What Policy Guidance Does the Literature Provide on the Relationship between School Quality and Child Labor?

by

Kenneth A. Swinnerton
Division of Economic and Labor Research
Office of Trade and Labor Affairs
U.S. Department of Labor

July 2013

Contact: swinnerton.kenneth@dol.gov; (202) 693-4916.

I thank Sarah Donovan and participants in an ILAB workshop for helpful comments.

Satisfying a mandate assigned by the Trade and Development Act (TDA) of 2000, the United States Department of Labor reports annually on the worst forms of child labor in over 140 countries around the world. Since 2009, the reports have included recommendations for actions that countries might take to reduce child labor. The purpose of this paper is to use the existing research literature to assess whether and when a discussion of school quality and a policy emphasis on it might be part of the TDA reporting and its follow-on recommendations.

This assessment finds little clear guidance from available empirical work. Although it is large, the literature generally does not suggest a consensus view[1] on what defines (or measures) school quality. As a result, the literature lacks robust empirical regularities to suggest when and how to intervene on school quality so as to affect child labor and other variables that might be related to it (e.g., school enrollment). Theory is more conclusive, at least in suggesting when. It suggests that in circumstances when the root cause of child labor is *not* poverty, improving school quality[2] can have a positive impact toward its elimination. But theory also suggests that when there *is* poverty, an emphasis on school quality may *increase* child labor. The *precise* line defining poverty needs itself to be established empirically[3], so cautious guidance from theory is to emphasize school quality as a tool for eliminating child labor only in situations that are –*by judgment call*-- undoubtedly *well-enough* above or below the "poverty line," and where poor school quality is an established fact.

[1] As discussed within the final section of this paper, this may be because school quality is not a static concept across (or within) countries or over time. However, literature reviews typically try to identify common findings in the literature and a remarkable finding in reviewing the school quality literature is how few commonalities there are in either what is assessed or found.

[2] An advantage to doing theory, but a draw back to implementing its implications, is that "school quality" need not be more concretely defined to derive results.

[3] Since the poverty line for this exercise depends on estimating the empirical relationship between school quality and child labor, which has yet to be done satisfactorily, precisely identifying this poverty line is not possible.

After reviewing the state of empirical and theoretical work, this paper concludes with a discussion on how two current trends in development policymaking and research – recipient involvement in the design of policy interventions and randomized control trial evaluation – might point the way to eventually being more precise about when and how an emphasis on school quality is important to eliminating child labor.

I. Empirical Literature

In an authoritative and broader survey of the literature on child labor, Eric Edmonds writes that "[c]ausal evidence on a link between school quality and child labor that would meet modern standards of evidence does not appear in the literature." (2008, p. 3681) Moreover, the studies he reviews do not directly assess the relationship. Rather, the impact of school quality on child labor is inferred from the relationship between another pair of variables.[4] Direct searches of the EconLit database and google scholar revealed no papers that suggest a more direct assessment of the relationship between school quality and child labor.[5] Reviewing related literatures on the impact of school quality on schooling and on test scores provides insight into why little effort appears to have been devoted to examining the direct empirical connection of school quality to child labor.

In the literature on the impact of school quality on some measure of school participation (enrollment, attendance or hours devoted), one approach is for researchers to specify equations with every indicator they can measure, and that they think contains information about school quality, entered as an array of explanatory variables or aggregated into a school quality index.

[4] For example, he cites a study by Case and Yogo (1999) in which the relationship assessed is between school attendance and pupil/teacher ratios.

[5] Searches conducted January 25, 2011 (EconLit); November 3, 2011 (google scholar); and, February 3, 2012 (EconLit again). The search terms entered were "child labor" and "school quality." For EconLit, the final search was limited to the years 2000 to 2012, directed to examine abstracts of peer-reviewed papers for these terms, and asked to return results for papers with abstracts that contained both terms Ten results were returned. None presented a direct empirical connection between school quality and child labor. The google-scholar search returned no papers since Edmonds' review that appear to have examined the relationship.

