

Assessing Designs for School Effectiveness Research and School Improvement in Developing Countries

ABBY RUBIN RIDDELL

The promise of an anticipated "third wave" of school effectiveness research in developing countries is in danger of being lost without ever having been explored.¹ Some of the first applications of multilevel analysis applied to questions of school effectiveness occurred in the late 1980s. Since that time, a trickle of research has utilized the new methodology, but rarely in a manner to illustrate its explanatory richness.² Researchers

¹ A. R. Riddell, "An Alternative Approach to the Study of School Effectiveness in Third World Countries," *Comparative Education Review* 33, no. 4 (November 1989): 481–97.

² S. Bashir, "Public vs. Private in Primary Education: Comparisons of School Effectiveness in Tamil Nadu" (Ph.D. diss., University of London, March 1994); Egypt Ministry of Education, Education Planning and Information Division, Research and Analysis Directorate, "General Report on National Survey of Teaching Practices and School Effectiveness" (Egypt Ministry of Education, Cairo, 1993); B. Fuller, H. Hua, and C. W. Snyder, Jr., "When Girls Learn More than Boys: The Influence of Time in School and Pedagogy in Botswana," *Comparative Education Review* 38, no. 3 (1994): 347–76; M. E. Lockheed and B. Bruns, "School Effects on Achievement in Secondary Mathematics and Portuguese in Brazil," Policy, Research, and External Affairs Working Papers WPS 525 (World Bank, Population and Human Resources Department, Washington, D.C., 1990); M. E. Lockheed and N. T. Longford, "A Multi-level Model of School Effectiveness in a Developing Country," Policy, Planning, and Research Working Papers WPS 242 (World Bank, Population and Human Resources Department, Washington, D.C., 1989); M. E. Lockheed and Q. Zhao, "The Empty Opportunity: Local Control of Secondary Schools and Student Achievement in the Philippines," *International Journal of Educational Development* 13, no. 1 (1993): 45–63; A. Loera and N. F. McGinn, "La repitencia de grado en la escuela primaria Colombiana: Resultados de una exploración sobre los factores asociados a la repitencia y las políticas de promoción," Education Development discussion paper (Harvard Institute for International Development, Cambridge, Mass., 1992); Namibia Ministry of Education and Culture, Florida State University, and Harvard University, *How Much Do Namibia's Children Learn in School? Findings from the 1992 National Learner Baseline Assessment* (Windhoek: Namibia Ministry of Education and Culture, 1994); L. M. Nyagura and A. R. Riddell, "Primary School Achievement in English and Mathematics in Zimbabwe: A Multi-level Analysis," Policy, Planning, and Research Working Papers WPS 1208 (World Bank, Population and Human Resources Department, Washington, D.C., 1993); S. W. Raudenbush, S. Kidchanapanish, and S. J. Kang, "The Effects of Preprimary Access and Quality on Educational Achievement in Thailand," *Comparative Education Review* 35, no. 2 (1991): 255–73; S. W. Raudenbush and C. Bhumirat, "The Distribution of Resources for Primary Education and Its Consequences for Educational Achievement in Thailand," *International Journal of Educational Research* 17, no. 2 (1992): 143–77; F. Reimers, "Influences on Student Achievement in Pakistan," and "The Role of Multigrade Education in Honduras" (papers presented at the Basic Research and Implementation in Developing Education Systems/Improving the Efficiency of Educational Systems [BRIDGES/IEES] Conference on School Effectiveness, Harvard Institute for International Development, Cambridge, Mass., September 1992); Riddell, "An Alternative Approach to the Study of School Effectiveness," and "School Effectiveness in Secondary Education in Zimbabwe: A Multilevel Analysis" (Ph.D. diss., University of London Institute of Education, 1988); A. R. Riddell and L. M. Nyagura, "What Causes Differences in Achievement in Zimbabwe's Secondary Schools?" Policy, Planning, and Research Working Papers WPS 705 (World Bank, Population and Human

Comparative Education Review, vol. 41, no. 2.

© 1997 by the Comparative and International Education Society. All rights reserved.
0010-4086/97/4102-0004\$02.00

in industrialized countries have focused on these new methods, challenged by the extended honeymoon of the difficult marriage between the formerly separate fields of school effectiveness and school improvement.³

A combination of factors threatens the development of further quantitative school effectiveness research in the Third World, not least of which is the disciplinary defensiveness of scholars whose findings require reexamination due to the doubt cast on them by multilevel reanalyses. Critiques of research are widely published and yet do not penetrate the received wisdom.⁴

School effectiveness research in the Third World may have reached a turning point. Judging by several recent reviews suggesting new directions,⁵ such research could die prematurely. If its early adolescence was spent copying its elders (i.e., the production-function models developed in industrialized countries), its late teenage years were relatively unproductive, failing even to replicate the more process-oriented research in the West. Now, just when it is increasingly possible to combine the contextualization of findings in more complex and realistic models, some researchers are dismissing such developments without a clear understanding

Resources Department, Washington, D.C., 1991); A. R. Riddell, "The Evidence on Public/Private Educational Trade-offs in Developing Countries," *International Journal of Educational Development* 13, no. 4 (1993): 373–86, and "Toward Modelling the National Study of Instructional Quality from a Multilevel Perspective" (Report to the National Center for Educational Research and Development, Amman, 1994); S. Singh and R. R. Saxena, "Achievement Difference and School Effects," *Indian Educational Review*, Special Number 1995 on District Primary Education Programme (New Delhi: National Council of Educational Research and Training, July 1995); D. P. Warwick and H. Jatoti, "Teacher Gender and Student Achievement in Pakistan," *Comparative Education Review* 38, no. 3 (1994): 377–99.

³ That is, the first International Congress of School Effectiveness and Improvement held in London in January 1988 and the inauguration of the journal *School Effectiveness and School Improvement* in 1990.

⁴ For example, despite the fact that serious doubt has been cast on the supposed differentiation of Third World school effectiveness (Riddell, "An Alternative Approach"), e.g., that "school-related factors are more important than out-of-school factors in explaining differences in student achievement" (J. P. Farrell, "International Lessons for School Effectiveness: The View from the Developing World," in *Teachers in Developing Countries: Improving Effectiveness and Managing Costs*, ed. J. P. Farrell and J. B. Oliveira [Washington, D.C.: World Bank, Economic Development Institute, 1993], p. 29), such "gospel" continues to be pronounced, as above, most often unquestioningly; see B. Fuller and P. Clarke, "Raising School Effects While Ignoring Culture? Local Conditions and the Influence of Classroom Tools, Rules, and Pedagogy," *Review of Educational Research* 64, no. 1 (1994): 122; H. Black, R. Govinda, F. Kiragu, and M. Devine, *School Improvement in the Developing World: An Evaluation of the Aga Khan Foundation Programme* (Glasgow: Scottish Council for Research in Education, 1993), p. 10, citing B. Fuller, "What School Factors Raise Achievement in the Third World?" *Review of Educational Research* 57, no. 3 (1987): 255–92; D. Reynolds, B. P. M. Creemers, J. Bird, S. Farrell, and F. Swint, "School Effectiveness—The Need for an International Perspective," in *Advances in School Effectiveness Research and Practice*, ed. D. Reynolds, B. P. M. Creemers, P. S. Nesselrodt, E. C. Schaffer, S. Stringfield, and C. Teddlie (Oxford: Pergamon), pp. 221–22.

⁵ See Fuller and Clarke; J. Jansen, "Effective Schools?" *Comparative Education* 31, no. 2 (1995): 181–200; and Reynolds, Creemers, Nesselrodt, et al., eds., for continuation, then see P. Dalin, *How Schools Improve: An International Report* (London: Cassell, 1994), for school improvement; and then J. Schwille, A. Beefu, R. Navarro, R. Prouty, S. Raudenbush, W. Schmidt, M. Tsang, and C. Wheeler, "Recognizing, Fostering, and Modeling the Effectiveness of Schools as Organizations in Third World Countries" (Project BRIDGES/IEES, Cambridge, Mass., 1986).

of what is being discarded. Instead, school effectiveness research appears to be seeking its lost youth and a return to the "second wave": class observational case studies characteristic of American and British studies during their heyday but absent in Third World studies.

This article takes stock of research into school effectiveness in developing countries, focusing on five issues: (1) Why has Third World research not used the analytic techniques applied in industrialized countries? (2) How do school effectiveness and school improvement approaches differ? (3) What indicators should researchers be using? (4) How has multilevel analysis been used in Third World research? and (5) How can we combine school effectiveness and school improvement concerns in a single research design?⁶

The End of School Effectiveness or the Consummation of Its Marriage with School Improvement?

