PART I

The Historical Development of Modern International Comparative Assessments

CHAPTER 1

Beginning the Modern Investigation of the Role of Schooling across the Globe

Schooling is ubiquitous across the globe. Virtually all of the almost 200 countries in the world have schooling as one of their central responsibilities. It is expected to prepare students to be responsible, informed citizens who will also contribute to the nation's economy. In that respect, schooling represents an investment in human resources that serves as a source of economic development as well as supporting the stability of society. Consequently, societies expect schooling to address both excellence and equality for their citizens. It is probably for this reason that around one-half of the world's countries have participated in one or more of the seventeen international assessments that have focused on mathematics, science, or both. Countries want some objective criteria on which to determine how well they have done on these two dimensions. This has led to the use of student achievement in these academic subjects, which have become ever more strongly related to productive and successful economies and that increasingly influence citizens' daily lives.

The metaphorical black box defining the key elements of schooling includes the student home and family background, which is brought to school; the content (including skills and reasoning) deemed by the society to be taught and learned; and the teacher who, with knowledge of the content and pedagogical skills, engages with the student over the content that is to be learned. To study schooling is to study these three key features.

A small group of university professors recognized that in order for comparisons to be made based on international assessment results, measures in addition to these assessments would be essential, since without them country comparisons would be meaningless. Using the black box metaphor, this led to the development and implementation of international measures of the home and family background of the student as well as measures that characterize what opportunities a student has had to

learn appropriate academic content. These two factors – student socio-economic background (SES) and opportunities to learn (OTL) appropriate content – have been present in various instantiations in the international studies in a fairly consistent manner.

The issue of teacher quality, how well prepared a teacher is to understand the academic content and to prepare cogent and coherent lessons around this content as well as managing and maintaining a classroom environment conducive to student learning, is vitally important for student learning. Unfortunately, this is what has not been consistently measured in international studies and is often left out. International Association for the Evaluation of Education (IEA) studies have collected many different teacher background and instructional measures, yet perhaps due in part to the limits of the amount and kind of data that can be collected in cross-sectional studies, the types of measure included such as degrees earned, years of experience, and number of professional development activities attended are weak proxies for teacher quality.

As important as teachers are in schooling, few international assessment studies have developed meaningful measures of teacher quality that have demonstrated a relationship to student learning. One international study addressed this issue through an examination of teacher preparation that gathered an assessment of future teachers' knowledge of mathematics and related pedagogy at the end of their teacher preparation program. In this book we consider this groundbreaking international study of tertiary education as well as an additional study of the training, instructional practices, and beliefs of practicing teachers.

The focus, however, is on examining K–12 schooling across the globe by using seventeen assessment studies in mathematics and science that focus on OTL, SES, and student achievement as measured by curriculum-based tests and literacy tests.

This chapter provides the historical development of international assessments in mathematics and science in relation to the IEA from the formation of the organization and the first international education assessments through the Second International Mathematics Study (SIMS). This will allow the reader to position the IEA assessments in a historical context before moving to a greater understanding of the evolution of the international assessment stages that include the addition of the Organisation for Economic Co-operation and Development's (OECD) Programme for International Student Assessment (PISA).

Setting the Stage for the IEA

The origins of "comparative education" are mostly unknown, although the tradition can be traced as far back as the days of the Roman and Greek empires (Hans, 1949; Noah & Eckstein, 1969). Comparative education as a field has evolved since the Roman and Greek era, as have the theoretical and methodological underpinnings of the field. Arguably, the field has followed a similar developmental pattern expressed in the history of all comparative studies wherein "they all started by comparing the existing institutions. . . .. Gradually, however, these comparisons led the pioneers of these studies to look for common origins and the differentiation through historical development. It unavoidably resulted in an attempt to formulate some general principles underlying all variations" (Hans, 1949, p. 6). For the purposes of this chapter, we break down the development of the field of comparative education into five stages and provide a brief overview of what each stage entailed until reaching the point at which the IEA was formed. This will provide a clear lens with which to understand how assessment came to be viewed as an important tool in comparative education.

The five stages, adopted from Noah and Eckstein's book *Toward a Science of Comparative Education* (1969), as illustrated in Figure 1.1, move from general observations of foreign schools to borrowing aspects of schooling methods, finally developing into a field characterized by the

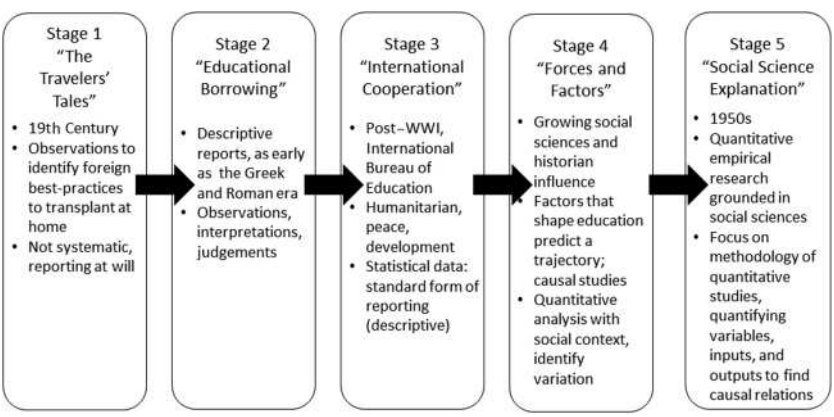

Figure 1.1 The field of "comparative education:" a brief history from approximately eighth century BC to 1969[1]