Examples of the "array" approaches are Lloyd, Mensch and Clark (2000) and Lloyd, El Tawila, Clark and Mensch (2003). In the former, there are specifications with as many as 43 different indicators of school quality. The problems such an approach implies for measuring impact are explained later, the point now is just to note that the approach does not suggest a practically implementable definition of school quality for program or policy purposes. There are too many possible levers.[6]

More typically, researchers make some determination that limits their array to a smaller set of variables, but there is not a great degree of overlap in the determinations made across researchers. Glewwe, Hanushek, Humpage and Ravina (GHHR, 2011) systematically analyzed the literature in a way that makes possible a quick illustration of, and a high degree of confidence in, this assessment. They used two databases and a number of prominent working paper series to search for literature on school inputs (many of these are arguably "quality" inputs) and school outcomes. The databases were EconLit for the economic literature and the Education Resources Information Center (ERIC) for the education literature. They searched for papers from between 1990 and 2010 that were indexed with "education" and, singly, each of 72 different educational inputs as keywords. They next screened into their sample papers only that included the name of

[6] In both papers, the school participation variable is the odds of dropping out or exiting school. Odds are measured separately for boys and girls using the same empirical model within each paper. In the first (on Kenya), the "school quality" (related) variables that are significant are: higher school fees which *lower* the odds of dropout for girls; better credentialed teachers which *increase* the odds of dropout for girls; teachers with more in-service training *increase* the odds of dropout for boys; and more advisors for boys than girls, perceived gender discrimination by girls, and reported sexual harassment by girls, all increase the odds that girls dropout. In the second (on Egypt), a shorter-length school day raises girls' odd of exit; a high ratio of borrowed or temporary to full time regular teachers increase odds of exit for both boys and girls; in-service training for teachers lowers odds of girls' exit; regular meeting of a home economics class lowers odds of girls' exit; the presence of an adult confidant *increases* the odds of girls' exit; reporting that students are treated equally lowers the probability of *boys'* exit; and for exit during primary school, the presence of extra-curricular activities lowers girls' odds and being told by a teacher that they are a failure raised their odds (these results do not extend to higher grades). In addition to identifying a long list of possible "school quality" levers that appear to matter differently for boys and girls, note that these papers also yield some counter-intuitive results. This could be from a failure to recognize any number of econometric problems, or it could simply be that with such a large number of variables entered into the model's specification a few falsely significant results are bound to turn up.

at least one developing country or that contained the phrases "developing country" or "developing countries." At this point, their sample included around 9,000 papers.

The authors then examined abstracts for each of the articles, culling out the vast majority because they were not relevant to examining the relationship between school inputs and educational outcomes. The remaining 253 papers were read in their entirety, and culled further to 112 on the basis of relevance. Methodological quality screens were then applied, eliminating another 33 that did not use certain common and widely accepted econometric or statistical methods. The remaining 79 studies made up the sample they analyzed. Within that sample they identified 43 studies as "high quality" and further disaggregated out 13 that were based on random control trial (RCT) methods. Because an RCT is often viewed as the gold standard for assessing a causal relationship, these are presumably the "highest quality" studies. The majority (69) of the 79 studies that GHHR identified took student test scores as the outcome of interest. Eighteen examined "time in school variables" (e.g., attendance or enrollment) in addition or instead.

In their analysis, GHHR are interested in assessing which individual variables that measure school inputs (and in many cases aspects of school quality) are best supported by the literature as explanatory of school outcomes.[7] The primary focus here is bigger picture: is there really that much support for *any* variable? The other question of interest for this paper is whether there is consistency across individual contributions to the literature in the explanations

[7] GHHR provide no explicit criteria by which they decide a result is, e.g., "strong" or "very strong". They appear to use some combination of whether results are "what almost anyone would expect" or what "common sense suggests" and significance of the variable in more than 10 percent of cases as their screen. The latter is based on notion that if the underlying parameter is zero, the sample parameter should be insignificant in 90 percent of samples drawn, which is true when sampling from a *fixed* population (see GHHR p. 19, particularly footnote 12). Contestable in this assertion is the notion of a fixed population: why not view samples drawn from different countries and different time periods as draws from different populations?

that researchers seek. Tables 1 and 2 reorganize the results from GHHR to answer these questions.