The impetus for examining why Third World research has not used the analytic techniques applied in industrialized countries and the development of separate school effectiveness and improvement approaches comes from two recent reviews.⁷ Bruce Fuller and Prema Clarke update and expand one of the most widely cited reviews of Third World school effectiveness research,⁸ elaborating a dichotomy between "policy mechanics" and "classroom culturalists," one similar to Jonathan Jansen's distinction between school effectiveness and school quality research.⁹

Fuller and Clarke

Fuller and Clarke argue for a union of the "classroom culturalists" and the "policy mechanics." The former are a small group of researchers from a strong developed-country research tradition focused on the "culturally constructed meanings attached to instructional tools and pedagogy" in Third World schools, while the latter are the old production-function-based, universal solution-hunters. Out of their linkage some new breed of school effectiveness researcher is to be born. Fuller and Clarke argue convincingly for a culturally situated model of school effectiveness but balk at defining its new methodology.¹⁰ This is in part because they seek to incorporate criticisms of an earlier draft of their paper, as well as earlier work, but fail to see the contradictions in their response.¹¹ Ironi-

⁶ I am indebted to an anonymous reviewer of this article for excellent points concerning the reorganization of my arguments.

⁷ Fuller and Clarke; Jansen.

⁸ Fuller, "What School Factors Raise Achievement."

⁹ Jansen.

¹⁰ Fuller and Clarke, pp. 119-20.

¹¹ See the critique of B. Fuller, "Raising School Quality in Developing Countries: What Investments Boost Learning?" World Bank Discussion Paper no. 2 (World Bank, Washington, D.C., 1986), in Schuille et al.

cally, given their primary message, that the cultural meaning of classroom events conditions learning, it is surprising that they still accept as "proven" the decontextualized research findings they report.¹² They end up as the "reluctant butterfly": wanting to emerge, yet "safe" in the cocoon of donor-supported research.

What are the "givens" that Fuller and Clarke derive from school effectiveness research generated by policy mechanics? First, that school effects on achievement at times are greater than family background within impoverished settings.¹³ Although toned down from earlier pronouncements, this "gospel" is no more proven by the empirical evidence than it had been before.¹⁴

Second, they update a widely cited and harshly criticized review of school effectiveness research without addressing the fundamental criticisms.¹⁵ They list many of the reasons for invalidating the meta-analysis they report but then produce a table citing the number of significant effects for 30 variables constituting school inputs, teacher attributes, and pedagogical practices. They refer to the ability of multilevel models to address some statistical problems inherent in school effectiveness research, but they do not follow this through by dissecting the results they report. They warn that the use of culturally insensitive indicators of socioeconomic status can lead to the overestimation of the aggregate effect of schooling vis-à-vis family background,¹⁶ but they neglect the overestimation of the school effect due to fundamental model misspecification, rife in Third World school effectiveness research and, indeed, on which the dichotomy between industrialized and developing countries' school effects is drawn. The authors also mention the scarcity of longitudinal data, but nonetheless state that "we continue to rely on cross-sectional associations and limited [socioeconomic status] measures, rather than truly modeling how school factors influence growth in learning."¹⁷

Examining the weak underpinnings of a body of research without weeding out the more dubious findings does not strengthen the general-

¹² For instance, they state that "the majority of studies reveal significant effects within primary schools" of "the influence of teachers' own social class background" (Fuller and Clarke, p. 129). While sounding extremely reasonable, the basis for such a generalization is merely a tally: that in seven out of ten studies, this variable had a significant effect. We know nothing of the size or the validity of these "significant" effects because there is no discussion of the research designs used in these studies.

¹³ *Ibid.*, p. 122.

¹⁴ See Riddell, "An Alternative Approach" (n. 1 above), for full discussion. It is ironic that this gospel is contradicted by Fuller himself in Fuller et al. (n. 2 above), p. 367, in the context of a multilevel analysis of school effectiveness carried out in Botswana, in which only 12 percent of the variance in English and 16 percent of the variance in math scores can be attributed to between-school differences.

¹⁵ Schwille et al.; Fuller, "Raising School Quality," and "What School Factors Raise Achievement" (n. 4 above).

¹⁶ Fuller and Clarke (n. 4 above), p. 125.

¹⁷ *Ibid.*

izations, however “reasonable” they may seem, when there is insufficient skeleton to keep the body upright. Yet, this is what Fuller and Clarke ask of their readers, as they point to “consistent school effects” that include (1) the availability of textbooks and supplementary reading materials, (2) the general background and preservice training of teachers, and (3) instructional time and work demands placed on students. Although they state that these findings identify factors more consistently related to higher achievement,¹⁸ the research evidence is not so clear.¹⁹

The issue of clarity here, however, is problematic because the variables delineated are such obvious, intuitive candidates for policy attention, being easily manipulable resource inputs for policy making. Indeed, their obvious importance is fodder for those arguing *against* further quantitative analyses of school effectiveness, given that such studies confirm what is already known without necessarily increasing our understanding. To take textbooks as an example, Fuller and Clarke select it from their list of variables explaining achievement in developing countries because textbooks were found to be significant in 19 of 26 studies of primary schools. At the secondary level, however, they point to “less consistent effects” as textbooks were found to be significant in only seven out of 13 studies.²⁰ Although Fuller and Clarke raise the possibility of family background confounding the variance between schools, and thus lowering the effect of textbooks, this does not trouble them when accounting for achievement at primary schools—even though it is impossible to separate out the sources of overall variance in achievement and thus the real effect of textbooks in most studies reporting a significant positive effect. The vote tally method also present in Fuller’s 1987 review of research is used here,²¹ but precedence does not overcome the objections to this approach at meta-analysis.²²

Fuller and Clarke also accept uncritically the controversial “finding” from school effectiveness research in the Third World that private schools appear to outperform government schools in certain settings.²³ Yet the

¹⁸ *Ibid.*, p. 132.

¹⁹ Without reanalyzing the data contained in the individual studies, such as has been done to some of the more controversial school effectiveness research in industrialized countries—see the reanalysis of James S. Coleman, Thomas Hoffer, and Sally Kilgore, *High School Achievement: Public, Catholic, and Private Schools* (New York: Basic, 1982), in Stephen Raudenbush and Anthony S. Bryk, “A Hierarchical Model for Studying School Effects,” *Sociology of Education* 59 (1986): 1–17, or the reanalysis of S. N. Bennett with J. Jordan, G. Long, and B. Wade, *Teaching Styles and Pupil Progress* (London: Open Books, 1976), in M. Aitkin, S. N. Bennett, and Jane Hesketh, “Teaching Styles and Pupil Progress: A Reanalysis,” *British Journal of Educational Psychology* 51 (1981): 170–86—one cannot derive valid generalizations from this body of research, despite the pressures or, indeed, the research money incentives, applied especially by donors, for many of us to do so.

²⁰ Fuller and Clarke, p. 128.

²¹ Fuller, “What School Factors Raise Achievement.”

²² See Schwille et al. (n. 5 above).

²³ Fuller and Clarke, p. 141.

evidence for this conclusion is highly questionable.²⁴ Paying lip service to the methodological inadequacies of research but then ignoring those flaws when drawing conclusions based on such research does not set the stage for the more in-depth, cultural reflections on school organization that Fuller and Clarke suggest as the next step.

If Fuller and Clarke were not trying to approach a new paradigm of school effectiveness research, then perhaps I would not be so critical. Here it is important to distinguish the more questionable parts of their argument from the main thrust, with which I am extremely sympathetic. Thus, they are quite right that "simply accumulating more evidence from production-function studies—without specifying local conditions and without linking inputs to teaching practices—is becoming a less useful exercise."²⁵ They also are correct in stating that "a serious weakness of the production-function line is that instructional tools and even teaching practices are seen as cultureless, technical instruments for raising achievement."²⁶

However, their incomplete metamorphosis is evident in another conclusion they draw: "The production-function line of research may continue to provide useful policy guidance to the poorest nations and to impoverished areas of so-called middle-income countries."²⁷ This is because, in their view, "central policies hold a greater potential in actually touching the quality of instruction and the achievement of students. . . . Yet the influence of these policy-manipulable inputs will likely diminish as overall school quality rises."²⁸ What are we to make of this? On the one hand, we are told that we must situate our model of school effectiveness within its cultural parameters. On the other hand, interpreting how classroom teachers are viewed or what is being taught are deemed irrelevant in poor areas because the influence of trained teachers or having textbooks will be so great as to obviate such factors.²⁹

The basis for Fuller and Clarke's acceptance of defunct models of school effectiveness research for poorer countries is the too little questioned gospel that school effects are greater than family effects in such contexts. Yet the evidence does not substantiate this view. Table 1 shows that between-school differences are usually smaller than those between

²⁴ See Riddell, "The Evidence on Public/Private Educational Trade-offs," pp. 373–86, for a full discussion (n. 2 above).