[1] The Stage titles in this model are directly quoted from Noah & Eckstein, 1969, pp. 4–7.

scientific method and its identification of variables and causal relationships between inputs (both within and outside the school system) and outputs of the school system. The first formal definition of the field has been credited to Marc-Antoine Jullien de Paris who, in 1817, saw comparative education as a way to analytically study education in other countries in order to modify and perfect national systems (Hans, 1949). Initially, the field (from a North American and European perspective) relied on descriptive comparisons that were used to help shape national education or to assess the values behind schooling while seeking best practices and eventually incorporating other educational philosophies (Cowen, 1996; Hans, 1949).

The first four stages increasingly incorporated the social sciences, building a foundation for the fifth stage of comparative education in the 1950s when the field sought deeper statistical understandings of various education systems. Following the mainstream argument that Sputnik and the Cold War altered the trajectory of the purpose of education, and therefore shifted the goal of comparative education, the move for comparative education to include more scientific comparative data through testing and statistics is not surprising (Cowen, 1996; Lundgren, 2011; Mitter, 1997). However, if we take a step back from the Sputnik theory and look at the history of the field we can see a gradual evolution into this stage, a trajectory that was already in place – Sputnik merely intensified the focus and movement in this direction.

Interest in the international comparison of education systems was already increasing around World War II when education began to be viewed as an investment in human resources and a source of economic development (Husén, 1967a). Political agendas increasingly focused on the need to identify scientific methods of comparison that served to highlight progress or areas of weakness as political competition between nations increased. The field of comparative education was a way to view the world guided by a lens of what is important in a specific context and at a specific time (Cowen, 1996). In 1949, eight years prior to Sputnik, comparative educationalists were calling for common statistics as a foundation for future comparison, including administration, organization, and tests of intelligence and achievement, at a time when "each country has its own terminology, based on national history, its own classification and its own method of collecting and compiling statistical tables" (Hans, 1949, p. 7). Such a situation made meaningful comparisons all but impossible.

While this is an incomplete picture of the state of the field at the time, it sets the stage for the 1958 research memorandum authored by Arthur Foshay of Teachers College, Columbia University, and sent to the United

Nations Educational, Scientific and Cultural Organization (UNESCO), which ultimately launched the initial phase of the development of the International Association for the Evaluation of Educational Achievement (IEA) (Husén, 1967a; Husén & Postlethwaite, 1996). The comparative education field is now replete with international assessments such as those found in the IEA studies, which adapt in purpose and structure to reflect a field that has become increasingly policy-oriented, increasingly competitive, and increasingly market-oriented (Broadfoot, 2010; Grek, 2009).

The Creation of the IEA

This fifth stage of "social science explanation" for comparative education sets the stage for the formation of the IEA and the launch of international comparative assessments in education that respond to the need for empirical evidence about student achievement and internationally comparative data. The consensus on which the IEA founding fathers based "the need to introduce into comparative educational studies established procedures of research and quantitative assessment" (Husén, 1967a, p. 13) was that previous research in comparative education provided only qualitative and descriptive information of education and culture, such as the descriptions from UNESCO, the International Bureau of Education (IBE), and the OECD, which were unable to provide insight into causal relationships among the educational inputs and outputs (Husén, 1967a). Before exploring the formation of the IEA, it would be pertinent to identify the founding fathers, together with their institutional affiliations, in order to better comprehend the context behind the conceptual framework of the IEA.

The IEA Founding Fathers

While there are many influential researchers who have been part of the IEA since the early days, there are a few who stand out as having taken an active and direct role in the actual formation of the organization. Often dubbed "the paternity," these founding fathers of the organization (listed here) are worthy of more space than this section allows:[2]

- C. Arnold Anderson, Professor, University of Chicago
- Benjamin Bloom, Professor, University of Chicago

[2] Founding fathers are listed with their academic position and affiliation held during the early days of the IEA and the First International Mathematics Study.

