

Chapter 22

The Missing Link—Incorporating Opportunity to Learn in Educational Research Analyses

William H. Schmidt, Leland S. Cogan, and Michelle L. Solorio

Abstract International comparative education has developed a rich and complex history to address a variety of questions about how education systems are similar or different in their organization, operation, and outcomes. Despite different approaches and the exploration of different research questions, a common challenge is making these comparisons both trustworthy and credible. In the language of research methodology these are issues of reliability and validity, and are of particular interest as researchers attempt to compare what some consider to be the incomparable (Husén, 1983). Our contention is that whatever approach, method, or question informs the research, a critical element to be considered is the subject matter that is the focus of the educational enterprises being compared.

Keywords Opportunity to learn • Statistical bias • Student performance

International comparative education has developed a rich and complex history to address a variety of questions about how education systems are similar or different in their organization, operation, and outcomes. Despite different approaches and the exploration of different research questions, a common challenge is making these comparisons both trustworthy and credible. In the language of research methodology these are issues of reliability and validity, and are of particular interest as researchers attempt to compare what some consider to be the incomparable (Husén, 1983). Our contention is that whatever approach, method, or question informs the research, a critical element to be considered is the subject matter that is the focus of the educational enterprises being compared.

W.H. Schmidt (✉)

Education Policy Center, Michigan State University, East Lansing, MI, USA

Center for the Study of Curriculum Policy, Michigan State University, East Lansing, MI, USA

e-mail: bschmidt@msu.edu

L.S. Cogan • M.L. Solorio

Center for the Study of Curriculum Policy, Michigan State University, East Lansing, MI, USA

© Springer International Publishing AG 2017

J.-W. Son et al. (eds.), *What Matters? Research Trends in International Comparative Studies in Mathematics Education*, Research in Mathematics Education, DOI 10.1007/978-3-319-51187-0_22

411

A hypothetical example may help clarify our contention. Imagine two photographs arranged side by side, both featuring students clad in white coats and engaged in measuring, mixing, and heating various mixtures. In one picture these activities are being conducted in a large room that has a number of stoves, ovens, and common kitchen pots and pans; in the other there are a number of open-flame “cooking stations” surrounded by various glass cylinders, bowls, and cups. Comparing the laboratories in the two pictures seems a bit odd if we assume all the students are studying the same thing. However, the differences become immediately meaningful once we realize the first is a laboratory kitchen for future chefs while the other is a laboratory for organic chemistry students. Although a picture may well be worth a thousand words, these pictures require a few words for a faithful and reasonable comparison.

The need to provide important contextual information about the subject matter in comparative education research is not limited to any particular set of research methods, goals, or questions. Indeed, such studies encompass a variety of approaches and methodologies that have been strongly influenced by the disciplinary traditions informing the research: sociology, psychology, philosophy, economics, political science, and of course various traditions within education. This history reveals a changing focus on various goals and accompanying methods. Although not strictly chronological there is a sense of development, with contemporary research typically embracing more than a single goal and often reflecting a multidisciplinary or interdisciplinary orientation.

Among the earliest comparisons of education were descriptions made during the Greek and Roman eras that documented differences among foreign—“xeno” or “barbarous”—peoples. This goal of rich description of differences continues to find expression in ethnographies and in the compendia of education indicators produced, for example, by UNESCO and OECD. A second goal has been to examine foreign education systems and practices specifically to discover novel practices, approaches, or structures that could be employed in one’s own context. Sometimes studying the mundane in an unfamiliar context, i.e., a different culture or social setting, can spark insights and perspectives that would otherwise remain tacit and unexamined. The World Bank and other international agencies often examine educational practices with the explicit humanitarian goal of improving education in order to improve people’s well-being and the overall economies in developing countries. As these goals have been pursued, an additional goal arose: to examine specific factors thought to shape education. The advent of relatively cheap and powerful computers has led many scholars to elucidate more specifically quantitative explanations, and to examine causal relationships among many different education resources, practices, and products.¹

¹The broad goals summarized here have been gleaned from histories of international comparative education by Hans (1949), Noah and Eckstein (1969), Cowen (1996), Mitter (1997), and Lundgren (2011).