Table 1 summarizes the results of the GHHR search of the literature for impacts of 12 school inputs on "time in school."[8] The input variables are listed in the first column. The second column gives the expected impact of the row variables on time in school.[9] Each entry in columns three through five gives a ratio. The numerator is the number of times the associated row variable was to found to have the expected and statistically significant effect[10] on time in school. The denominator is the number of empirical specifications in the GHHR sample that have considered the variable. It is common for a study to contain more than one specification, so the number of studies that consider a given row variable is fewer than the number of specifications. With that in mind, look at the column for "All 79 studies."

Scanning the denominator of each reported entry, notice that the most times any variable was included in a specification was seven (Textbooks/Workbooks, Teacher Experience and Pupil-Teacher Ratio). Now note that the number of variables included at all decreases with the increasing quality of methodology of specifications. For the high quality, two variables have no entry and for the RCTs, eight have no entry. Based on the heterogeneity of variables across these specifications, there does not appear to be a consensus set of variables that researchers think ought to be part of their empirical specifications.

Next notice that only eight (out of a possible 36) entries indicate that the row variable had the expected and statistically significant effect in half or more of the specifications studied. At

[8] GHHR (p. 16) report including a variable in their tables if it appeared in at least one RCT specification or at least two specifications using other methodologies.
[9] For most entries across Tables 1 and 2, these expectations come from the discussion in GHHR. In instances where an expectation cannot be gleaned from GHHR's discussion, it was assigned by the author using the same apparent criteria as GHHR, viz. intuition informed by a broader reading of the literature or common sense.
[10] GHHR define significance at the 10 per cent level or better.

100 percent "library" seems proportionally to be the most successful variable. But that is based on just two specifications. At 80 percent, the proportionally next largest rates among all 79 studies were for Building New Schools and School Quality Index. The first of these is arguably not a school quality measure, but a school existence measure. The second is clearly intended to be a school quality measure. But, it turns out to be based on just two papers. Also, it vanishes when the tighter methodological screens are applied to whether the specifications that include it remain in the review sample (note the missing entries in the last two columns of Table 1).

Now turn to Table 2. The studies under consideration attempt to measure the impact of school inputs, including some hypothesized measures of school quality, on student test scores. Again each entry in columns three through five gives a ratio, which is defined analogously to the ratio used in Table 1. Even though the set of input variables considered in Table 2 is much larger (at 34 versus the 12 in Table 1), unlike in Table 1, Table 2 shows that many researchers appear to agree that there are at least some common ones that should be included in empirical specifications. From the column for "All 79 studies" the top five variables in terms of the number of specifications in which they appear are pupil-teacher ratio (101 specifications), teacher education level (70), teacher experience (63), textbooks/workbooks (60), and female teachers (39). Remarkably, of these variables only pupil-teacher ratio crosses the threshold for being statistically significant in 50 percent or more of the specifications in which it was included.

More pupils per teacher shows a statistically significant negative effect on test scores in about 59 percent of the specifications assessed for Table 2, according with an intuition that smaller class sizes leads to better outcomes. But perhaps reflecting the fact that it is the most commonly used input variable and the fact that it sometimes (15 percent of the specifications reflected in Table 2) has a statistically significant *positive* effect, there have been caveats in the

literature about what it measures, and what to expect about its impact. For example, GHHR (p. 24) write about the possibility "that schools that are of high quality due to unobserved characteristics will attract more students, raising the pupil[-]teacher ratio and thus leading to a positive correlation between that ratio and student test scores." Alternatively, if pupil-teacher ratios reflect the needs of the population of students, then more needy[11] students, who may be expected to perform worse on tests, should be found in smaller classes (Lazear, 2001). This too would introduce a positive correlation between pupil-teacher ratios and test scores.

