The Use of Cross-National Comparisons for Local Education Policy

STEPHEN P. HEYNEMAN
Vanderbilt University
Nashville, Tennessee, USA

Cross-national education survey research was first just an experiment, born out of a chance visit of Torsten Husen from the University of Stockholm to the Comparative Education Center at the University of Chicago in the mid 1950s. There Husen met C. Arnold Anderson, Mary Jean Bowman, and Ben Bloom, whose view was that the whole world should be seen as a single educational "laboratory." From this meeting emerged the International Association for the Evaluation of Educational Achievement (IEA), which, for diplomatic purposes, was managed from Sweden. First results appeared in 1964. Since then there have been 33 cross-national studies, 29 of which have been associated with IEA (Table 1).

From the beginning, a myriad of problems emerged—the logistics of managing such a massive enterprise, the complexities of agreeing on common definitions, methodologies, sampling, and data management. These problems have been the focus of a great deal of official and unofficial assessments of the state of the art (Olkin & Searls, 1985; Horvitz, 1992; Medrich & Griffith, 1992; Goldstein, 1995; Postlethwaite, 1999; Linn, 2002; Chromy, 2002; Buchmann, 2002; Floden, 2002; Smith, 2002; Rowen, 2002), and these assessments have helped generate a number of significant improvements in the standards expected for cross-national surveys (National Research Council, 1985, 1990, 1995, 1999; Martin, Rust, & Adams, 1999). In spite of this progress, doubts and skepticism remain (Rotberg, 1990) and these have generated carefully constructed replies (Bradburn et al., 1991).

Reviews of cross-national policy implications have focused attention on better understanding of the generalizations associated with the influences on academic achievement. One of the most persistent generalizations has been that the influence of the home has been greater than the influence of the school itself (Coleman et al., 1966). Heyneman and Loxley, however, found that the degree of this influence varied across nations, and that the lower a nation's gross domestic product (GDP), the more influence the school seemed to have (Heyneman & Loxley, 1983). This finding has been

Curriculum Inquiry 34:3 (2004)
Published by Blackwell Publishing, 350 Main Street, Malden, MA 02148, USA, and 9600 Garsington Road, Oxford OX4 2DQ, UK.

TABLE 1
Selected International Comparative Studies in Education: Scope and Timing

Sponsor	Description	Countries	Year(s) Conducted
IEA	First International Mathematics Study (FIMS)	12 countries	1964
IEA	Six Subjects Study:		1970–1971
	Science	19 (systems)	
	Reading	15 countries	
	Literature	10 countries	
	French as a foreign language	8 countries	
	English as a foreign language	10 countries	
	Civic education	10 countries	
IEA	First International Science Study (FISS) (part of Six Subjects Study)	19 systems	1970–1971
IEA	Second International Mathematics Study (SIMS)	10 countries	1982
IEA	Second International Science Study (SISS)	19 systems	1983–1984
ETS	First International Assessment of Educational Progress (IAEP-I, Mathematics and Science)	6 countries (12 systems)	1988
ETS	Second International Assessment of Educational Progress (IAEP-II, Mathematics and Science)	20 countries	1991
IEA	Reading Literacy (RL)	32 countries	1990–1991
IEA	Computers in Education	22 countries	1988–1989
		12 countries	1991–1992
Statistics Canada	International Adult Literacy Survey (IALS)	7 countries	1994
IEA	Preprimary Project:		
	Phase I	11 countries	1989–1991
	Phase II	15 countries	1991–1993
	Phase III (longitudinal followup of Phase II sample)	15 countries	1994–1996
IEA	Language Education Study	25 interested countries	1997
IEA	Third International Mathematics and Science Study (TIMSS):		
	Phase I	45 countries	1994–1995
	Phase II (TIMSS-R)	About 40	1997–1998
IEA	Civic Education Study	28 countries	1999
OECD	Program for International Student Assessment	32 countries	2000 (reading) 2003 (math) 2006 (science)

Source: Chromy (2002).

a principal rationale for the investment in school quality by the World Bank, USAID, UNESCO, and many other development-assistance agencies, and although recent reanalyses have challenged the strength of the earlier findings, the current conclusions are that the influence of socioeconomic status on achievement is by no means uniform across nations, age/grade levels, genders, and subject matter (Baker, Goesling, & LeTendre, 2002).

More recent policy reviews have tried to speculate on the meaning of these cross-national projects on local education policy within the United States. Schmidt, for instance, has helped Americans focus on the weaknesses of having a "splintered curriculum" in which a scattering of topics are presented to students with insufficient attention to the progression, sequencing, and review that one finds in more academically successful nations (Schmidt, 2002; Schmidt et al., 2001).