The root of current large-scale comparative studies sponsored by the International Association for the Evaluation of Educational Achievement (IEA) and OECD, such as TIMSS and PISA, can be found in the interest in the late 1950s among university research professors and education ministry officials to investigate education practices and outcomes in a systematic manner. The initial 12-country pilot administered in 1960 (Foshay, Thorndike, Hotyat, Pidgeon, & Walker, 1962) was the result of a consensus in this group of “the need to introduce into comparative educational studies established procedures of research and quantitative assessment” (Husén, 1967, p. 13). Their goals included providing rich qualitative descriptive data situating education in its social, cultural, and political context, but also moving beyond this to provide insight into possible causal relationships between educational inputs and outputs.

Benjamin Bloom was a member of this group and was selected to lead the initial pilot study. To move beyond mere description required a theoretical or conceptual model that would identify constructs of interest and would inform the creation of instruments. Carroll and Bloom were an integral part of the early discussions. Consequently, the constructs embedded in Bloom’s (1974) mastery learning model and Carroll’s (1963) model of school learning served central roles in the research. More specifically, in thinking about influences on student achievement to include in the instrumentation, “one of the factors which may influence scores on the achievement examination was whether or not the students had an opportunity to study a particular topic of how to solve a particular type of problem” (Husén, 1967, pp. 162–163). This opportunity to learn (OTL) construct, termed “time actually spent on learning” in Carroll’s model, was conceptualized at the student level as his was a psychological model. Given the practical challenges of a large-scale research endeavor, the decision was made to measure OTL through a teacher survey rather than burdening students with greater response time. Measuring OTL at the classroom level through teachers’ survey responses has been a hallmark of large-scale comparative surveys. Most recently, PISA 2012 for the first time included an OTL measure and it was measured through student responses (Cogan & Schmidt, 2015).

Conducted appropriately, comparisons can lead to deep insight into a researcher’s own education system. Done poorly, however, researchers and others are left with shallow observations regarding superficial differences and similarities, observations that fail to provide insights that may be leveraged to make sense of the resultant data. Common to all of the research goals in comparative education identified earlier, implicitly if not explicitly, is a desire to learn about different education systems in order to gain insights toward potentially improving one’s own education system. The consistency of this foundational purpose of comparative education is striking and important, and underscores the important question posed earlier: “how have researchers made their comparisons both trustworthy and credible?” Throughout the evolution of international assessment studies, a problematic issue has been the tendency for researchers, policy makers, and others to use country means from the assessments to create an unsubstantiated ranking system, also referred to as a league table or a cognitive Olympics (Burststein, 1993; Husén, 1979a, 1979b). The danger is that this simplistic compilation may be used to

leverage policy objectives based on comparisons of countries' mean scores alone, while failing to take into account differences in critical educational factors—educational structures, cultures, and student learning. Those researchers who developed and analyzed SIMS acknowledged this issue explicitly: “We cannot escape the ideological use and misuse of cross-national data for political purposes. We can only hope to overwhelm the most base misrepresentations with the wealth of knowledge and understanding international studies can provide” (Burstein, 1993, p. xxxi).

Whether a comparative study uses large-scale international data to look at multiple systems of education or focuses in on one or two education systems, what actually makes comparative work meaningful and useful, is a true exploration of the learning experiences students have, which provide them the opportunity to learn the material represented on assessments researchers use to compare education system outcomes. In order to make sense of comparisons that large assessments and other comparative methods allow for, researchers must pay attention to the content and substance of the education being communicated to the students whose education is being measured. Broadly speaking, we are discussing the opportunity to learn construct as a measure of the implemented curriculum, which allows for these meaningful comparisons. Referring again to the researchers who developed the IEA and the early large-scale quantitative studies, they recognized that comparison was not possible when the absence of curricular commonality existed or was adjusted for, and as the assessments were developed the opportunity to learn construct evolved:

But the early leaders were not so naïve as to think that wishing for equity made it so. Rather they were prescient enough to introduce what may be IEA's most powerful contribution of all to the literature on educational achievement surveys; namely, the measurement of **opportunity to learn (OTL)**. (Burstein, 1993, p. xxxiii, emphasis added)

By choosing to look at educational attainment or achievement on international assessments through the lens of opportunity and with an exploration of differences in curriculum, comparisons can make sense and shed light onto why different systems of education have a different distribution of scores. That is not to say that examining different curriculums will make comparison between education systems simple; rather, the comparisons will be more meaningful if we are able to see what students have been given the opportunity to learn through intentional studies of differences in curriculum. Comparing educational opportunity itself is complicated due to the nature of the work required, and is made all the more complicated by the diverse meanings attributed to the concept of “curriculum” by educators, researchers, and teachers. In different contexts, “curriculum” can refer to textbooks, lesson plans, education frameworks, national guidelines, educational expectations, classroom activities, and a number of other attributes that make up a single system of education.

In comparative studies that incorporate factors related to opportunity to learn, “curriculum” that directly influences OTL refers to the content presented to students, the instructional opportunities that students experience, or in technical

TIMSS terms the “implemented” part of the tripartite model of curriculum. This model is comprised of the intended curriculum (what students are expected to learn as stated in national or regional goals, written frameworks, and standards), the implemented curriculum (what happens in the classroom), and the attained curriculum (what students learn). Textbooks and other learning materials comprise the potentially implemented curriculum, creating items that students will potentially have the opportunity to be exposed to, thus adding yet another element to the delicate yet complex understanding of a students’ exposure to educational opportunity.

What we have learned through the phases of international assessment work, particularly through researcher-driven developments to include information about what happens inside schools through opportunity to learn measures and other practice-relevant data points, is that what makes comparisons useful is an understanding of what material students have been exposed to, in what ways, and how often. These internal workings of education get at the heart of learning, and remain the foundation for differences in student outcomes, education systems, and educational similarities and differences. Without knowing what happens in school, comparative work is reduced to meaningless numbers in a formula or words on a page, with no foundation upon which to derive understanding. Just like mathematics requires units of measure to define the meaning of a value, comparative education requires educational opportunities and exposure to subject matter to define an outcome.

Ultimately, despite the challenges faced in comparative education studies, the subject matters (Stodolsky & Grossman, 1995). Consequently, researchers need to attend to the learning opportunities that differ across education systems, classroom practices (pedagogies), and school activities in order to draw trustworthy and credible comparisons. Although Hans (1949) argued that “the application of the findings of these studies [of comparative education] is outside the scope of Comparative Education proper and belongs in its theory to the philosophy of education and in its practice to the administration and organization of education” (p. 11), we submit that the framework informing the comparative exercise is within the scope of comparative education and that this plays a central role in the proper interpretation of the research. The substance of the education enterprise—the focus and content of the curriculum—can only be excluded from consideration to the peril of the reliability and validity of the comparisons in view.

In truth, this threat to education research does not exist solely for cross-national or cross-cultural comparisons. One of the major insights from the 1995 TIMSS curriculum analysis was the great variation in what passes for eighth-grade (13-year-olds’) mathematics across countries (Schmidt, McKnight, Valverde, Houang, & Wiley, 1997). Mathematics is studied around the world in every school system yet it is not all the same; math is not math. Much greater specification is needed to balance that equation. Our analysis of the US data underscored this, as we realized the great variation in what students studied in eighth-grade mathematics in our country was every bit as great as the variation across all participating TIMSS countries (Cogan, Schmidt, & Wiley, 2001).

From the disciplinary viewpoint of statistics, ignoring subject matter introduces bias. Many different studies have documented a relationship between students' motivation and their academic performance, and many investigate this relationship specifically in the context of mathematics. Regression analysis yields a numerical estimate of this relationship. However, motivation may well be related to the specific mathematics studied, i.e., students' mathematics OTL, as well as to students' achievement. In this case, if OTL is left out of the analysis model the estimate of the strength of the relationship between motivation and performance will be biased by the indirect effect. Consequently, one of the most critical contextual issues to be addressed in any piece of educational research is the substance (subject matter) that is the focus of what teachers are teaching and students are expected to learn.