Among those variable represented less frequently in Table 2, there are some that have attained statistical significance more than half the time. In terms of percentage of specifications significant, the most successful of these variables appear to be tutoring, contract teacher, student attendance and hours of school. Each is significant 100 percent of the time in at least one of the three sets of studies.

Student attendance is significant in the largest number of specifications (8). These specifications are from two studies (see GHHR, Table 9) among the lowest quality studies that GHHR consider. In terms of saying much informative about the effect of school quality, these results can also be discounted on the grounds that student attendance is a student input measure, rather than a school input or quality measure. Even making the argument that it proxies for school quality –say, because students are more likely to show up at higher quality schools – does not provide any real information, because there is no way of knowing what quality aspect it is that attracts the attending students. Similarly, hours of school is arguably not a quality measure. Rather, it measures intensity.

"Contract teachers" is significant in all four of the RCT specifications in which it appears. Contract teachers are employed on short-term renewable contracts. GHHR (p. 26)

[11] Where the need could be for, say, individualized instructional attention or to discipline disruptive behavior.

highlight the argument that hiring teachers on such contracts provides them with better incentives than tenured teachers to perform well because they want to have their contracts renewed. But the four specifications are from just one study (see GHHR, Table 11), which may just say something about incentive systems in the context of the one country studied. Indeed, the fact that the other two specifications, and the one other study (see GHHR, Table 9) that includes this variable show it to be insignificant, tend further to suggest that concerns about a country fixed effect should not be easily dismissed. Similarly, the 100 percent of the two RCT specifications that find tutoring to be significant are from one study (see GHHR, Table 11), and tutoring also does not fare so well using other methods. In short, it seems premature to say *anything* about the external validity or robustness of either contract teachers or tutoring as explanatory for student test scores.

Among the other input variables surpassing the threshold of significant in at least fifty percent of specifications examined, grounds for being cautious about the general nature of the results include, that the variable appears in no higher quality specification (school infrastructure index, teacher assigns homework, expenditure per pupil), or that the results are based on six or fewer specifications and even fewer papers (library, teacher job training, teacher absenteeism, school provides meals, merit-based scholarships). The most successful of these variables in terms of the significance threshold and having been examined in a fair number of specifications is teacher knowledge as determined by testing. It is significant in about 55 percent of all specifications and 65 percent of the high quality specifications in which it appears. Moreover, there are 33 total and 20 high quality specifications considered. Still, these specifications come from a small number of papers, 9 and 5 respectively (see GHHR, Tables 8 and 10).[12]

[12] Interestingly, none of these papers make use random control trial methods.

The very big picture summary of these empirical results is that there are few (if any) robust regularities, and in many cases that is because there are not enough results to judge. But, rather than immediately suggesting the production of "more results," it is helpful to focus a bit on methodological points as one source of constraint on the ability to establish robust findings. Start with multivariable econometric approaches. With these, the main constraint generates from the fact that researchers in this area do not appear to have a common sense of the input variables that they should examine. This makes "omitted variable bias" a valid criticism of any somewhat parsimonious specification. That is, any significant result found may not be due to a measured variable itself, but instead may be due to an omitted variable with which the included variable is correlated and for which the included variable, in essence, serves as a proxy.[13] Indeed, the GHHR caveat about the pupil-teacher ratio describes one possible example of omitted variable bias. To avoid this, it may be tempting to try to include all variables that have appeared in the literature. But remember that GHHR identified 72 that may be relevant. In general, the more variables in a specification, the less precise are estimates of the effects of individual variables. If the sample is not sufficiently large, then trying to tease out the effects of many variables in too small a sample can be the source of statistical insignificance rather than true economic insignificance of a variable *per se.*[14]

Random control trial (RCT) methods show better promise of producing more usable results than traditional econometric methods. In principle, the randomization of the sample into treatment and control groups means that there is no variation between the groups on average for

[13] For example, suppose that a specification contains teacher training but not teacher tested knowledge, but that the two are sufficiently highly correlated. Then teacher training may pick up an effect that actually attributes to teacher knowledge. Emphasis may then be place upon training methods whereas the appropriate emphasis would be on the knowledge teachers accumulate.