²⁵ Fuller and Clarke (n. 4 above), p. 135.

²⁶ *Ibid.*, p. 139.

²⁷ *Ibid.*, p. 135.

²⁸ *Ibid.*, p. 134.

²⁹ To be fair, in correspondence, Fuller disputed the allegation I am making. However, it would seem that there remains a dichotomy, in Fuller and Clarke's view, between "school effectiveness" in "impoverished areas" and "school effectiveness" in more affluent areas, whether or not this is *between* countries.

TABLE 1
BETWEEN-SCHOOL VARIATION IN ACHIEVEMENT ACROSS COUNTRIES

Country (Study) Outcomes	% Variation between Schools	% Variation after Pretest Control	% Variation after Socioeconomic Status Control
Primary:			
Colombia (Loera and McGinn):			
Spanish 1	18
Spanish 3	29
Egypt (Egypt Ministry of Education):			
Math 5	60	...	59
Science 5	41	...	41
Arabic 5	53	...	51
Honduras (Reimers, "The Role of Multigrade Education"):			
Reading 1	33
India (Bashir):*			
Math 4	55	...	55
Reading comprehension 4	54	...	54
India (Singh and Saxena):†			
Math 4/5	20-61	...	20-54
Language 4/5	14-45	...	15-42
Jordan (Riddell, "Toward Modelling the National Study"):			
Arabic 5	27	23	23
Namibia (Namibia Ministry of Education and Culture et al.):			
English 7	66
Math 7	64
Pakistan (Warwick and Jatoi; Reimers, "Influences on Student Achievement"):‡			
Math 4	54	...	54
Science 4	47	...	47
Math 5	52	...	52
Science 5	50	...	49
Thailand:			
Math 3 (Raudenbush et al.)	31
Thai 3 (Raudenbush et al.)	35
Overall (Raudenbush and Bhumirat)	48
Zimbabwe (Riddell, "School Effectiveness," "An Alternative Approach"):			
English language 7	42
English literature 7	42
Math 7	42
Zimbabwe (Riddell and Nyagura):			
English 7	47
Math 7	60
Zimbabwe (Nyagura and Riddell):			
English 7	56	...	47
Math 7	74	...	36
Secondary:			
Botswana (Fuller et al.; Snyder 1994):			
English form II	12
Math form II	16
Brazil (Lockheed and Bruns):			
Math 3d year	62
Portuguese 3d year	36
Egypt (Egypt Ministry of Education):			
Math 8	42	...	40

SCHOOL EFFECTIVENESS AND SCHOOL IMPROVEMENT

TABLE 1 (Continued)

Country (Study) Outcomes	% Variation between Schools	% Variation after Pretest Control	% Variation after Socioeconomic Status Control
Science 8	35	...	32
Arabic 8	29	...	26
English 8	43	...	39
Philippines (Lockheed and Zhao):			
Math 8	52
Science 8	43
Thailand (Lockheed and Longford):			
Math 8	32	11	...
Zimbabwe (Riddell, "School Effectiveness," "An Alternative Approach"):			
English language form IV	42	27	21
English literature form IV	48	26	26
Math form IV	44	23	18
Zimbabwe (Riddell and Nyagura):			
English form II	65	47	...
Math form II	61	51	...

SOURCES.—See footnotes 1 and 2.

* A pretest was included in this study, but not curriculum related, of word knowledge. Separate results are not reported for the regressions on socioeconomic status and pretest results. "The between-school variances are not substantially reduced" (p. 136), although the pupil-level variances in both subjects were reduced by about 16% in both subjects.

† A range of results is reported because separate analyses were conducted in each of eight states. In math and language, the low variance component for the between-school variance is for Kerala, the high for math is Karnataka, and the high for language is Madhya Pradesh. The average between school variance in math for the eight states is 41%, the average for language, 34%.

‡ Results are those reported in Reimers, "Influences on Student Achievement in Pakistan." He acknowledges that "insufficient measures of background characteristics" were chosen. Hence the surprising lack of influence reported in the table.

students in accounting for overall variation in achievement. Essentially, the argument regarding the case of the Third World is *not* a reversal of the Coleman and Plowden reports' findings that the influence of the home is greater than that of the school.

Even if one denies the relative importance of between-student and between-school differences, the greater differentiation between schools in the Third World in general cannot be denied. This is the second point on which Fuller and Clarke base their suggested continuation of the policy mechanics' line of research. This greater differentiation is understandable given how quickly mass education systems have been constructed, certainly, relative to the longer trajectories taken by the already industrialized countries. Expansion under increasing resource constraints has produced stark differences in educational quality both across and within countries—a characteristic accelerated by the increasing introduction of user fees and the popularity of privatization as the state's role as provider of education diminishes.

Here the argument by Fuller and Clarke becomes tautological: with such vast differences in the quality (read resources) between schools in developing countries, we need not concern ourselves with the refined school effectiveness models used in industrialized nations because funding priorities can be based on equalizing school resource distribution. Limited resources necessitate a simpler research approach to determine which "inputs" have the greatest influence on student achievement. Indeed, this has been the basis of most school effectiveness research in the Third World in recent decades. But the evidence is not as clear-cut as Fuller and Clarke suggest, precisely because it has been so crude. On the one hand, they argue that we have the answers: textbooks, teachers, and time. On the other hand, their support for better understanding the local meanings of classroom situations goes against the policy mechanics' approach. Which is right?

Crude production-function analyses neither answer questions about school effectiveness nor shed light on the narrower issues of educational efficiency because the effects that purportedly are related to the identified "inputs" cannot be separated from the preexisting conditions of student populations. Therefore, the greater differentiation between schools in developing countries could be attributed to the differentiation in access to schools by different social groups as much as to the differentiation in school resources. In fact, private/public differentiation between schools has been shown to be confounded by peer group effects.³⁰ Further, to reanalyze those data sets that form the basis of such evidence would be a thankless task given the amount of flux in the educational systems in the Third World over the last 20 years; policy inferences from such outdated research would have little application today. It would be far better to convince donors to fund research embracing and unifying the two paradigms dissected by Fuller and Clarke in rich and poor countries alike. If Fuller and Clarke caution developing countries against introducing this new "classroom culturalist" paradigm, can it be surprising that others have avoided new techniques of the "policy mechanics"?

If the very sources of educational demand are being called into question by deterioration in educational quality, the increasing differentiation of schools within developing countries cries out for the contextualized research arising from appropriate melding of the two research traditions. Of course resources matter, and no doubt, they matter most where they are least. Yet does the evidence at hand enable us to evaluate the relative importance of, say, the three "T's" versus peer group effects in our competitive, selective educational environments? I think not.

³⁰ Riddell, "The Evidence on Public/Private Educational Trade-offs," and Bashir (n. 2 above).

Jansen and Others

Jansen puts the knife more deeply into the tradition of effective schools research, writing that it reached a cul-de-sac in the 1990s.³¹ He contends that studies of *effectiveness* and studies on *quality* represent competing and incompatible agendas for school and classroom-based research³² and distinguishes between approaches examining effectiveness and quality, maintaining that the latter is “concerned with [1] processes of teaching, learning, testing, managing and resourcing which must be [2] investigated on its own terms, i.e., through in-depth *qualitative investigations* of such processes, and [3] drawing more deliberately on *insider perspectives* of what happens inside schools and classrooms.”³³

Unlike Fuller and Clarke, who allow for the policy mechanics’ methods to be retained as they “broaden and culturally situate” their studies, Jansen dismisses quantitative approaches as incompatible with studies of educational quality—a curious stance given his comments on “the intensity with which the Anglo-American research on effective schools was replicated in the developing world.”³⁴ As I have argued elsewhere, the second wave of school effectiveness research, which emphasized process variables such as teaching style rather than physical inputs alone, was virtually overlooked in the Third World.³⁵ Indeed, Fuller and Clarke observe that North American educators and scholars have largely shifted to microschool factors and classroom-level processes in defining school effectiveness.³⁶

Curiously, Third World school effectiveness research is emerging as if from a time warp. Begun in the style of large-scale, quantitative research in industrialized countries, Third World studies continued in that tradition, but from a perspective that stressed efficiency rather than equity, taking as gospel that school mattered more than the home.³⁷ While research in industrialized countries focused on how schooling could compensate for the inequalities inherent in students’ backgrounds, the emphasis in developing countries was how to provide the best education for all with limited resources. Belatedly, researchers in developing countries are recognizing the need for contextualized, qualitative studies. The way this

³¹ Jansen (n. 5 above), p. 186.

