

---

Multilevel Methods for Analyzing School Effects in Developing Countries

Author(s): Stephen P. Heyneman

Source: *Comparative Education Review*, Vol. 33, No. 4 (Nov., 1989), pp. 498-504

Published by: The University of Chicago Press on behalf of the Comparative and International Education Society

Stable URL: <http://www.jstor.org/stable/1188451>

Accessed: 12-11-2017 20:17 UTC

## REFERENCES

Linked references are available on JSTOR for this article:

[http://www.jstor.org/stable/1188451?seq=1&cid=pdf-reference#references\\_tab\\_contents](http://www.jstor.org/stable/1188451?seq=1&cid=pdf-reference#references_tab_contents)

You may need to log in to JSTOR to access the linked references.

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://about.jstor.org/terms>



JSTOR

*Comparative and International Education Society, The University of Chicago Press* are collaborating with JSTOR to digitize, preserve and extend access to *Comparative Education Review*

## Multilevel Methods for Analyzing School Effects in Developing Countries

STEPHEN P. HEYNEMAN

### Background and Summary

When predicting academic achievement, one important problem is how to handle the “location” from which the sample is drawn since it varies considerably—different classrooms, schools, districts, states, and occasionally, different countries. Pupils in the same “units” tend to share common experiences (and educational inputs) that make their results more like each other than would be the case if pupils were to be drawn from a random population.

How should this unit of analysis problem be handled? Many studies in the 1970s used ordinary least squares (OLS). This technique assumed that the variability of each variable was identical. This assumption was clearly a problem since variance within one level was naturally very different from other levels. Mother’s educational background will certainly differ within the sample as a whole, but it will differ differently according to the classroom, the state, or the country. The same may be true of educational inputs, textbooks, and the like.

The use of OLS had to beg the question of different variability at different levels. Now there is a way to incorporate such differences using a statistical technique called multilevel analysis (MLA). What follows is a comment on the results (and the tone) of one recent experiment using MLA.<sup>1</sup>

The tone in the Riddell article implies some dismay about the good judgment of users of ordinary least squares (OLS) analytic techniques in the 1970s. But this is like faulting Charles Lindbergh for not using radar. There is little doubt that the new computer packages that allow easy access to MLA of pupil, teacher, classroom, district, and state differences is an improvement over OLS techniques of 10 years ago. Nor is there any doubt that “the story” presented as a result of using MLA techniques is different from using OLS alone. The question is whether previous results are null and void and whether, as implied by Riddell, previous analyses were deficient in their use of tools available at the time.

I would like to acknowledge the helpful comments received from Marlaine E. Lockheed, but the views are mine alone, and, in particular, they should not be interpreted as necessarily consistent with any policy of the World Bank.

<sup>1</sup> Abby Rubin Riddell, “A Multilevel Analysis of School Effectiveness in Zimbabwe: A Challenge to Prevailing Theory and Methodology,” in this issue.

Permission to reprint this commentary may be obtained only from the author.

## MLA Results

Multilevel methods for analyzing hierarchically structured data were theoretically available in the early 1970s but were hampered by the absence of computationally efficient algorithms.<sup>2</sup> In recent years, computational methods have been developed that address this problem and have allowed the practical use of multilevel analysis, of which Riddell's study is an example.<sup>3</sup> What is certain, however, is that experience is expanding and impressions are changing rapidly. In fact, the whole idea of a methodological discussion as though it were bivariate is probably passé, the more relevant questions now concerning which combination of techniques to utilize and under which circumstances to use them—LISREL, OLS, partial least squares, iterative generalized least squares, hierarchical linear modeling, among others. The most comprehensive comparison of different modeling techniques and their different academic achievement results will soon appear in Cheung, Keeves, and Sellin.<sup>4</sup>

Using MLA enables the researcher to first partition the variance in some indicator—say, scores on an achievement test—into “between-individual” and “between-group” (classroom, district, etc.) components, with the levels of groups determined by the sampling design. Fixed effects for each level can then be estimated conventionally. Next, variable slopes between groups are examined and fitted. Most MLA research to date has employed only a few individual characteristics (e.g., gender, race, cognitive capacity) and a few group characteristics (e.g., classroom, school, and school district) to model variance at either individual or group level or to model within-group variance. The method identifies the total variance and its components. So far as I am aware, the most elaborate estimates so far are in Bryk and Raudenbush, Lockheed and Komenan, or in Lockheed and Longford.<sup>5</sup>

<sup>2</sup> A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum Likelihood for Incomplete Data via the E. Algorithm (with Discussion),” *Journal of the Royal Statistical Society* 39 (1977): 1–38; D. V. Lindsey and A. F. M. Smith, “Bayes Estimates for the Linear Model (with Discussion),” *Journal of the Royal Statistical Society* (1972): 1–41.