[14] See Orazem and King (2008, section 6) and GHHR, pp. 9-11 for more detailed discussions of potential problems using multivariate econometric techniques.

variables not part of the experiment. This deals with omitted variable bias issues and other concerns about identifying causality. However, there are three caveats. The first is that randomization may be difficult to do perfectly and there may still be need for traditional econometric tools to implement controls *ex post*. More important is the fact that RCTs only allow variation of one or a very few variables within the context of a study. If each school quality variable that has been proposed in the literature is to be analyzed according to this method, and if done in many country contexts to establish more general results, it will take a great deal of time and money to generate enough additional results to establish empirical regularities. Finally, because it is not possible to rule out that school quality is a different concept in different countries, empirical regularity spanning different countries may never be found.

II. Theoretical Literature

Given that the empirical research literature provides little general policy guidance on the effect of school quality on child labor, reviewing theoretical results may be of more interest than usual to policymakers. It turns out that theory research is no more helpful on what the specific levers to ensure school quality are. It does however, say when –it may not be always-- a policy focus on school quality may be helpful to the elimination of child labor, and how helpful.

An understanding of the relationship between child labor and school quality is most easily grasped when the theoretical child labor literature is understood as a part of the literature on "consumption smoothing." It is well known that the variation over time in household consumption patterns is less than the variation in their income streams. Consumption smoothing is the actions households take to allocate their income across time so as to limit the variation in consumption.

The basic behavioral principle behind the desire to consumption smooth is the "law of diminishing returns", which in this case suggests that each additional unit of a good that household may consume at a point in time is valued less by the household than the unit it just consumed. Because of this law, households (of fixed composition and with preferences that are constant across time[15]) are happiest consuming around the same amount in every period. If not, they could make themselves happier by cutting back in periods of high consumption to allocate more to low consumption periods.

There are three ways that households may receive and allocate their income so as to achieve their desired degree of consumption smoothing:

- They may have their entire stock of wealth at the time they are planning their consumption path and draw down that wealth at appropriate times to finance that path; essentially, this means that they can save;
- They can borrow against future income if there would be "too much" consumption in the future and "too little" at present, where too much and too little are defined with reference to the desired consumption path for their lifetime income or wealth; and/or
- Their income at all points in time is always enough to finance a smooth consumption path.

These will be referred to as "wealth scenarios." These wealth scenarios can be thought of as describing a household that lives forever, or as linking generations together financially by allowing for parent households to bequeath financial inheritances to their adult children's households. Note that in the latter conceptualization, both bequests and schooling are gifts that parents make to their children's future.

[15] Such strong assumptions are needed to ensure a steady state stream of actual consumption. If these assumptions are relaxed, consumption smoothing is achieved by making "horizontal" the stream of "marginal utility" enjoyed across time.

In households where child labor is observed, if any of the wealth scenarios listed above is in place, then increased school quality leads to more schooling and lower child labor. Assume that a child's time is allocated between just two activities, work or school. Work generates income now, school in the future. But if the scenarios hold then it does not matter when income is earned in determining when it is consumed. Saving, borrowing, or a reduction in planned bequests can be used to smooth consumption as desired across time. When an increase in school quality raises future earnings, parents reallocate their children's time from work to school and make up for the lost current income by adjusting as needed savings, borrowing, or bequests. Taking these actions raise household lifetime wealth and consumption.

Now consider a situation where a parent household can not borrow, does not save, and does not plan to make bequests, perhaps because the household exists just at subsistence. It is still the case that an increase in school quality raises the household's potential lifetime wealth and that this provides an incentive to allocate children's time to schooling. But the imperative for consumption smoothing creates a countervailing incentive. This is because increased consumption in the future that follows from more schooling must come at the expense of less consumption today because the child works less. In order to restore the balance sought in consumption smoothing, there is an incentive not to increase schooling and perhaps even decrease it so as to have the child work more to raise current consumption. In general, it is not possible to predict whether the potential wealth effect or the consumption smoothing effect will dominate, so the overall predicted impact of increased school quality on time devoted to child labor is ambiguous.[16]

[16] Even if child labor in a household goes up, household welfare increases in this class of models because their lifetime income has increased and they can allocate it optimally given their constraints.