³² *Ibid.*, p. 194.

³³ *Ibid.*, p. 195.

³⁴ *Ibid.*, p. 190.

³⁵ Riddell, “An Alternative Approach” (n. 1 above), pp. 483–84.

³⁶ Fuller and Clarke (n. 4 above), p. 134.

³⁷ C. Teddlie analyzes “the lack of contextually sensitive studies in school effectiveness research for most of the first 20 years of its existence [as] a testament to the power of the equity ideal that long dominated the field.” He sees the shift to school improvement as being a reflection of the abandonment of the equity ideal in favor of efficiency: improving schools for all (see C. Teddlie, “The Study of School Context in School Effects Research: History, Methods, Results, and Theoretical Implications,” in *Advances in School Effectiveness Research and Practice*, ed. Reynolds, Creemers, Nesselrodt, et al. (n. 4 above), pp. 87–88.

need is being expressed, often as a wasteful confrontation between qualitative and quantitative research, reflects some of the same underlying tensions that characterize the unconsummated marriage between school effectiveness and school improvement research in industrialized countries. The division between these approaches is not identical to the one Jansen delineates between school effectiveness and school quality in developing countries, however, because of the more persistent quantitative orientation of school effectiveness research in the Third World.

The friction between these very different research traditions in industrialized countries is relevant to Third World studies and is not unlike the confrontation Fuller and Clarke paint between the "classroom culturalists" and the "policy mechanics" or the division Jansen makes between those concerned with school effectiveness and those concerned with school quality. Understanding this division helps explain why new techniques such as multilevel analysis have not been well received in Third World countries and why research carried out in either tradition is put to different uses. Quantitative school effectiveness research in developing countries neither embraced the process variables prominent in the second wave of such research in industrialized nations nor the more qualitative contextualization that has characterized more recent school effectiveness research in industrialized countries. As a result, many researchers in the Third World—increasingly concerned with the quality, as opposed merely to the quantitative provision of education—have been put off by the crudeness and economic focus of quantitative school effectiveness research that continued, seemingly impervious to developments in the rest of the world. When more complex models embracing education's multilevel realities became amenable to analysis, such researchers already had been lost to the ethnographic classroom studies which similarly had been neglected earlier in most Third World countries. Jansen's article is testimony to this divided world of educational research. The opportunity of unifying the increasingly contextualized, quantitative research tradition with the newly embraced ethnographic tradition was not seized.

It is vital to break this schizophrenia of having a school effectiveness camp focused on the evaluation and measurement of progress across schools, feeding those involved in policy formulation at the macro level, and a school improvement faction concentrating on policy implementation and its effects at individual schools.

School effectiveness research developed from its antecedent "determinants of achievement" focus, seeking to explain differences in student achievement across schools. Until the convergence of new multilevel models and computers capable of coping with their underlying algorithms, school effectiveness research was hamstrung by having to analyze a nested, hierarchical social reality—of students nested in classes, schools,

neighborhoods, and so on—by models constrained to analyze such complex social realities within a single class, school, or district. Thus, while researchers understood that particular teachers would affect student achievement, analyzing a *school's* effectiveness meant dealing with the aggregate effect of all teachers rather than being able to disaggregate the differential effects of individual teachers as well as the effects of the overall characteristics of the teaching body—something a multilevel framework can do. Single-level models also forced researchers to ignore the nonrandom selection of students into schools in different neighborhoods, despite knowing that where a student lived would constitute a significant “determinant” of that student’s achievement and might well confound the effects of the school on subsequent achievement.³⁸ While researchers in industrialized countries have been using multilevel models to extend the areas of school-effectiveness inquiry and hone their measurements of school quality, such models have been little used in the Third World. The dependence of Third World research on donor funding and the fact that most studies—at least in the early days—were carried out by economists accounts for its economic focus and the distortion of school improvement into a school “quality” camp.³⁹

Indicators of Effectiveness and Improvement: Definitions and Case Studies

The friction between the school effectiveness and the school improvement camps is rooted not only in their different origins⁴⁰ but also in the type of indicators they choose for monitoring change. These differences have important implications, particularly for the application of research to policy making, given the widespread decentralization of educational governance now being experienced in developing and industrialized countries alike. Indicators that would be useful for teachers, school administrators, parents, and students are not always those deemed essential by district, regional, or central office personnel. Indeed, a lack of communication is likely between different levels of educational administration where either a school improvement or a school effectiveness faction dominates decisions concerning data collection and the production of indicators. The

³⁸ This is not the place for a long methodological treatise. Suffice it to say that economists who have been exposed to multilevel models see them as a minor subset of generalized least squares models that do not solve the overriding problems of selectivity bias that hinder such research. (I am indebted to an anonymous reviewer for this particular point, which clarified for me why it has been so difficult to interest economists in multilevel models of such intrinsic interest to educationists: they miss their substantive importance.)

³⁹ See K. King, *Aid and Education in the Developing World: The Role of the Donor Agencies in Educational Analysis* (Harlow, Essex: Longman, 1991); and P. Jones, *World Bank Financing of Education: Lending, Learning and Development* (London: Routledge, 1992).

⁴⁰ See, e.g., the useful analysis of these two “schools” in D. Reynolds, D. Hopkins, and L. Stoll, “Linking School Effectiveness Knowledge and School Improvement Practice: Towards a Synergy,” *School Effectiveness and School Improvement* 4, no. 1 (1993): 37–58.

macro systems orientation of school effectiveness researchers, many of whom come from backgrounds in educational administration or sociology of education that emphasize the effect of complex organizations and downplay teacher/student interactions, separates them from school improvement researchers focusing on individual schools. Many members of this latter group are from educational psychology, which stresses individual differences and teacher/student interactions while downplaying the role of organizations. Although some school effectiveness and school improvement researchers worry that an emphasis on teacher effects would lead to a focus on classroom rather than school reform, such turf guarding is unproductive since true change must occur simultaneously at both the school and class levels.⁴¹

Can these two perspectives be reconciled? Ideally, research would inform central, regional, or district policy makers about the effect of decisions made above the level of the school—for example, changes in resource allocation, or patterns of governance—while also ensuring good practices within schools in areas that may not be directly influenced by central policies. What is needed from each side is a shared recognition that not all answers will be found at either the level of the school or the level of the policy maker. In practice, however, research within each tradition has remained separate and generally has met at the level of neither data collection nor policy. An examination of indicators proposed by the two camps follows, using examples to illustrate the survival of earlier “schools” of school effectiveness, Third World school improvement approaches, and recent combinations.

*IIEP Primary School Quality Study on Zimbabwe*⁴²

An International Institute for Educational Planning (IIEP) Research Report describes indicators of educational quality developed in a national study of primary schools in Zimbabwe. As this study is used as a template for in-country training of research skills as well as a model for other countries,⁴³ it is a worthy example of a school effectiveness approach.

Twenty sixth-graders at each of the 150 primary schools in a stratified random sample were given a reading test. These schools—representative of the different Zimbabwean school types and with the probability of their selection proportional to the size of the grade 6 enrollment—were ranked

⁴¹ C. Teddlie and S. Stringfield, *Schools Make a Difference: Lessons Learned from a 10-Year Study of School Effects* (London: Teachers College Press, 1993), pp. 188–89.

⁴² K. N. Ross and T. N. Postlethwaite, *Indicators of the Quality of Education: A Summary of a National Study of Primary Schools in Zimbabwe*, Research Report no. 96 (Paris: International Institute for Educational Planning, 1992).

⁴³ See the Southern African Consortium for Monitoring Educational Quality (SACMEQ): D. Kulpoo, M. Nkamba, and T. Machingaidze, “SACMEQ—an Innovative and Exciting Concept,” *IIEP Newsletter* 13, no. 3 (July–September 1995): 4–6.

according to their average scores in achievement. They also were ranked according to the school's average socioeconomic level, determined by responses of the 20 sixth-graders to a questionnaire concerning seven possessions at home. School averages of the residual scores, after controlling for student socioeconomic level, then were computed to gain a rank order.⁴⁴ Differences in inputs between the most and least effective schools then were studied, as well as the correlations between inputs and achievement for the entire sample.⁴⁵ On this basis, implications for action by the Ministry of Education and Culture were drawn up, and visits to the most effective schools in the poorest areas were suggested. This sort of school effectiveness research, driven by variables manipulable by central policy, unlikely to move head teachers to much action, given that they probably are all too aware of their schools' resource deficiencies.

