1993, mimeographed).

10. Mark R. Rosenzweig and Kenneth J. Wolpin, ''Evaluating the Effects of Optimally Distributed Public Programs,'' *American Economic Review* 76 (June 1986): 470–87, model the problems with program evaluation if program choices are endogenous. Jere R. Behrman and Raaj Kumar Sah, ''What Role Does Equity Play in the International Distribution of Aid?'' in *Economic Structure and Performance,* ed., Moises Syrquin, Lance Taylor and Larry E. Westphal (New York: Academic Press, 1984); and Jere R. Behrman and Steven G. Craig, ''The Distribution of Public Services: An Exploration of Local Governmental Preferences,'' *American Economic Review* 77 (March 1987): 37–49, have modeled the allocation of such public services among different constituencies (recipients) as reflecting the implicit constrained maximization of a governmental social welfare function that reflects productivity versus equity trade-offs. The application of this framework is summarized in this paragraph for Brazil (from Jere R. Behrman and Nancy Birdsall, ''Implicit Equity-Productivity Tradeoffs in the Distribution of Public School Resources in Brazil,'' *European Economic Review* 32 [October 1988]: 1585–1601) and Mexico (from Alec Ian Gershberg and Til Schuermann, ''Education Finance in a Federal System: Changing Investment Patterns in Mexico'' [Cambridge, Mass.: National Bureau of Economic Research, 1994]). The cross-national estimates are in Schultz (n. 1 above).

11. N. Birdsall, C. Pinckney, and R. Sabot, ''Inequality, Savings and Growth,'' (Williams College, Williamstown, Mass., 1995, mimeographed). The other side of this statement is that, if schooling were viewed only as an investment, households were risk neutral, and capital markets were perfect, household income would not affect schooling demands.

12. Better information regarding the expected rates of return to schooling may be valuable for at least two reasons: first, even in a static world, better information may lower the uncertainty and thus increase investments if households are risk averse; second, in the context of rapidly changing markets and technology, as some claim is appropriate for rural Pakistan, better information may permit a better assessment of how the distribution of returns to schooling is changing. If this distribution is shifting up, as conjectured by Richard Sabot, ''Human Capital Accumulation in Post-Green Revolution Pakistan: A Progress Report,'' *The Pakistan Development Review* 31, pt. 1 (Winter 1992): 449–90 (and as seems consistent with some other studies of developing agriculture, see, e.g., Andrew Foster and Mark R. Rosenzweig, ''Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture,'' *Journal of Political Economy* 103 [December 1995]: 1176–1209), better information would imply greater schooling demands, even if there were no risk aversion.

13. The text is written as if households (parents) make the schooling investment decisions, which probably is plausible for the schooling levels being considered, given the nature of society in rural Pakistan. If children themselves made the marginal schooling decisions, however, this same association might hold because the children's preferences may be shaped considerably by those of their parents.

14. The ordered probit estimates control for the self-selection into schooling implied by the probit estimates of the probability of starting school.

15. Beyond directly influencing cognitive skills, school quality and the home learning environment may also increase the demand for schooling by increasing the expected benefits from schooling.

16. The predicted cognitive skills of an average individual will not, in gen-

eral, be the average cognitive skills of all individuals, unless cognitive skills are a linear function of all observed characteristics (and our estimates incorporate a number of nonlinearities).

17. In the only province not represented, Baluchistan, the rural population constitutes a small portion of its overall population.

18. Children in many of the remaining villages have a school available in a nearby village. However, there is no way to sort out the degree to which village characteristics may have influenced the location of a nearby school.

19. The actual distance was doubled if the entire distance was on unpaved roads and was increased by a factor of 1.5 if the trip was on mixture of paved and unpaved roads.

20. Estimates of real permanent household income at the time older respondents were of school age for Dir are significantly greater than those for the other regions. When the sample is limited to the 20–24-year age cohort, as we do below, there is no significant regional difference in permanent household incomes.

21. Schooling often starts at an older age in rural Pakistan than it does in high-income countries. Moreover, schooling progress is slowed down by grade repetition. Therefore, students tend to be older when they complete each schooling level.

22. Useable observations contain nonmissing values for all the variables used in the schooling-demand and cognitive-skills production functions.