Formal theoretical models demonstrating these results are found in at least five publications in the literature. Glewwe (1999, 2002) allows each unit of schooling to raise the immediate income a child can earn. In this setting, he shows that an increase in school quality has an ambiguous effect on child labor because there is some incentive to quit school to begin earning the returns from better quality school at an earlier age. Jafarey and Lahiri (2005) use their insight about the effects of school quality on consumption smoothing when borrowing and saving are difficult to suggest that policy focused on supplementing current consumption (specifically, a food for education program) will have a bigger effect on encouraging schooling and reducing child labor. Donovan and Swinnerton (2010) model how the general effect of raising the returns to education in the market for adult labor affect child labor, showing that when there are constraints on borrowing and saving, the impact on child labor is again ambiguous. Modeling constraints on borrowing for poor households by assuming that the interest rate for loans decreases with household income, Orazem and King (2008) show that any factor that exogenously raises the per-unit-of-time-devoted productivity of schooling has an ambiguous effect on the total time devoted to schooling and child labor.[17]

III. Policy Guidance

This concluding section returns to the title question of the paper and discusses what the literature implies for an emphasis on school quality as a policy instrument for lowering child labor.

[17] This discussion of theory suggests another possible problem or explanation for the current state of the empirical work discussed earlier: a failure to find an impact of school quality may be due to the theoretical ambiguity, particularly if there are (borrowing, saving and bequest) constrained and also unconstrained households in the data sample. The average across households may not be different from zero. Empirical specifications that lack proper methods for addressing the possible existence of such constraints will not be able to identify school quality effects. Often the data do not make it possible to implement such methods. For a good and creative illustration of appropriate methods to address the constraints, see Edmonds (2006).

The theoretical literature provides the most general guidance. It establishes that a policy emphasis on school quality is likely to be most effective in situations where child labor exists but where households are not using child labor as a consumption smoothing strategy. In general, these are households that have other mechanisms for allocating their income or wealth across time. Practically speaking, these are likely to be households that either have a regular and relatively high (among households with child labor) flow of income or households that are able to exploit mechanisms to save. Wealthier households and households with mechanisms to save are most likely to be found in wealthier countries. So where there is child labor and an issue of poor school quality in a relatively well off country, an emphasis on improving school quality may have an important impact on reducing child labor. In a poorer country, emphasizing school quality is likely to have a smaller and possibly opposite impact on child labor. In such a country, addressing the causes of poverty and constraints on consumption smoothing is more of a priority both in and of itself and in terms of sequencing. That is, issues of pervasive poverty likely should be addressed before an emphasis is placed on school quality, not concurrently.

Lacking in the policy guidance from theory is precision: what constitutes "a relatively well off country" and separates it from a poor one for the purposes of determining school quality effects, precisely how big the effect on child labor would be, and what precisely is the lever for addressing school quality. These are empirical questions. But the current empirical literature is incapable of answering them with any degree of generality. As earlier discussion in this paper showed, the overriding explanation for this is a lack of general understanding of what constitutes good measures of, or levers for affecting, school quality. Additionally, theory suggests that there may be no unambiguous general expectation about the direction of impact. It may be the case that there is no general empirical result to be discovered.

School quality may be a highly contextual factor varying from country to country or even within a country. If this is the case, then two current trends may be the key to making progress in determining where and when an emphasis on school quality may affect child labor, and indeed other outcomes for children. Many current approaches to development programming and policymaking emphasize inclusion of targets of the program or policy in the design of interventions.[18] Following this approach, the potential beneficiaries themselves would identify the lever or levers in the intervention. Assessing the impact of these levers (as long as there are sufficiently few for the data that can be collected), could be the mission of well-designed RCT studies. This marriage of two trends – grass roots involvement in policy and program design and RCT research methodologies – may only provide information that may be valid in specific contexts, but at least it provides better guidance for what to do on a larger scale *in those contexts* than may ever be obtained if the mission of research is to try to identify results that generalize on a global or substantially global scale. And even if there are general empirical results to be found, using the beneficiaries of school quality to identify interventions and RCTs to test them is likely a more productive way to search for those results.