While this study may have been useful as a first stage in training MEC officials in survey techniques, it is fundamentally flawed as a template for school effectiveness research in the absence of indicators of the distribution of key resource-based inputs. Even acknowledging that time constraints may lead to a limited data set, basing policies on the derived indicators ignores research that has focused on producing sound measures of school effectiveness.⁴⁶ Controlling for students' socioeconomic status is a move in the right direction, but a multilevel analysis—even of the limited data at hand—would have been preferable to the single-level analysis performed, especially if the aim was to produce a ranking of school effectiveness.⁴⁷ The most important omission, however, is an intake achievement measure of students. Several studies of Zimbabwean schools have used a multilevel framework,⁴⁸ and these, like the wider research literature of which they are a part, point to several elements of good research design for school effectiveness studies: that data are measured at the level of the individual pupil, that controls are made for pupils' intake, and that the research design accounts for the inherent hierarchical organization of pupils in classes and in schools so that bias is not introduced in the measurement of school effects.⁴⁹ Without reanalyzing the

⁴⁴ Ross and Postlethwaite, p. 41.

⁴⁵ Data were collected on school buildings, teachers and their living conditions, and school and classroom provisions.

⁴⁶ My research in Zimbabwe was similarly constrained by time. Yet it was possible to produce a multilevel school effectiveness study in significantly less time than the 40-week period specified as necessary for studies of the IIEP variety.

⁴⁷ See H. Goldstein, "Better Ways to Compare Schools?" *Journal of Educational Statistics* 16, no. 2 (1991): 89–91, on the broad brush use of rankings, as opposed to a finer approach, given the inseparability of school effects given their relatively large standard errors.

⁴⁸ Riddell, "An Alternative Approach" (n. 1 above), and "School Effectiveness in Secondary Education in Zimbabwe"; Riddell and Nyagura; and Nyagura and Riddell (all in n. 2 above).

⁴⁹ See J. D. Willms, *Monitoring School Performance: A Guide for Educators* (London: Falmer, 1992), chap. 4; H. Goldstein, *Multilevel Statistical Models* (London: Edward Arnold, 1995); and Raudenbush

IIEP's data set using student-level intake data, it is impossible to determine how biased the results are.⁵⁰

The IIEP study belongs to a large family of Third World educational research that continues, despite the availability of practical alternatives, to be characterized by methodologies whose weaknesses have been highlighted repeatedly. The infrequent use by schools of the indicators derived from the study, as opposed to their use by higher level administrators, applies to multilevel as well as single-level school effectiveness research.⁵¹ The limited application of multilevel analysis to school effectiveness research in the Third World will be discussed later.

Willms's Type A and Type B School Effects

As noted, educational indicators serve different purposes for different people. Students or parents rarely want to know the same things about a school as teachers or administrators. Jon Douglas Willms details two types of school effect that together can only be generated from a multilevel analysis.⁵² These illustrate the uses to which any such indicators can be put.

The type A effect, or "adjusted school differences," which Willms suggests should be of special interest to parents, relates to school policies and practices, school composition, exogenous social and economic factors, and unmeasured effects associated with a given school.⁵³ This type A effect—essentially an intake-adjusted school ranking—measures variation between schools including, but without adjustment for, the effects of factors beyond the control of teachers and administrators.

and Bryk (n. 19 above), among many other more recent applications of a multilevel analytic approach in the literature. See Reynolds, Creemers, Nesselrodt, et al., eds. (n. 4 above), for the most recent review of such research.

⁵⁰ Judging by other reanalyses, however, it is more than likely that the effectiveness rankings both include and exclude schools that otherwise would be notable utilizing a multilevel framework. For instance, see P. Sammons, S. Thomas, P. Mortimore, C. Owen, and H. Pennell, *Assessing School Effectiveness: Developing Measures to Put School Performance in Context* (London: Institute of Education for the Office for Standards in Education, 1994), for a comparison of different models of school effectiveness. This study investigated the use of "nationally available data sources [in the United Kingdom] to create indicators of secondary schools' intakes to . . . enable comparisons of General Certificate of Secondary Education (GCSE) performance to be placed in better context" (p. i). "It was concluded that OLS [ordinary least squares] techniques can provide an *interim* method of putting secondary schools' performance in context. Nonetheless, a major limitation is the absence of confidence limits for individual schools. Using the OLS approach, it is not possible to identify for which schools actual GCSE performance differs significantly from that expected on the basis of their intake" (p. iii). "False positives and false negatives will be encountered using the less sophisticated OLS methodology and aggregated school-level information" (p. 38).

⁵¹ For instance, Jansen includes my multilevel work, legitimately, in the older tradition of school effectiveness research, bound by resource-driven modeling. Lack of donor funding, however, has meant that multilevel applications in the Third World have been extremely limited—both in number and, importantly, in design. The fact that richer, more complicated studies have not been funded should not be taken as the basis for discarding multilevel analysis. Indeed, Jansen himself ends with a plea for adequate donor finance of educational research. See King (n. 39 above) on the bias of donor-driven and supported educational research in developing countries.

⁵² Willms, chap. 4.

⁵³ *Ibid.*, p. 40.

Of interest to teachers and administrators is the type B effect, which is associated only with school policies and practices and unmeasured effects associated with a given school. Consequently, the estimates of the type B effect include statistical adjustment for the contextual influences arising from the school's social-class composition, and for effects arising from outside social and economic factors.⁵⁴

Thus, different policies, resourcing, and processes can be assessed in terms of both the average adjusted school effect (type A) and the particular school effect (type B) for pupils of different backgrounds. As Willms points out, "If comparisons of type A effects were used for accountability purposes, they would be unfair in that they would include factors beyond the control of teachers and administrators."⁵⁵

Peter Cuttance identifies three models of school effectiveness in the literature: the standards model, school-level intake adjusted model, and pupil-level intake adjusted model.⁵⁶ The last is the only one capable, in principle, of separating out the different effects that Willms details.⁵⁷ The standards model is simply a league table "comparing the average performance of pupils in a given school with the average performance of pupils across all schools."⁵⁸ The IIEP ranking of reading test scores is an example of a standards model.⁵⁹ Because it is impossible to separate the measure of a school's effectiveness from the intake characteristics of its pupils, change cannot be measured. In the IIEP study, if school achievement were regressed on the school-level socioeconomic data instead of simply comparing the rankings of schools by socioeconomic level with their reading test score rank, the residuals would constitute a school-level intake-adjusted measure. However, "school-level standards and intake-adjusted models of effectiveness may lead to contradictory conclusions about the effectiveness of individual schools."⁶⁰ They also "incorrectly [imply] that it is possible to summarize the effectiveness of schools for different types of pupils on a given outcome score on a single unidimensional index."⁶¹

*Namibian National Learner Baseline Assessment*⁶²

The Namibian National Learner Baseline Assessment goes much further than the IIEP Zimbabwe Primary Quality Study in producing new

⁵⁴ Ibid., pp. 40–41.

⁵⁵ Ibid., p. 146.

⁵⁶ Peter Cuttance, "Evaluating the Effectiveness of Schools," in *School Effectiveness: Research, Policy and Practice*, ed. D. Reynolds and P. Cuttance (London: Cassell, 1992), pp. 76 ff.

⁵⁷ This is the same point, essentially, as was made in the last section concerning Fuller and Clarke's suggested continuation of the policy mechanics' tradition.

⁵⁸ Cuttance, p. 76.

⁵⁹ Ross and Postlethwaite (n. 42 above).

⁶⁰ Cuttance, p. 79.

⁶¹ Ibid., p. 82.

⁶² Namibia Ministry of Education and Culture et al. (n. 2 above).

information allowing schools to be compared on the basis of English and mathematics proficiencies. Originally intended to inform curriculum, instructional design, and teacher training,⁶³ the assessment's use for planning purposes came afterward. From a representative sample of 136 primary schools nationwide, all students in one randomly chosen fourth-grade class and one seventh-grade class at each school were given criterion-referenced English and mathematics tests developed to serve as baseline assessment instruments. Background information on the pupils was limited to age, sex, and home language.