<sup>3</sup> S. W. Raudenbush and A. S. Bryk, “A Hierarchical Model for Studying School Effects,” *Sociology of Education* 59 (1986): 1–17; N. T. Longford, “A Fast Scoring Algorithm for Maximum Likelihood Estimated in Unbalanced Mixed Models with Nested Random Effects,” *Biometrika* 74 (1987): 817–27; H. Goldstein, “Multilevel Mixed Linear Model Analysis Using Iterative Generalized Least Squares,” *Biometrika* 73 (1986): 43–56.

<sup>4</sup> K. C. Cheung, John Keeves, and Norbert Sellin, “Multi-level Analysis in Educational Research,” *International Journal of Educational Development* (in press).

<sup>5</sup> Anthony S. Bryk and Stephen W. Raudenbush, “Toward a More Appropriate Conceptualization of Research on School Effects: A Three Level Linear Model,” *American Journal of Education* 97 (November 1988): 65–109; M. E. Lockheed and A. Komenan, “Teaching Quality and Student Achievement in Africa: The Case of Nigeria and Swaziland,” *Teaching and Teacher Education* 5, no. 2 (1989): pp. 93–113; M. E. Lockheed and N. T. Longford, “Multi-level Models of School Effectiveness in Thailand” (Population and Human Resources Department, World Bank, Washington, D.C., October 1988).

From a secondary school sample in Zimbabwe, Riddell finds that, using MLA techniques, social background is a more important influence on academic achievement than school quality. Moreover, she suggests that this may be generally typical of developing countries. To test the question, fairly, of whether school inputs are more powerful in developing countries using MLA techniques, one would need to incorporate into the equation the variance of school inputs within each country.<sup>6</sup> Based on past analysis, it is clear that school inputs do vary significantly within countries, though I would hesitate to guess whether that variance was greater in the developing countries or greater within the Organization for Economic Cooperation and Development (OECD) countries.<sup>7</sup> The point is that a true MLA test of the generalization of school effects being larger in developing countries would require intracountry information on variance.

According to Riddell, the effect of secondary school quality in Zimbabwe is modest, but is her measure net of all school effects? The problem is that she has labeled the student academic intake score from grade 7 as though it were a pupil "background characteristic." This is inaccurate. Although it is a characteristic that the pupil brings to the secondary school and, except for selection itself, it is a characteristic over which the *secondary* school has little or no control, it is hardly a characteristic with which a student was born. Can one accept that a pupil's elementary school achievements in arithmetic, science, and reading comprehension are solely due to the pupil's home? Again, Riddell would have been wiser to follow the rules proposed by Aitkin and Longford for use of MLA techniques and use an intake score of intellectual ability, such as the Verbal Reasoning Quotient or perhaps Raven's Progressive Matrices, with which I experimented when collecting data in Uganda in 1972.<sup>8</sup>

There is little question that the results of using MLA differ from using OLS alone. But different results do not always suggest that past results are wrong. While examples emerge of increased pupil effects, it remains to be seen whether these increased effects are limited to less industrialized countries. It may very well be true that the power of pupil characteristics using the MLA techniques will still be less in the less industrialized societies by comparison to OECD countries. Neither the Riddell article nor any other has sufficient evidence to answer this question.

<sup>6</sup> This is analogous to the suggestion made by Aitkin and Longford that, as a minimum requirement, an MLA sampling design must include the group variance among schools districts. See M. Aitkin and N. Longford, "Statistical Modelling Issues in School Effectiveness Studies," *Journal of the Royal Statistical Society* 48 (1986): 25.

<sup>7</sup> Stephen P. Heyneman and William A. Loxley, "The Distribution of Primary School Quality within High- and Low-Income Countries," *Comparative Education Review* 27 (February 1983): 108–18.

<sup>8</sup> Aitkin and Longford, p. 4; Stephen P. Heyneman, "A Brief Note on the Relationship between Socioeconomic Status and Test Performance among Ugandan Primary School Children," *Comparative Education Review* 20 (February 1976): 42–47.