23. If there are important liquidity constraints, current income at the time of the schooling investments might be relatively important, but this information is not included in the data set so we cannot explore this possibility. Other recent studies of rural areas in south Asia (including Pakistan) suggest that there are considerable possibilities for smoothing expenditures across households and over time, though with some limitations within agricultural production cycles (e.g., Harold Alderman, ‘‘Saving and Economic Shocks in Rural Pakistan,’’ *Journal of Development Economics* 51 [1996]: 343–66; Harold Alderman and Marito Garcia, *Poverty, Household Food Security, and Nutrition in Rural Pakistan,* Research Report 96 [Washington, D.C.: International Food Policy Research Institute, 1993]; Harold Alderman and Christine Paxson, ‘‘Do the Poor Insure? A Synthesis of the Literature on Risk and Consumption in Developing Countries,’’ Policy Research Working Paper 1008 [World Bank, Washington, D.C., 1992]; Jere R. Behrman, Andrew Foster and Mark R. Rosenzweig, ‘‘Dynamic Savings Decisions in Agricultural Environments with Incomplete Markets,’’ *Journal of Business and Economic Statistics* [in press]; Jere R. Behrman, Andrew Foster, and Mark R. Rosenzweig, ‘‘The Dynamics of Agricultural Production and the Calorie-Income Relationship: Evidence from Pakistan,’’ *Journal of Econometrics* 77 [1996]: 187–207; Mark R. Rosenzweig and Oded Stark, ‘‘Consumption Smoothing, Migration, and Marriage: Evidence from Rural India,’’ *Journal of Political Economy* 97 [August 1989]: 905–26; Mark R. Rosenzweig and Kenneth J. Wolpin, ‘‘Credit Market Constraints and the Accumulation of Durable Production Assets in Low-Income Countries: Investments in Bullocks,’’ *Journal of Political Economy* 101, no. 2 [April 1993]: 223–45; and Robert Townsend, ‘‘Risk and Insurance in Village India,’’ *Econometrica* 62 [May 1994]: 539–92).

24. We have investigated the substantial gender gap in cognitive skills in Harold Alderman, Jere R. Behrman, David Ross, and Richard Sabot, ‘‘Decomposing the Gender Gap in Cognitive Skills in a Poor Rural Economy,’’ *Journal of Human Resources* 31 (Winter 1995): 229–54.

25. We consider the substantial regional gaps in cognitive skills in Harold Alderman, Jere R. Behrman, Shahrukh Khan, David Ross, and Richard Sabot,

‘‘Decomposing the Regional Gap in Cognitive Skills in Rural Pakistan,’’ *Journal of Asian Economics* 7 (Spring 1996): 49–76.

26. We verified in villages in which we administered the school survey that the proxy was an accurate indicator of the year in which the school was first available.

27. As discussed in Subsec. IV.A, this appears to reflect regional differences in the distribution of schools rather than any direct effect of household income on the supply of schools.

28. Since tests were administered only to those with at least 4 years of schooling, scores had to be imputed for those with less schooling. Those with no education were assigned scores of zero. (The scores of a subsample of the uneducated who were given the tests confirmed the appropriateness of this assignment.) The means and standard deviations for cognitive achievement in table 5 include the scores for these individuals. Respondents with 1–3 years of schooling and qualified respondents who failed to take the test are kept in the sample only for the estimates of the schooling-attainment relations. Both tests were used successfully in research on human capital accumulation and the labor market in east and west Africa (e.g., Maurice Boissiere, John B. Knight, and Richard H. Sabot, ‘‘Earnings, Schooling, Ability and Cognitive Skills,’’ *American Economic Review* 75 [1985]: 1061–30; Jere R. Behrman and Victor Lavy, ‘‘Child Health and Schooling Achievement: Association, Causality, and Household Allocations,’’ World Bank Living Standards Measurement Study Paper no. 104 [Washington, D.C.: World Bank, 1994]; Paul Glewwe, ‘‘The Relevance of Standard Estimates of Rates of Return to Schooling for Education Policy: A Critical Assessment,’’ *Journal of Development Economics* 51 [December 1996]: 267–90; Paul Glewwe and Hanan Jacoby, ‘‘Student Achievement and Schooling Choice in Low Income Countries: Evidence from Ghana,’’ *Journal of Human Resources* 29 [Summer 1994]: 842–64; Paul Glewwe and Hanan Jacoby, ‘‘An Economic Analysis of Delayed Primary School Enrollment and Childhood Malnutrition in a Low Income Country,’’ *Review of Economics and Statistics* 77 [February 1995]: 156–69; and John B. Knight and Richard H. Sabot, *Education, Productivity, and Inequality: The East African Natural Experiment* [New York: Oxford University Press, 1990]). We assume that cognitive skills so measured, perhaps several years after the completion of schooling, reflect the cognitive skills at the time of termination of school. That is, there is neither subsequent further augmentation nor depreciation in cognitive skills. Our preliminary estimates indicate that time and experience subsequent to schooling do not affect cognitive achievement in our data.