While the study serves its original purpose, its adaptation for planning is problematic. Instead of serving as a "baseline" from which further measurements could be taken on even a subsample of the more than 7,000 original test takers, the study's longitudinal part is limited to pupils in only 20 of the schools. Thus, although ostensibly capable of serving as the basis for constructing appropriate pupil-level intake-adjusted models, such a small number of schools would result in unwieldy estimates.

It would seem that the assessment is intended to construct school-level intake adjusted measures from the larger data set in the hopes of "disentangl[ing] the influence of (a) family and community factors, versus (b) various school characteristics on learner achievement."⁶⁴ Learner achievement data will be linked with information on local communities from the 1991 Population and Housing Census.⁶⁵ It will *not* be possible to accomplish this aim reliably, however, in the absence of student-level data.⁶⁶

The ministry embarked on a second national assessment of upper primary school students in 1995.⁶⁷ One hopes that the design will reflect the needs of curriculum design and planning, enabling proper longitudinal measures on the achievement of the same cohort of pupils, rather than being restricted to essentially comparing two large-scale, cross-sectional surveys.

⁶³ Specifically for identifying the levels of English and mathematics.

⁶⁴ Namibia Ministry of Education and Culture et al., p. 1.

⁶⁵ *Ibid.*, p. 1, no. 4.

⁶⁶ Attempts to use data from the baseline assessment for planning purposes not only have been frustrated by the original research design but also have been compounded by questions on the analyses made of the existing cross-sectional data. Thus, a multilevel, variance components model is used to analyze—albeit in the absence of both pupil-level intake measures and socioeconomic background information—the proportion of the variation in English and math achievement attributable to between-school vs. between-pupil factors. It is found that, in both subjects, about two-thirds of the variation can be attributed to between-school differences, which the authors go on to state reflect "the vast inequality in family wealth and school quality across Namibia's local communities" (Namibia Ministry of Education [n. 2 above], p. 64). However, despite the extensive clustering between schools, the authors do not point out the problems of utilizing a single-level model in the remaining analyses.

⁶⁷ This is an adaptation of the SACMEQ initiative to which reference has been made in n. 43 above.

The IIEP Zimbabwe Primary School Quality Study focused primarily on the distribution of physical inputs to schools and ultimately the relationship of such inputs to reading achievement. Its interest, other than to central administrators, has been questioned. The Namibian National Learner Baseline Assessment provides information beyond that already in the hands of teachers, but the more interesting analyses either will be based on what is virtually a set of case studies or will lack reliability because the research design is inappropriate for planning purposes. Other research, of interest to schools themselves, focuses on the relationship of different pedagogical practices to student outcomes.

Botswana Junior School Study

Examples of more sophisticated school effectiveness research may be found in only a handful of studies in the Third World, among them, the Fuller, Hua, and Snyder investigation of Botswana's junior schools.⁶⁸ It combines extensive qualitative information from classroom observations of 214 teachers in 31 junior secondary schools with the outcomes of achievement tests in English and mathematics administered to the same students with a year's gap, in form 1 and form 2. It thus goes considerably beyond the physical input-driven school effectiveness studies and has a pupil-level intake measure. Four sets of variables are used: (1) material conditions and classroom inputs, (2) teacher characteristics and training, (3) teachers' pedagogical beliefs and efficacy, and (4) teaching practices and classroom rules. Yet, aside from gender, no pupil background characteristics were collected (or at least included). Further, although some multilevel analysis is carried out, and strong arguments are used in its favor, the majority of analyses utilize OLS estimates.⁶⁹

The study's primary purpose was to examine how much girls' achievement is affected by classroom practices. The limited findings were disappointing: girls' achievement on the math posttest is reduced by teachers'

⁶⁸ Fuller et al. (n. 2 above). It is somewhat ironic that I should cite another article by Fuller as an exemplification of more sophisticated school effectiveness research, given the discussion above in relation to Fuller and Clarke (n. 4 above), which depicts their two "classes" of school effectiveness research. It is certainly true that Botswana stands out as having much lower between-school differentiation than most Third World countries, as can be seen in table 1. Perhaps the mineral wealth that financed the relatively equitable expansion of the Botswanan educational system has brought it into the top league meriting such research. However, following Fuller and Clarke's argument, resource-based inputs in a higher quality system should be less "effective." Yet the authors pick out class size, in-service math teacher training, and supplementary math readers for further analysis.

⁶⁹ The authors explain that this is due to the "limited number of degrees of freedom" resulting from the 61 "teacher-form clusters" into which information on the 214 observed classrooms was reduced because "initial analyses showed that teacher behaviors were more strongly related to achievement when all observed teachers within a form were pooled, rather than matching pupils to their particular English or math teacher" (Fuller et al., p. 361). Yet the cost was the absence of potential association of particular pedagogical practices with student achievement in English or math.

use of open-ended questions.⁷⁰ This led the authors to question the cultural meaning of particular pedagogical practices deemed “effective” in the industrialized country literature and is the important point they make regarding the application of a set of “factors of effectiveness” untested in a new context. Their tentative (or perhaps post hoc) adoption of a multilevel methodology belies the importance of their research topic, which marries detailed classroom observations with macro systems-level data. Their use of OLS estimates for the purpose of cost-effectiveness measures further weakens the case made for utilizing these more sophisticated techniques.⁷¹

*Heneveld's Primary School Indicators*⁷²

A certain frustration with either the simplistic or commonsense policy implications derived from quantitative school effectiveness research or the lack of linkage between of what is known about school effectiveness and the implementation of school improvement has led some researchers away from quantitative indicators toward more qualitative monitoring-based inquiry (e.g., case studies). Those leaning in this direction often are ready to accept as given different sets of factors that determine school effectiveness. This contrasts, for instance, with querying the accepted pedagogical practices related to a learner-centered approach in the Botswana junior school study reported above. Commonly, school improvement is monitored along the accepted dimensions of school effectiveness, irrespective of these constituting a tested theory as such. Edward Heneveld appears to belong to this second school, as he bases his extensive set of indicators on 16 factors that constitute his model of school effectiveness: (1) parent and community support, (2) effective support from the education system, (3) adequate material support, (4) effective leadership, (5) a capable teaching force, (6) flexibility and autonomy, (7) high time-in-school, (8) high expectations of students, (9) positive teacher attitudes, (10) order and discipline, (11) organized curriculum, (12) rewards and

⁷⁰ It would have been preferable if the posttest score, *controlling* for the pretest score, had been used, rather than a gain score as the independent variable, given that the variance of the pretest score, which is of interest, is lost in the gain score.

⁷¹ The authors utilize class size for English (which does not survive a multilevel analysis and has borderline significance in the OLS equation), in-service math teacher training, and supplementary math readers—the latter two found to be significant in their final multilevel equation. They then use the coefficients found for these variables in the OLS equations, not the “final” multilevel equations, to calculate the cost of implementing a strategy that would produce a 1-year gain in learning by each of these routes. They argue that “the single-level OLS model is entirely appropriate” because the three “policies would be applied to a school form” and “no pupil-level processes or interventions are being introduced” (Fuller et al., p. 374, n. 36). The OLS estimates may well be applied uniformly, but it is their derivation, not their application, which is in question.

⁷² Edward Heneveld, “Planning and Monitoring the Quality of Primary Education in Sub-Saharan Africa,” Technical Note no. 14, Human Resources and Poverty Division, Technical Department, Africa Region (Washington, D.C.: World Bank, 1994), p. 1.

incentives, (13) high learning time, (14) variety in teaching strategies, (15) frequent homework, and (16) frequent student assessment and feedback. He divides the outcomes into four types: participation, academic achievement, social skills, and economic success.

Heneveld makes an important point for this discussion when he states that the use of his framework for monitoring and evaluation does "not look directly at pupil achievement. Instead, [it] seeks to assess the presence and dynamics of conditions that have been identified as conducive to effective education." He suggests, however, that "a more formalized research design is needed . . . one that links changes in the indicators to changes in pupil performance."⁷³ Moving from the sort of straightforward monitoring that Heneveld advocates to monitoring school effectiveness on the basis of such linkages, however, is a minefield for those unschooled in its dangers, as could be seen in the IIEP Zimbabwe study. For instance, ranking schools using Heneveld's descriptive indicators (or the more limited IIEP indicators) to monitor educational quality based on the presence of "conducive conditions" is entirely different from ranking effective schools based on the *relationship* of such conditions to pupil outcomes.