In an extraordinary new paper, Baker has used cross-national studies to generate novel hypotheses across a wide variety of education policy characteristics. He finds that American school systems distribute school resources less fairly and are less able to educate the most disadvantaged students compared to other countries. He also uses cross-national studies to challenge common assumptions, for instance, that American schools are more likely to experience violence and classroom disruption. He finds that more homework is often a proxy for poor educational quality and that nations with effective family-assistance policies are more able to overcome the handicap of social poverty. He uses these cross-national studies as background for speculation that the reason why American school systems have such significant policy difficulties stems from their public governance by which educational elections determine policy direction. Because the nature of political campaigns have a single and simplistic focus that attracts the voters for school board elections (sex education, get God back in school, more time on task, class size, etc.), the nature of American education reform seems to be as fractured as the sequencing in its curriculum (Baker, 2002).

Within this extraordinary flurry of data and interpretation comes the article by Howard Russell on the "Connections among Factors in Education." The article raises four questions: (1) Whether inferences about performance are likely to be valid; (2) What happens when content becomes more rigorous, and time for learning does not; (3) Why school graduates complain that they are not sufficiently prepared for employment; and (4) Why many students fail to complete their schooling. This rather strange mixture of questions is partially relevant to cross-national studies but more often related to the local educational debates in Canada (and a few other industrialized nations) independent of the choice of methodology. Briefly, this note will assess each of these questions and the answers provided. By way of a conclusion, I will offer a brief summary on the appropriate use of cross-national studies.

Question 1, about inferences, covers many, but certainly not all, of the issues of sampling and the degree or representativeness of cross-national

studies. This set of problems and answers are well covered elsewhere (see work referred to near the beginning of this article). In essence, bemoaning the weaknesses of cross-national methodologies does little to address the problems inherent in all large-scale survey research. The principle, which all beginning graduate students must master, is that each research technique is associated with particular strengths and weaknesses. An experimental design may be more adequate to answer the question of causality, but not generalizability. A cross-sectional survey may be more adequate for the latter but not the former. The larger the survey coverage, the weaker the degree of content depth. The fact that less than 100 percent of the Ontario mathematics curriculum was covered in a study that attempts to cover the curricula in 45 countries is inevitable. The less-than-perfect content coverage is simply part of the sacrifice one makes when one chooses to enlarge the scope of analysis.

Given identical time and resources, a survey of students in Catholic schools within Toronto may offer more depth of understanding than if the survey had to encompass all schools in Toronto, or all schools in Ontario, or all schools in Canada, and certainly all schools in the world. Given constant resources, the larger the area of generalization, the thinner the analytic depth in content. On the other hand, there are virtues to enlarging the area of a survey. There are important reasons why one might extend a survey beyond Catholic schools in Toronto, and quite good reasons why one might extend a survey beyond Canada. Key to the use of these data has little to do with whether they are "valid" or not. Given the degree of progress already made on their methodological design, the TIMSS and other cross-national findings should be judged on the basis of whether they are sufficiently interesting to generate new and perhaps compelling hypotheses about the nature of local school policies that deserve more close scrutiny employing more targeted methodologies. Cross-national surveys will never supplant more targeted data-collection efforts and to judge their utility by faulting their comparative differences in content depth is not very constructive. It is like complaining that the elephant is big. After all, it *is* an elephant. Scholars and policy analysts should learn to live with it, and use it for things only an elephant can do well.

Question 2, about the increase in curricular substance not being matched by changes in the opportunity to learn, is genuinely interesting for any kind of educational research, but has little to do with cross-national surveys. The thesis is rather straightforward, and stems from one unfortunate but universal characteristic of school systems: the propensity to design content independent of the resources necessary to deliver it effectively (Heyneman, 1987). Russell's truly interesting suggestion has to do with the analysis of individual-item results in conjunction with data on opportunity to learn. In this way, one might better understand the difference across countries, and reduce distortions that result from inevitable differences in curriculum coverage. This "time-equivalent scale" in which each item

might be coded according to its degree of coverage, might then lead to a measure in the rate of learning, holding constant differences in opportunity. This would be no small accomplishment.

Russell moves from a fairly detailed methodological discussion into speculation over how teachers "game" the system in response to the (perverse?) incentives generated by government-imposed performance standards. Teachers might limit their teaching to test items (i.e., reducing pedagogical coverage). They might increase the amount of time devoted to learning, a strategy with ample precedent, particularly during times of austerity (Heyneman, 1983). Or they might limit breadth of students who sit for statewide examinations (i.e., lower retentively).