[18] See de Silva (forthcoming) for a discussion of this issue as part of a "sustainable livelihoods approach" to development and the White House (2010) for inclusion in the theme, at least at the country level, in United States global development policy.

References

Case, Anne and Motohiro Yogo (1999) "Does School Quality Matter? Returns To Education And The Characteristics Of Schools In South Africa," National Bureau of Economic Research Working Paper 7399, http://www.nber.org/papers/w7399.

de Silva, Theresa "Principles in International Development: Sustainable Livelihoods and Human Rights Based Approaches," ILAB Economic and Labor Research Paper, forthcoming.

Donovan, Sarah and Kenneth A. Swinnerton (2010) "Adult Returns to Schooling and Children's School Enrollment: Theory and Evidence from South Africa," in Randall K.Q Akee, Eric V. Edmonds and Konstantinos Tatsiramos (eds.) *Research in Labor Economics: Child Labor and the Transition Between School an Work* (31) pp. 297-319.

Edmonds, Eric V. (2006) "Child Labor and Schooling Responses to Anticipated Income in South Africa, *Journal of Development Economics,* 81, pp. 386-414.

Edmonds, Eric V. (2008) "Child Labor" in T. Paul Schultz and John Strauss (eds.) *Handbook of Labor Economics* (Amsterdam: Elsevier B.V.) pp. 3607-3709.

Glewwe, Paul (1999) *The Economics of School Quality Investments in Developing Countries: an Empirical Study of Ghana,* (New York: St. Martin's Press).

Gleewwe, Paul (2002) "Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes," *Journal of Economic Literature*, XL, June, pp. 436-482.

Glewwe, Paul W., Eric A. Hanushek, Sarah D. Humpage and Renato Ravina (2011) "School Resources and Education Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010," National Bureau of Economic Research Working Paper 17554, http://www.nber.org/papers/w17554.

Jafarey, Saqib and Sajal Lahiri (2005) "Food for Education versus School Quality: A Comparison of Policy Options to Reduce Child Labor," *Canadian Journal of Economics*, 38(2), pp. 394-419.

Lazear, Edward (2001) "Education Production," *Quarterly Journal of Economics* 116(3), pp. 777-803.

Lloyd, Cynthia B., Sahar El Tawila, Wesley H. Clark. and Barbara S. Mensch (2003) "The Impact of Educational Quality on School Exit in Egypt," *Comparative Education Review*, 47(4), pp. 444-467.

Lloyd, Cynthia B., Barbara S Mensch, and Wesley H. Clark (2000) "The Effects of Primary School Quality on School Dropout among Kenyan Girls and Boys," *Comparative Education Review*, 44(2), pp. 113-147.

Orazem, Peter F. and Elizabeth King (2008) "Schooling in Developing Countries: The Roles of Supply, Demand and Government Policy" in T. Paul Schultz and John Strauss (eds.) *Handbook of Labor Economics* (Amsterdam: Elsevier B.V.) pp. 3475-3559.

United States Department of Labor (annually since 2002) *Findings on the Worst Forms of Child Labor*, http://www.dol.gov/ilab/programs/ocft/tda.htm.

The White House (2010) *Fact Sheet: U.S. Global Development Policy*, http://www.whitehouse.gov/the-press-office/2010/09/22/fact-sheet-us-global-development-policy.