*Dalin's "How Schools Improve" Study*⁷⁴

Unlike those moving on a trajectory from quantitative toward qualitative investigations of school effectiveness, the "How Schools Improve" study conducted in Bangladesh, Colombia, and Ethiopia is firmly based in a qualitative tradition. Thirty-one rural primary schools were studied in depth, sorted into three categories (excellent, very good, and good) on the basis of their outcomes in three main areas: (1) degree of implementation of key aspects of the reform, (2) degree of effect on students, teachers, and the school as an organization, and (3) degree of institutionalization of the reform.⁷⁵ A comprehensive model of school improvement is developed that helps specify the particular contexts that determine different outcomes. The study begins by detailing country-level factors contributing to the success of educational reforms, then works through a panoply of factors such as the linkage among the country, district, and school levels; the assistance given in the course of the project; the different roles of head teachers, administrators, teachers, parents, government, and external donors; resources (as well as pressure); local adaptation of particular practices; and so on.

Like other school improvement projects, the focus is on accounting for educational change. This means that "success" is broader than the relatively limited student outcomes used as measures of school effective-

⁷³ Ibid., p. 14.

⁷⁴ Dalin (n. 5 above).

⁷⁵ Ibid., p. xi.

ness in the quantitative tradition. What distinguishes the study—even from others in a similar qualitative tradition—is the importance given to the different *perspectives* on change at different levels.

There seems to be a trade-off between the depth of information and the availability of reliable indicators in studies of school effectiveness and improvement. On the one hand, school improvement studies are characterized by the collection of extensive information that fills out a comprehensive model of improvement. However, as the authors of the “How Schools Improve” project acknowledge, such information also suffers from an absence of baseline data on student and school-level outcomes, while containing errors associated with data based on opinions.⁷⁶ On the other hand, in the case of school effectiveness research, not only are the outcomes relatively narrow, but many short-cuts also are still being taken—especially the use of aggregate data rather than the pupil-level data (including intake) required by multilevel models to generate reliable estimates of school effectiveness.

I have used some examples of research into school effectiveness and school improvement to illustrate the different approaches taken in the development of indicators. However, I began this article by raising the concern that a movement away from quantitative toward more qualitative research may prevent the “third wave” of school effectiveness research from ever reaching fruition in Third World contexts. What remains is to take stock of studies that have been carried out using multilevel analysis and to discuss possibilities for improving research and helping bridge the gap between school effectiveness and school improvement studies.

Multilevel Analysis of School Effectiveness in Third World Countries

A trawl of the literature revealed 16 studies of Third World school effectiveness that apply a multilevel analytic framework in some form.⁷⁷ Although all but three venture beyond initial variance components models,⁷⁸ a smaller number of the 16 go beyond a cross-sectional analysis,⁷⁹ and just one utilizes an outcome measure other than academic achievement.⁸⁰ Educational reform is about changing schools, and judging school effectiveness without a baseline measurement cannot hope to reveal what matters in that process whether from a school effectiveness or a school

⁷⁶ Ibid., p. 181.

⁷⁷ See n. 2 above.

⁷⁸ Loera and McGinn; Namibia Ministry of Education and Culture et al.; and Fuller et al. (all in n. 2 above).

⁷⁹ Fuller et al.; Bashir; Riddell, “Toward Modelling the National Study,” and “School Effectiveness”; Lockheed and Langford; and Riddell and Nyagura (all in n. 2 above); and Riddell, “An Alternative Approach” (n. 1 above).

⁸⁰ Fuller et al.

improvement perspective. Yet only a handful of studies have attempted to model even one age-cohort at more than one point in time. Taking a snapshot is much easier than a moving picture. It is certainly less costly and less complicated in terms of research design, but we know that to derive reliable measures of school effectiveness, the minimal requirement is "time-lapse photography" of the students, not a collection of snapshots.

It is the behavioral change such an approach would reveal that is at the root of appropriate design of school effectiveness research and the development of indicators to monitor and evaluate that change process. Because multilevel analysis can separate out the variance between students from the variance between schools, even researchers loathe to adopt different methodological approaches and embrace multilevel techniques often utilize a variance components analysis to judge the extent of between-school variation in the outcomes of interest, such as the initial analyses in the Fuller et al. study or the Namibia Ministry of Education and Culture study.

In their review of recent school effectiveness studies in developing countries, Fuller and Clarke give too little attention to research design.⁸¹ By contrast, I focus only on those 16 studies using a multilevel framework, with the objective of distilling an approach to the development of appropriate indicators. Table 1 illustrates the amount of variation in achievement attributable to between-school differences, divided into subjects and grades examined. The first thing to note is the much higher level of variation between schools than is generally found in industrialized countries. Although R. J. Bosker and J. Scheerens reported the average proportion of between-school variance for such studies as being between 9 and 12 percent,⁸² only in Botswana and Kerala, India, were such low figures recorded.⁸³ The average for countries represented in table 1 is 46 percent at the primary level and 41 percent at the secondary level, revealing a greater heterogeneity in achievement between schools in developing countries than in developed countries, whether or not this is due to student background differences or selection into schools on the basis of prior achievement. In the few cases in which information on pupils' prior achievement is available, the variation in achievement attributable to the school is much less, averaging 30 percent of the total.

It is also important to recognize in table 1 that notwithstanding the greater heterogeneity in schools in developing countries, the proportion

⁸¹ Fuller and Clarke (n. 4 above).

⁸² R. J. Bosker and J. Scheerens, "Issues in the Interpretation of the Results of School Effectiveness Research," *International Journal of Educational Research* 13 (1989): 745.

⁸³ "This may be due to a uniform level of quality across Botswana's junior-secondary schools in terms of teacher qualifications and instructional materials," the authors hypothesize (Fuller et al., p. 367), and similarly in Kerala, which is an exception to the norms for the rest of India.

of variation in achievement attributable to between-*student* differences, in most cases, is still greater than that between schools. In other words, as in developed countries, the attributes and background characteristics children bring with them have a greater influence on subsequent achievement than their experiences at school. Table 1 does not disaggregate that proportion of between-school variation in achievement attributable to classroom as opposed to school factors, so it is impossible to infer the importance of indicators that might be developed from variables at this level. As noted, few Third World studies have focused on the classroom.

The evidence from the three Zimbabwean studies reported in table 1 necessitates a cautionary note regarding the dangers of making inferences based on data from only one cohort of pupils, even when pupil-level intake measures are utilized. As the sampling frames for these studies are comparable and cover the gamut of school types, the increase in school-level differentiation from 1981 to 1990 is notable.⁸⁴ Though judging the stability of school effects over time from these studies is impossible, this is one of several issues on which further research attention in industrialized countries is likely to focus. Clearly, with such limited multilevel Third World educational research, many of these tangents cannot be investigated, for example, the differential effectiveness of schools for students of particular backgrounds, or the context specificity of school effectiveness, whether across levels or, indeed, across countries.⁸⁵

A comparison of the indicators used in different types of studies would help to consolidate the lessons for further research relating to school effectiveness or school improvement (and thus feed into the policy debate on funding priorities at the central, regional, or local levels) and which classroom factors are conducive to "improved learning." A proposal for a single research design for school effectiveness and school improvement follows.

Implications for Monitoring and Evaluating School Effectiveness and Improvement

The marriage between school effectiveness and school improvement has the potential for answering some of the communication problems

⁸⁴ The 1988 study contains grade 7 examination scores for 1981, the 1991 study for 1987, and the 1993 study for 1990. This increased differentiation is no doubt due to the rapidity of the expansion of the educational system following independence in 1980, together with the reliance on the private sector to subsidize education, resulting in a diversity of resource endowments and differential access. It is also of interest that the increase in differentiation is greater for mathematics than for English, rising from 42 to 74 percent over the period, compared to a rise for English from 42 to 56 percent. Perhaps this is another way of saying that background factors equalize achievements in language more than in mathematics, where school factors are more significant.

⁸⁵ See Bosker and Scheerens on stability of school effects; D. L. Nuttall, H. Goldstein, R. Prosser, and J. Rasbash, "Differential School Effectiveness," *International Journal of Educational Research* 13 (1989): 769–76; and D. Jesson and J. Gray, "Slants on Slopes: Using Multilevel Models to Investigate Differential School Effectiveness and Its Impact on Pupils' Examination Results," *School Effectiveness*

associated with the decentralized educational management movement worldwide. The lessons from reform efforts focused on national policies and those based on classroom-level changes need to be united in order to enhance *overall* understanding between central managers and the school or district-level administrators.⁸⁶ Typically, a wide gap exists between the perspectives of these groups.