To counter this latter strategy, a new formula is discussed that is designed to control the degree of "exclusivity" of students sitting for examinations. Russell incorrectly assumes that exclusivity based on a student's social background has a strong and constant association with academic performance. Because of this mistake, the formula presented (achievement × coverage × exclusivity) would be problematic were it to be applied. The reality—as evidenced by cross-national studies—is that higher performance is not associated with identical characteristics. In some instances, socioeconomic status might be a primary associate. In other instances, the primary associate (and, hence, the primary criteria of exclusivity) might be student motivation. Associates of better student performance shift in response to age, grade level, gender, and subject matter. Results are not the same for mathematics as for language, arts, or civics.

Moreover, the formula presented by Russell is described as being necessary to counteract a logical set of teacher- or school-initiated responses to (perverse) incentives. It is true, as he suggests, that drop-out rates affect performance. Given equal coverage at entrance, and equal provision of educational quality, a school in which only 60 percent of the students complete a normal cycle is at an advantage over a school in which 90 percent complete it, because the students left in the school are usually the brighter ones and hence may score better on external tests. But how often is it the case that all other factors are equal?

Normally, schools with higher drop-out rates are also associated with lower levels of educational quality, lower levels of social background, higher levels of poverty, greater demand for outside incomes from students, and higher levels of student illness and disease To obtain the "true achievement scores" that Russell argues is necessary, one would have to control for all the major intervening influences, not only some of them, and if precedent is any guide, these controls would elicit opposite results. Higher exclusivity places schools at an advantage, but an advantage associated with lower school quality, poverty, and the like. Thus the formulae as presented by Russell designed to overcome the differences in achievement associated by differences in student exclusivity will remain interesting, but as operational policy will remain solely speculative.

Question 3 as to why students complain about not being well prepared for employment is spurious. Perhaps Canadian students complain because, unlike other societies (Japan?), they are taught to blame someone else for their problems. Perhaps the degree of complaint is unassociated with the degree to which there is a problem. Perhaps schools are blamed when the real culprit for unemployment is elsewhere. Who knows? But situating this discussion in the context of an article on "connecting factors in education" is misplaced.

Question 4 as to why so many students fail to complete their education appears to be a "straw man." The problem is raised, but it is then concluded that it is a "nonproblem." In an era in which a higher proportion of young people attend postsecondary education than at any time in history, and especially in Canada, which has one of the higher rates of postsecondary attendance in the world, it is not clear why one would suggest that low coverage (low being defined perhaps as less than 100 percent?), is a relevant policy topic.

Is it the case that less than 100 percent of Canadian youth finish secondary school? Yes. But finishing secondary school is a cultural construct, and many industrialized societies are now belatedly discovering that, however laudable, this social objective may not be consistent with normal human development. Many young people want to leave school and go to work. Some want to join the military. Some want to travel. Compulsory education legislation is often not coordinated with finishing secondary school. Often, the ages between 16 and 18 appear as an anomaly, in which a young person may have the legal right to work but has the social expectation of finishing the secondary school cycle. In some school districts in the United States, this anomaly is being addressed through significant reforms in the provision of secondary education through 24/7 stop and study institutions in which students between the ages of 16 and 18 can complete their secondary cycle at a schedule convenient to the other demands on their time. Perhaps the problem in Canada is similar. Perhaps the drop-out rate (the rate of retentively) is not a student problem at all but a signal that there is insufficient innovation in the provision of Canadian education.

SUMMARY

Torsten Husen, C. Arnold Anderson, Mary Jean Bowman, and Benjamin Bloom should be proud today. Almost 50 years after their meeting, their dream of using the world as a research laboratory to better understand the process of education is here. Cross-national projects, which attempt to calibrate the degree of curriculum coverage and the influences on achievement, are a normal, regular, and permanent part of the world of education.

What they did not expect, however, is that some would try to interpret the results of this open laboratory as if it would have a direct application

to changes in local education policy. The concern expressed by Russell over the likelihood of policy distortion from overgeneralizing from cross-national studies is legitimate. Not one of the founders of cross-national education studies would have dreamed of using their cross-national projects to replace more focused local analytic work having more relevant applications to local policy. The appropriate function of cross-national work is to inform us, it is not to direct us. It is the "elephant" of education research. Cross-national education research is a marvelous and informative instrument, but it would be irresponsible of us to expect the elephant to fly.

REFERENCES

Baker, D. P. 2002. "Should we be more like them"?: Reflections on causes of cross-national high school achievement differences and implications for educational policy reform. Paper presented at the Education Policy Conference on the American High School Today. Washington, DC: Brookings Institution.