The issue of bias can be framed mathematically for greater clarity. Assume that the following model defines the true relationship between two variables—OTL and student motivation, for example—and mathematics achievement:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e$$

where x_1 is a measure of mathematics content coverage of (OTL) and x_2 is another variable describing a different aspect of schooling such as motivation or teacher quality.

Now imagine that the researcher does not have a measure of OTL and as such analyzes the data using the following model:

$$y = \beta_0 + \beta_2 x_2 + e$$

The consequence of this, given the true relationship as described in the previous equation, is that in reality:

$$\beta_2 = \beta_2 + \beta_1 \frac{\sigma_{x_1 x_2}}{\sigma_{x_2}^2}$$

where $\frac{\sigma_{x_1 x_2}}{\sigma_{x_2}^2}$ indicates the bias that results if x_2 is related to x_1 (e.g., student motivation is related to OTL) and that OTL is related to academic achievement in mathematics, which has been well established in the literature (Schmidt & Maier, 2009; Schmidt et al., 2001; Schmidt, Burroughs, Zoido, & Houang, 2015).

Furthermore, it is our contention that content coverage in mathematics is very likely related to most other school, teacher, and student characteristics, which are also related to learning. If this is the case, then most data analyses relating those characteristics to outcome measures without the inclusion of a measure of content coverage (OTL) will produce biased estimates of the relationships of those variables to student outcomes. The direction and magnitude of the bias, however, is not known.

This suggests two important roles that the measurement of content coverage plays in educational research related to practice and policy. First, it can be

conceived of as an important outcome in and of itself. The first four chapters of this book have such a focus as they characterize country differences in textbook and classroom coverage. This coverage reflects the policies of the country as to what content should be covered in what grades, and differences can be used to inform potential policy or practice reforms. Such characterizations of content coverage are outcomes of educational policy, and many countries monitor this coverage as they do other achievement measures. Part II of the book demonstrates the same use of content coverage as an important indicator of schooling, but in the context of teacher preparation.

The other major use of OTL measures goes to their relationship to academic achievement. International studies have a long tradition of measuring student achievement in mathematics, and the results of TIMSS and PISA testing provide a rich source for country comparisons toward learning “what works”—or more precisely, determining the important variables that are related to achievement both across and within countries. Additional variables are included in such studies to characterize countries, schools, classrooms, teachers, and students. Many research studies have been published using this data, as well as those from TEDS-M and TALIS. We also find analyses in this book using all of these international data sets. But here we also find a shortcoming prevalent in the research literature: most of the authors do not control their analyses for differences across countries in terms of content coverage. This is especially true in Part IV of the book, as Sarah Lubienski’s comments on Chaps. 15–19 confirm; she discusses the limitations of cross-sectional data sets, especially in terms of confounding variables that are not measured or are ignored in the analyses.

This is a serious limitation of the studies reported on in Chaps. 15–19. Without adequate measures of OTL we do not know if the relationships described are characterizing the variables identified or are biased coefficients resulting from no accurate control of the variation in the content coverages, both within but especially across countries where we know how different content coverage can be (Schmidt et al., 2001).

TIMSS has always had measures of OTL but unfortunately they have become less specific and are not as detailed as in the original 1995 TIMSS. PISA in 2012 had OTL measures for the first time in mathematics. In general, like much of the educational research of this sort, the studies included in this book do not include these measures of OTL either, with the notable exceptions of Chap. 9 (using TEDS-M data) and Chap. 5 (using PISA data).

Despite the limitations of the studies reported on in this book, the book does make very visible the use of mathematics content coverage in international comparative research focusing on determining differences in content coverage as an important policy variable, as well as its use in reducing the potential bias associated with characterizing the relationship between various other schooling variables and academic performance.