Besides good communication, there also needs to be some belief in the validity and reliability of the measures of effectiveness. The correct use and interpretation of statistical data requires skills lacking among many educational managers, who are used to "knowing" their schools through inspections or supervisory visits rather than numbers. Thus, bridging these views of reality requires both training and experience. If measures of effectiveness are biased due to design faults, such as in the rankings used in the Zimbabwean study discussed earlier,⁸⁷ the potential misuse of these data is worrisome.⁸⁸ The communication gap does not disappear with the production of reliable indicators, but the lack of appropriate contextualization of "success" or "failure" risks making the whole exercise inherently dishonest. To put it starkly: school effects *cannot* be inferred in the absence of some baseline measures of student outcomes.⁸⁹

The studies reviewed here run the gamut from a centralized-ministry focus to a school-level focus. The centrally focused IIEP Zimbabwean and the Namibian studies aim to produce a list of effective schools that can be studied further.⁹⁰ Ideally, each wants to relate "effectiveness" to particular attributes of those schools that top their lists. Their shared problems are rooted in the use of statistical techniques that make it impossible to draw appropriate inferences, a point clearly made in the Namibian study.⁹¹

and School Improvement 2, no. 3 (1991): 230–71, for differential school effectiveness; and discussion of context specificity in Reynolds, Creemers, Bird, et al. (n. 4 above).

⁸⁶ See W. K. Cummings and A. R. Riddell, eds., "Alternative Policies for the Finance, Control, and Delivery of Basic Education," *International Journal of Educational Research* 21, no. 8 (1994), for a comparative discussion of different aspects of decentralization.

⁸⁷ Ross and Postlethwaite (n. 42 above).

⁸⁸ One also needs to be aware of unintended effects of the use of indicators. For instance, the raw, unadjusted league tables publicized in Great Britain have resulted in some schools not entering weaker students for examinations that they might fail, and thus bringing down the school's "pass rate" and their position in the ranking.

⁸⁹ Of course, the situation is further exacerbated if unreliable rankings are then used as the basis for resource allocation decisions.

⁹⁰ Ross and Postlethwaite, p. 45; Namibia Ministry of Education et al. (n. 2 above), p. 72.

⁹¹ On p. 75, the authors of the Namibian study point out the difficulties of making valid inferences on the basis of such a small longitudinal sample (20 schools), but the two avenues they pursue for overcoming the limitations of their research design are not tenable. On the one hand, they discuss the possibility of "equating" the grade 4 and grade 7 tests as a means of expanding the number of classrooms in the longitudinal study from 40 to 80. On the other hand, they discuss returning to the original 136 schools and assessing learners once more in grades 4 and 7. Although they would acquire a school-level measure at a further point in time, unless they followed these particular students (as, in the ideal world, they should have done with the first cohorts, had the uses of the baseline assessment for planning purposes been embraced at the start), they still will not

The Heneveld and Dalin school-level studies, by contrast, have different aims.⁹² Heneveld's list of indicators in some ways reads as a much more sophisticated version of the IIEP Zimbabwean stock-taking exercise. Hence, although focused on the school and certainly capable of being utilized effectively by schools as a monitoring device, it also feeds a central ministerial interest in attributes conducive to effectiveness. Like the monitoring and evaluation instruments of many donor education projects, measurements of Heneveld's indicators taken at different times would reveal a trajectory of quality enhancement or deterioration *on the basis of inputs or processes but not of outcomes*.

Dalin's study is aimed specifically at evaluating the progress of educational reform rather than monitoring outcomes in general. The design matrix is useful for further monitoring of an ongoing reform and would be confirmatory rather than investigative. Both Heneveld and Dalin start from a list of givens and then seek to detail whether the particular attributes are in place,⁹³ and, in the case of Dalin's study, whether the success of the reform is related to the presence of these attributes. Neither specifically addresses the relationship between these attributes and measures of *student outcomes*.

Can the different aims of school effectiveness and school improvement research be reconciled? Would such a reconciliation, no less a marriage, be fruitful? What would a joint research project look like?

In industrialized countries, where more multilevel educational research has been undertaken, one can begin to see the way forward. Much educational research has been utilized to uncover "answers"—for instance, factors that make a school "effective"—rather than the complex relationships of which specific findings are but a small part. The results of such school effectiveness research have been fed into donor educational projects worldwide. To paraphrase, "class size doesn't matter; trained teachers and textbooks do." Learner-centered instruction is the favored "process."

If production functions were the craze of the 1960s and 1970s and educational processes received more attention in the 1980s, it is the cultural context that is drawing increasing interest during the 1990s. Do the same lists of effective factors apply across contexts for all students? This question is starting to be investigated more carefully in industrialized

derive reliable measures of school effectiveness, even if—all things being equal—they could perceive a "trend."

⁹² Heneveld (n. 72 above); Dalin (n. 5 above).

⁹³ Heneveld and Dalin both build in flexibility for the local determination of appropriate indicators.

countries,⁹⁴ but not, unfortunately, in developing countries.⁹⁵ Instead, the approach tends to be piecemeal research in a handful of countries, all too often with questionable research designs, driven by a small number of industrialized-country academics utilizing donor funds and with donor timelines that emasculate the measures or make case studies out of what otherwise could be major longitudinal studies.

It is neither appropriate nor possible to design a generic school effectiveness study outside the context of a particular country. Indeed, Heneveld suggests that his conceptual framework needs to be tailored specifically to those factors that are "most important in the education system concerned, and most amenable to change within the existing context."⁹⁶ Nonetheless, there is an approach that can go beyond the monitoring of school quality in Heneveld's framework by relating it to student outcomes and also can transcend the narrowness of the exclusively quantitative studies of school effectiveness that have been carried out to date.

Before sketching the outlines of such an approach, however, it is important to make a clear distinction between monitoring for accountability and for further investigation.⁹⁷ Rather than assume, as does much of the literature, that a particular set of factors makes for effective schools, the initial stage would be to produce a rich, multilevel baseline study that identifies the factors, contexts, and processes that contribute to effective learner outcomes. Much of what this would uncover would be confirmatory in nature, but it would be grounded in the actual context of the country. This first study would be the basis for follow-up studies of the same cohort to produce measures of effectiveness over time and could be complemented by further cohort analyses. One could use this baseline to target the "most effective" and the "least effective" schools (also longitudinally), in order to follow up such identification with qualitative case studies to examine the relationships at work in such schools. With such a baseline study, one could pursue different school-based monitoring and

⁹⁴ For example, international school effectiveness and improvement research being proposed under the European Union's framework on targeted socioeconomic research.

⁹⁵ Although the school effectiveness and improvement research that forms a part of the District Level Primary Education Programme in India is an exception.

⁹⁶ Heneveld, p. 12.

⁹⁷ H. Goldstein and S. Thomas, "Using Examination Results as Indicators of School and College Performance," *Journal of the Royal Statistical Society*, ser. A, 159, pt. 1 (1996): 149-63, discuss the use of examination results for monitoring education and the identification of outliers as a "screening" device. Sally Thomas and Peter Mortimore, "Report on Value-Added Analysis of 1993 GCSE Examination Results in Lancashire" (University of London, Institute of Education, Curriculum Studies Department, London, 1994) demonstrate the use of particular residuals rankings, pointing out that "without any information on practice within the schools the analysis cannot be taken forward but for those involved with the schools, this is the start rather than the end of the project" (p. 11). Parts of this study are reported in S. Thomas and P. Mortimore, "Comparison of Value-Added Models for Secondary-School Effectiveness," *Research Papers in Education* 11, no. 1 (1996).

evaluation, whether of the sort described by Dalin to investigate the lessons for implementation or of the sort Heneveld discusses to monitor the distribution and prevalence of those factors related to school quality (effectiveness) uncovered in the baseline study.

Such an integrated project has the potential for producing reliable measures of effectiveness that can be tracked over time and that should interest not only policy makers but also classroom teachers and school administrators because changes in “effectiveness” or “improvement” can be related to contexts and practices that have been validated locally and emerge from classroom practice as much as from central-policy directives. Communicating the results of such research and evaluation would still be a formidable challenge, however, as making this information understood or putting it into practice is often more difficult than producing the results. The inclusion of people from different levels in the educational hierarchy and across qualitative and quantitative “camps” would pose a further challenge, particularly if accountability is an issue as it is in many developed and developing countries alike. However, by forging a link between the research and monitoring of school effectiveness and those having to implement educational reform, the “third wave” of school effectiveness might be saved from an untimely death and some new, applicable lessons could be revealed as well.