29. $[(3.88 + 3.52) - (2.00 + 1.47)]/(3.88 + 3.52)$.

30. J. C. Raven, *Guide to the Coloured Progressive Matrices (Sets A, Ab, B)* (London: Lewis, 1956). An appendix in Knight and Sabot, *Education, Productivity, and Inequality: The East African Natural Experiment,* provides some examples from this test.

31. For some examples of economic studies, see the studies in n. 28. For examples in other literatures, see C. Nokes, S. M. Grantham-McGregor, A. W. Sawyer, E. S. Cooper, and D. A. P. Bundy, ‘‘Parasitic Helminth Infection and Cognitive Function in School Children,’’ *Proceeding of the Royal Society London B* 247 (February 22, 1992): 77–81, and the references therein.

32. For example, see Mohammed Aliuddin Khan, ‘‘Reports and Papers Prepared Under the Provisions of the HRD Component of the Pakistan Project: Comments and Queries’’ (Johns Hopkins University, Baltimore, 1993, mimeographed).

33. Because of the small number of women in the 30–44-year age cohort for whom a middle school was available, we find no variation in our distance proxy for this group.

34. We are not able to explore the determinants of middle schools, because so few villages in our samples have such schools. However, because middle schools tend to be concentrated in district centers, our initial conjecture is that household and village characteristics are less likely to affect the location of middle than of primary schools.

35. Other specifications, by use of the variables in table 4, were also explored (as well as still other variables, such as an indicator of the concentration of land holdings as suggested by Mark R. Rosenzweig and Robert E. Evenson, ‘‘Fertility, Schooling and the Economic Contribution of Children in Rural India,’’ *Econometrica* 45 [1977]: 1065–79), but they neither added to the explanatory power nor substantially altered the estimates of the relations that are presented here.

36. The Wald statistic for the four village characteristic variables becomes 6.3 for boys and 6.8 for girls.

37. The starting-school probits are used as controls for selectivity into schooling, and the estimation procedure controls for the truncation implied by the attainment of respondents still in school.

38. Standard errors are adjusted with White's correction for heteroskedasticity resulting from village cluster effects. Unless otherwise qualified, we adopt the convention of using the term ‘‘significantly’’ to refer to point estimates that are significantly nonzero at the standard 5% level.

39. We find no evidence of significant effects for maternal schooling. But this probably does not reveal anything about the possible importance of maternal schooling because the percentage of respondents whose mothers had at least primary schooling or more is very small—only 1.5% in the younger cohort and 0.5% in the older cohort (table 3).

40. The Wald statistic for the joint significance of β_2, γ_1, and γ_2 is 6.47 for math and 14.75 for reading. Behrman and Lavy (n. 28 above) similarly find for Ghana that household variables tend to affect reading and school variables tend to affect mathematics.

41. Primary schools were less available for low-income households in the 30–44-year age cohort.

42. The calculation is $[(51.5 - 44.1) + (16.8 - 13.8) + (12.8 - -3.3)]/49.4$.

43. The sum of the direct and the income-associated effects exceed 100% for the younger cohort because of simulated negative effect of school availability.

44. However, we qualify these estimates, as noted, because income may be capturing some income-associated household characteristics as well as direct effects in our cognitive achievement production-function estimates.