Nor are all MLA results different. Despite the paucity of data that lend themselves to MLA techniques, results suggest some overlap with previous findings: (1) a pupil's prior achievement has always been the best predictor of future achievement; (2) the predictive power of a pupil's socioeconomic status always seems to be greater in the case of language than arithmetic; and (3) the influence of a specific teacher always seems weaker than aggregating the influence of all teachers to which a given pupil has been exposed. Thus, though divergent in some respects, findings from MLA techniques are sometimes consistent with the results of using OLS.

These are the main points. Multilevel analysis results do not suggest that the predictive power of school inputs in less industrialized societies is identical to that of industrialized societies nor that the effects of those inputs are minuscule by comparison to the pupil. Besides these points, there are several side issues raised by the Riddell article also worthy of mention.

#### Side Issues

Riddell implies that school effects analysts of the 1970s—Joe Farrell, Ernesto Schiefelbein, and myself, among many others—were in some way deficient in our use of OLS, and, worse, that we used OLS to the exclusion of other available techniques. This characterization is inaccurate and unfortunate. Like Lindbergh, we struggled to get to Paris and used every available mechanism at our disposal. We used management analyses to get at the causes of school and district input variance.<sup>9</sup> We tried to employ achievement gain scores as opposed to cross-sectional scores in order to differentiate hypothesized, as opposed to real, changes in learning.<sup>10</sup> We used pupil affiliation with schools, as opposed to school inputs, in order to overcome misspecification of school and teacher measures.<sup>11</sup> We used time-series data, discrimination analysis, and cross-tabulations to ferret out the possibilities of error in our interpretations.<sup>12</sup> And we used new path models that incorporated changes in the labor market over time so as to avoid misspecification based on typically static models used in North America.<sup>13</sup> But most important, we went to some length to test

<sup>9</sup> Stephen P. Heyneman, "Changes in Efficiency and in Equity Accruing from Government Involvement in Ugandan Primary Education," *African Studies Review* (April 1975): 51–60.

<sup>10</sup> Stephen P. Heyneman and Dean T. Jamison, "Student Learning in Uganda: Textbook Availability and Other Factors," *Comparative Education Review* 24, pt. 1 (June 1980): 206–20.

<sup>11</sup> *Ibid.*

<sup>12</sup> Ernesto Schiefelbein and Joseph P. Farrell, *Eight Years of Their Lives* (Ottawa: International Development Research Centre, 1982).

<sup>13</sup> Joseph P. Farrell and Ernesto Schiefelbein, "Education and Status Attainment in Chile: A Comparative Challenge to the Wisconsin Model of Status Attainment," *Comparative Education Review* 29 (November 1985): 490–506.

the school-effects theories with actual experiments, as opposed to surveys.<sup>14</sup> Our most serious lack of information came from the absence of time-on-task and other classroom process characteristics with sufficient rigor to have been included on the regular surveys, though this gap is now being rectified.<sup>15</sup> In sum, it is certainly true to say that the analytic techniques of the 1970s were inadequate, but it is unfair to say that they were monotonal.

One argument in the Riddell article, a common criticism of sociological research in developing countries, is that the measures of socioeconomic status (SES) are misspecified, that there may exist alternative measures that better capture the SES differences.<sup>16</sup> This is quite possible and, in fact, normal. The difficulty comes when one seeks to transfer such measures across cultures. Having “vanity licence plates” may figure into social prestige in California, but not in France. The occupation of “headman,” while prestigious in Bugandan and Busogan cultures around Lake Victoria, implies something very different in the north, where headmen were chosen among colonial authorities and often imported from other ethnic areas. Similarly, land and cattle ownership mean different things in different cultures. It is true that the three standard sociological measures—parental occupation (carefully validated and scaled), income, and educational attainment—may assume different values in different cultures. It is also fair to say that they are more universal than other measures.