Baker, D. P., B. Goesling, and G. K. LeTendre. 2002. Socio-economic status, school quality and national economic development: A cross national analysis of the "Heyneman-Loxley" effect. *Comparative Education Review* 46: 291–313.

Bradburn, N., E. Haertel, J. Schwille, and J. Torney-Purta. 1991. A rejoinder to "I never promised you first place." *Phi Delta Kappan* June: 774–77.

Buchmann, C. 2002. Measuring family background in international studies of education: Conceptual issues and methodological challenges. In *Methodological advances in cross-national surveys of educational achievement*, National Research Council, ed., pp. 150–98. Washington, DC: National Academy Press.

Chromy, R. R. 2002. Sampling issues in design, conduct and interpretation of international comparative studies of school achievement. In *Methodological advances in cross-national surveys of educational achievement*, National Research Council, ed., pp. 80–117. Washington, DC: National Academy Press.

Coleman, J. S., E. Q. Campbell, C. J. Hobsen, J. McPartland, A. M. Mood, F. D. Wienfall, and R. L. York. 1966. *The equality of educational opportunity report.* Washington, DC: U.S. Government Printing Office.

Floden, R. E. 2002. The measurement of the opportunity to learn. In *Methodological advances in cross-national surveys of educational achievement*, National Research Council, ed., pp. 231–67. Washington, DC: National Academy Press.

Goldstein, H. 1995. *Interpreting international comparisons of student achievement.* Paris: UNESCO, 1995.

Heyneman, S. P. 1983. Education during a period of austerity: Uganda, 1971–1981. *Comparative Education Review* 27 (3): 403–13.

———. 1987. Curriculum economics in developing countries: An emerging crisis in developing countries. *Prospects* 18 (1): 63–74.

Heyneman, S. P. and W. Loxley. 1983. The effect of school quality on academic achievement across twenty-nine high- and low-income countries. *American Journal of Sociology* 88 (6): 1162–94.

Horvitz, D. 1992. *Improving the quality of international education surveys.* Draft prepared for the Board on International Comparative Studies in Education. Washington, DC: National Research Council.

Linn, R. L. 2002. The measurement of student achievement in international studies. In *Methodological advances in cross-national surveys of educational achieve-*

ment, National Research Council, ed., pp. 27–58. Washington, DC: National Academy Press.

Martin, M. O., K. Rust, and R. J. Adams. 1999. *Technical standards for IEA studies:* Amsterdam: International Association for the Evaluation of Educational Achievement.

Medrich, E. A. and J. E. Griffith. 1992. *International mathematics and science assessments: What have we learned?* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

National Research Council. 1985. *Summary report of the conference on October 16–17, 1985 (draft) Committee on National Statistics, Commission on Behavioral and Social Sciences and Education.* Washington, DC: National Academy Press.

———. 1990. *A framework and principles for international comparative studies in education.* Board on International Comparative Studies in Education, N. M. Bradburn and D. M. Gilford, eds., Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

———. 1995. *International comparative studies in education: Descriptions of selected large-scale assessments and case studies.* Board on International Comparative Studies in Education Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

———. 1999. *Next steps for TIMSS: Directions for secondary analysis.* Board on International Comparative Studies in Education, A. Beatty, L. Paine, and F. Ramirez, eds., Board on Testing and Assessment, Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Olkin, I. and D. T. Searls. 1985. Statistical aspects of international assessments of science education. Paper presented at the conference on Statistical Standards for International Assessments in Pre-college Science and Mathematics. Washington, DC.

Postlethwaite, T. N. 1999. *International studies of academic achievement: Methodological issues.* Hong Kong: University of Hong Kong.

Rotberg, I. 1990. I never promised you first place. *Phi Delta Kappan* December: 296–303.

Rowen, B. 2002. Large-scale, cross-national surveys of educational achievement: Promises, pitfalls and possibilities. In *Methodological advances in cross-national surveys of educational achievement,* National Research Council, ed., pp. 321–53. Washington, DC: National Academy Press.

Schmidt, W. H. 2002. Too little too late: American high schools in an international context. Paper presented at the Education Policy Conference on the American High School Today. Washington, DC: Brookings Institution.

Schmidt, W. H., C. C. Knight, R. T. Houang, H. C. Wang, D. E. Wiley, L. S. Cogan, and R. G. Wolf. 2001. *Why schools matter: A cross-national comparison of curriculum and learning.* New York: Jossey-Bass.

Smith, M. S. 2002. Drawing inferences for national policy from large-scale cross national education surveys. In *Methodological advances in cross-national surveys of educational achievement,* National Research Council, ed., pp. 2951–321. Washington, DC: National Academy Press.