References

- Bloom, B. S. (1974). Time and learning. *American Psychologist*, 29(9), 7.
- Burstein, L. (1993). *The IEA study of mathematics III: Student growth and classroom processes* (L. Burstein, Ed.). Oxford, England: Pergamon Press.
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, 64(8), 723–733.
- Cogan, L. S., Schmidt, W. H., & Wiley, D. E. (2001). Who takes what math and in which track? Using TIMSS to characterize U.S. students' eighth-grade mathematics learning opportunities. *Educational Evaluation and Policy Analysis*, 23(4), 323–341. doi:10.3102/01623737023004323.
- Cogan, L. S., & Schmidt, W. H. (2015). The concept of opportunity to learn (OTL) in international comparisons of education. In K. Stacey & R. Turner (Eds.), *Assessing mathematical literacy: The PISA experience* (pp. 207–216). Dordrecht: Springer. doi:10.1007/978-3-319-10121-7.
- Cowen, R. (1996). Last past the post: Comparative education, modernity and perhaps post-modernity. *Comparative Education*, 32(2), 151–170. doi:10.1080/03050069628812.
- Foshay, A. W., Thorndike, R. L., Hotyat, F., Pidgeon, D. A., & Walker, D. A. (1962). *Educational achievements of thirteen-year-olds in twelve countries*. Hamburg: UNESCO Institute for Education. Retrieved from <http://unesdoc.unesco.org/images/0013/001314/131437eo.pdf>.
- Hans, N. (1949). *Comparative education: A study of educational factors and traditions*. London: Routledge & Kegan Paul Limited.
- Husén, T. (1967). *International study of achievement in mathematics, a comparison of twelve countries* (Vol. II). (T. Husén, Ed.). New York: Wiley.
- Husén, T. (1979a). General theories in education: A twenty-five year perspective. *International Review of Education*, 25(2/3), 325–345.
- Husén, T. (1979b). An international research venture in retrospect: The IEA surveys. *Comparative Education Review*, 23(3), 371–385. Retrieved from http://www.jstor.org/stable/1187603?seq=1#page_scan_tab_contents.
- Husén, T. (1983). Are standards in U.S. schools really lagging behind those in other countries? *The Phi Delta Kappan*, 64(7), 455–461.
- Lundgren, U. P. (2011). PISA as a political instrument. In M. A. Pereyra, H.-G. Kotthoff, & R. Cowen (Eds.), *Pisa under examination: Changing knowledge, changing tests, and changing schools* (pp. 17–30). Rotterdam: SensePublishers.
- Mitter, W. (1997). Challenges to comparative education: Between retrospect and expectation. *International Review of Education*, 43(5–6), 401–412. doi:10.1023/A:1003084402042.
- Noah, H. J., & Eckstein, M. A. (1969). *Toward a science of comparative education*. New York: The Macmillan Company.
- Schmidt, W. H., & Maier, A. (2009). Opportunity to learn. In G. Sykes, B. Schneider, & D. N. Plank (Eds.), *Handbook of education policy research* (pp. 541–559). New York: American Educational Research Association, Rutledge.
- Schmidt, W. H., Burroughs, N. A., Zoido, P., & Houang, R. T. (2015). The role of schooling in perpetuating educational inequality: An international perspective. *Educational Researcher*, 44(7), 371–386. doi:10.3102/0013189X15603982.
- Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H. A., Wiley, D. E., Cogan, L. S., et al. (2001). *Why schools matter: A cross-national comparison of curriculum and learning*. San Francisco: Jossey-Bass.
- Schmidt, W. H., McKnight, C., Valverde, G. A., Houang, R. T., & Wiley, D. E. (1997). *Many visions, many aims, Volume I: A cross-national investigation of curricular intentions in school mathematics*. Dordrecht: Kluwer.
- Stodolsky, S. S., & Grossman, P. L. (1995). The impact of subject matter on curricular activity: An analysis of five academic subjects. *American Educational Research Journal*, 32(2), 227–249.
- Travers, K. J., & Westbury, I. (1989). *The IEA study of mathematics I: Analysis of mathematics curricula* (K. J. Travers & I. Westbury, Eds.). Oxford, England: Pergamon Press.