Is it possible that the effect of parental educational attainment on pupil academic achievement is less among pupils in developing countries because the measure is less valid or because its ability to capture SES is

<sup>14</sup> The experimental studies are particularly relevant to the MLA debate because we eliminated all the variation by unit. Students either had the new input or they did not. In other words, the variation was either 100 percent or it was zero based on whether a student was in a control or an experimental group. In both cases the result of the input produced a change in achievement many times what would have been expected were the experiment to have been conducted in an OECD country. In the case of having access to textbooks in the Philippines, the result was equivalent to what would have been the case were class size in the United States to have been reduced from 40 down to 10. More than anything else, these experimental studies proved what the multiple regressions of survey data could only infer—that the power of improved school inputs to improve academic achievement was highest where school quality was the lowest, in the least developed countries of the world. See Stephen P. Heyneman, Dean T. Jamison, and Xenia Montenegro, “Textbooks in the Philippines: Evaluation of the Pedagogical Impact of a Nationwide Investment,” *Educational Evaluation and Policy Analysis* 6 (Summer 1984): 139–50; Dean T. Jamison, Barbara Searle, Klaus Galda, and Stephen P. Heyneman, “Improving Elementary Mathematics Education in Nicaragua: An Experimental Study of the Impact of Textbooks and Radio on Achievement,” *Journal of Educational Psychology* 73 (1981): 556–67.

<sup>15</sup> Bruce Fuller and Stephen P. Heyneman, “Third World School Quality: Current Collapse, Future Potential,” *Educational Researcher* 18, no. 2 (March 1989): 12–20; M. E. Lockheed and A. Komenan (n. 5 above); E. Jiminez, M. E. Lockheed, and N. Wattanawaha, “The Relative Efficiency of Private and Public Schools: The Case of Thailand,” *World Economic Review* 2 (1988): 139–64.

<sup>16</sup> The Riddell argument is typical in this regard. It criticizes the use of such measures as being culture bound despite the fact that it uses them, too. But Marlaine Lockheed, Bruce Fuller, and R. Nyong'o, “Family Effects on Student Achievement in Developing Countries,” *Sociology of Education* (in press), provide examples of culturally more accurate measures of socioeconomic status for Malawi and Thailand.

more imperfect? Perhaps, but if it is less valid, that in itself would be interesting. However, I suspect that it is not, at least in a systematic way. The demand for education (and income) in developing countries remains high. There is a reason why the independent black government in Zimbabwe, the location of Riddell's study, doubled its secondary school population in 2 years. That reason was popular demand. It would be hard to believe that such demand would exist if education were not valued. And since education is highly valued, it would be hard to believe that achieving more of it was not prestigious. Therefore, the reason for the differences in the predictive power of parental education on pupil achievement is not likely to be due to the lack of systematic validity in the measure of educational attainment. It must be due to something else.

Riddell mentions only one of three explanatory theories, the one drawn from sociology. It might have been wiser to cite others drawn from economics and from social anthropology.<sup>17</sup> But the point is not whether one theory or another is correct. There is enough curiosity about these questions to keep scholars profitably engaged for the next few years. The point is whether there is reason for a theory at all. Riddell implies that there is not. I believe there is.

Nothing I have seen published using OLS, MLA, or any other technique suggests to me that the predictive power of pupil SES is identical in all countries. It differs by subject matter of the dependent variable. It differs by level of educational institution—primary, secondary, higher. It differs within different ethnic groups. It differs by school availability. And it differs by school quality.

No academic debate, or any new piece of computer software, can negate what is perfectly obvious to every minister of education in every developing country, including Zimbabwe—that even parents of low socio-economic status want more education for their children and will sacrifice a great deal to keep their children in school. While we may argue over the relative importance of one effect versus another, such arguments are irrelevant in the world of policy, where the only relevant questions are how to raise the availability of school quality inputs and how to distribute them more fairly. No one seriously argues that they should not be raised because academic achievement is conditioned by the home.

I believe a methodological discussion of this kind can help clarify the issues and the changes that have emerged in the means by which we are

<sup>17</sup> Stephen P. Heyneman, "Why Impoverished Children Do Well in Ugandan Schools," *Comparative Education* 15 (June 1979): 175–85; "Differences between Developed and Developing Countries: Comment on Simmons' and Alexander's 'Determinants of School Achievement,'" *Economic Development and Cultural Change* 28 (January 1980): 403–6; Stephen P. Heyneman and William W. Loxley, "The Effect of Primary School Quality on Academic Achievement across 29 High- and Low-Income Countries," *American Journal of Sociology* 88 (May 1983): 1162–94.

able to ask questions. But we must all remember that there is still a residual. No new technique has been able to achieve an  $R^2$  of one; no new method has solved our problem of predicting with perfect clarity why some children perform better in school than others. Home influences, intelligence, teaching techniques, and so forth, are all possibilities and will be the subject of our search for many years to come.