Table 1

Ratio of Empirical Specifications Finding an Expected and Significant Impact of the Row Variable on Time in School

	Expected Impact	All 79 Studies	43 "High Quality" Studies	13 RCT Studies
School Infrastructure Variables				
Textbooks/Workbooks	+	2/7	2/7	2/3*
Library	+	2/2*	a	a
Roof/Wall/Floor	+	1/2*	1/2*	a
Building New Schools	+	4/5*	4/5*	3/4*
School Quality Index	+	4/5*	a	a
Staff Characteristic Variables				
Teacher Education Level	+	1/5	1/5	a
Teacher Experience	+	2/7	1/6	a
Teacher Job Training	+	0/3	0/3	a
School Organization Variables				
Pupil-Teacher Ratio[c]	-|+	0/7 | 2/7[b]	0/7 | 2/7[b]	a
Cost of Attending	-	0/6	0/6	a
School Provides Meals	+	0/1	0/1	0/1
Merit-based Scholarships	+	0/3	0/3	0/2

Source: Tables 12-14 of Glewwe, Hanushek, Humpage and Ravina (2011). See text for a summary of how these authors identified and classified studies to consider. Denominator in each entry of this table is the total number of empirical specifications that considered the row variable. Some studies contained multiple specifications.

a = no specification considered this variable; b = first entry applies to negative and significant impacts, second to positive and significant; c = see text for a discussion of why the impact of this variable may be expected to be negative or positive; * = 50% or more of all specifications.

Table 2

Ratio of Empirical Specifications Finding an Expected and Significant Impact of the Row Variable on Test Scores

	Expected Impact	All 79 Studies	43 "High Quality" Studies	13 RCT Studies
School Infrastructure Variables				
Textbooks/Workbooks	+	26/60	3/21	0/4
Desks/Tables/Chairs	+	8/28	3/7	a
Computers/Electronic Games	+	7/28	4/22	4/20
Electricity	+	6/15	0/6	a
Blackboard/Flip Chart	+	7/25	2/6	0/1
Library	+	10/22	4/6*****	a
Roof/Wall/Floor	+	2/6	2/6	a
School Infrastructure Index	+	13/22*	a	a
Staff Characteristic Variables				
Teacher Education Level	+	24/70	2/13	a
Teacher Experience	+	17/63	5/28	a
Teacher Knowledge (tested)	+	18/33*	13/20*****	a
Female teachers	-|+	6/39 | 12/39[b]	1/8 | 1/8[b]	a
Teacher Job Training	+	11/29	3/6*****	a
Teacher Quality Index	+	6/14	a	a
Teaching Degree	+	2/6	a	a
Principal Education	+	1/6	a	a
Principal Experience	+	2/6	a	a
School Organization Variables				
Pupil-Teacher Ratio[c]	-|+	59/101*| 15/101[b]	14/46 | 3/46[b]	3/5* | 0/5[b]
Teacher Absenteeism	-	7/15	4/6*	a
Teacher Assigns Homework	+	12/16*	a	a
School Provides Meals	+	6/13	2/3*****	0/1
Multi-Grade Teaching	-|+	4/21 | 2/21[b]	4/10 | 1/10[b]	a
Hours of School	+	4/8*	4/4*****	a
Tutoring	+	2/5	2/4*	2/2*
Salaried Teacher	+	2/6	a	a
Contract Teacher	-|+	1/6 | 4/6*[b]	1/6 | 4/6*[b]	0/4 | 4/4*[b]
Expenditure/Pupil	-|+	2/3* |1/3[b]	a	a
Cost of Attending	-	1/6	a	a
Total School Enrolment	-|+	2/6 |1/6[b]	a	a
Group Work	+	4/13	a	a
Teacher Gives Examples	+	3/7	a	a
Student Attendance	+	8/8*	a	a
Community Information Campaign	+	1/14	1/14	1/14
Merit-based Scholarships	+	1/2*	1/2*	1/2*

Source: Tables 7-11 of Glewwe, Hanushek, Humpage and Ravina (2011). See text for a summary of how these authors identified and classified studies to consider. Denominator in each entry of this table is the total number of empirical specifications that considered the variable. Some studies contained multiple specifications.

a = no specification considered this variable; b = first entry applies to negative and significant impacts, second to positive and significant; c = see text for a discussion of why the impact of this variable may be expected to be negative or positive; * = 50% or more of all specifications.