- Arthur W. Foshay, Professor, Teachers College, Columbia University
- Torsten Husén, Professor, University of Stockholm, Past Chairman and Technical Director of the IEA
- Douglas A. Pidgeon, National Foundation for Educational Research in England and Wales
- T. Neville Postlethwaite, Lecturer, St. Albany College of Further Education; Research Officer (Test Services), National Foundation for Educational Research in England and Wales, London
- Robert L. Thorndike, Professor, Teachers College, Columbia University
- W. D. Wall, National Foundation for Educational Research in England and Wales
- Richard Wolf, Graduate Student in Measurement, Evaluation and Statistical Analysis, University of Chicago – Studying under Professor Benjamin Bloom

The Formation of the IEA

To understand the full arc of the IEA and its impact on international assessments over time and in the present, it is important to know the goals and scope of the studies that began with the Pilot. Husén and Postlethwaite (1996) break the history of the IEA into five stages, whereas Gustafsson (2008) analyzes the organization in two stages. In this section, we will address the development of the first and early stages of the evolution and then in Chapter 2 we will provide clarity on the transition into the more mature stages and goals. As we will see in Chapter 2, the organization shifted into an administrative undertaking; however, in the early stages the IEA operated with a strong research orientation, where studies were built from research questions and hypotheses (Lundgren, 2011). The educational and social science researchers (the founding fathers) at this time in the late 1950s were interested in researching educational achievement and its determinants with the underlying assumption that factors influencing achievement are complex (Gustafsson & Rosen, 2014).

The unofficial start to the IEA was in 1958 when C. Arnold Anderson first addressed an interest in a comparative education research project that would help establish unified metrics for testing educational hypotheses. Following this came the memo by Foshay calling a variety of educational researchers to action. This action was a UNESCO Institute of Education (UIE) meeting in Hamburg followed by subsequent meetings in Eltham

and other locales. The IEA was officially organized in 1959 with the aim "to look at achievement against a wide background of school, home, student and societal factors in order to use the world as an educational laboratory so as to instruct policy makers at all levels about alternatives in educational organization and practice" (Travers & Westbury, 1989, p. v).

Thus, as these founding fathers sought to fill a research gap with empirical evidence in a way that could easily be understood by specialists and nonspecialists alike, they focused on constructing instruments to evaluate problem areas related to school failure. The original purpose/mission of the IEA was to research education achievement and its determinants, to test hypotheses relating to educational outcomes based on social and cultural contexts, and to establish a "science of empirical comparative education" (Gustafsson & Rosen, 2014; Husén, 1967a, 1979, p. 371).

The initial IEA meetings were moved from UNESCO's UIE in Paris to their offices in Eltham and Hamburg as their Paris offices were too constraining for the organization. The IEA coordinating center moved to Stockholm in 1969 in order to accommodate a growing staff at the same time as the UIE's interest in the IEA waned. Furthermore, the Swedish University Chancellor was able to offer free computer time on the Ministry of Defense mainframe computer to the IEA while the data cleaning, weighting, and analyses were housed at the University of Chicago. Eventually, part of the data processing needed to move to the United States as part of a US Office of Education (USOE) grant requirement, at which time Thorndike established a data processing unit at Teacher's College, Columbia University. In this arrangement, data were cleaned and weighted with initial descriptive statistics at Teacher's College then sent to Stockholm for data analysis (Husén & Postlethwaite, 1996).

The Pilot Twelve-Country Study

As the IEA began to gain momentum following the initial meetings in Hamburg and Eltham, the first major (and logical) step was to test whether or not an international assessment of the type proposed would be feasible. This included identifying logistical issues that would need to be addressed prior to launching the first official assessment, and creating an assessment that is contextually appropriate and yields reliable data. The IEA's first international assessment was the Pilot Twelve-Country Study (Pilot). The stated goal of the study was to test various factors related to achievement, specifically "to be able to test the degree of universality of certain

relationships which have been ascertained in one or two countries – for example, sex, home background, or urban-rural differences as related to achievement" (Husén, 1967a, p. 28). The group that had met in Hamburg to discuss the idea of the international assessment and form the IEA met three times between 1959 and 1961 in order to create the assessment, define the target student population, and plan the logistics of data collection and analysis (Foshay, Thorndike, Hotyat, Pidgeon, & Walker, 1962).

What sets the Pilot apart from the subsequent assessments is that it tested multiple areas (mathematics, reading comprehension, geography, science, and nonverbal ability), whereas the following assessments focused on single subjects until merging mathematics and science in 1995 (in the Third International Mathematics and Science Study).[3] The Pilot chose to focus on students aged thirteen years old, since that age group represented the final year that students would still be in school in all the participating countries, and ideally the samples were to be constructed based on students who were close to the national mean and standard deviation of achievement as understood in each country (Husén, 1967a).

The data were collected in 1960 and analyzed between 1961 and 1962, and from these analyses researchers presented the data in the form of national profiles. The main finding was that an international assessment of student achievement was feasible (Husén, 1967a). However, significant issues were recognized throughout the process, from the developmental stage to the analytical stages. The main issues related to data and expertise: Countries were not equally equipped with the skill set necessary to administer the test or to identify and conduct the appropriate sampling of schools and students, and the test items themselves had some translation issues. The key finding remains, however, that "what was most significant was that it proved that the project could be completed as planned" (Husén, 1967a, p. 29), so that the researchers at the IEA were encouraged to develop and launch the first formal international assessment of student achievement, the First International Mathematics Study.

First International Mathematics and Science Studies (FIMS and FISS)

As the Pilot demonstrated feasibility of such a large-scale international assessments of educational achievement, the IEA researchers were able to

[3] Even the Six Subject Survey, discussed on the following pages, broke the research into single-subject assessments rather than testing multiple subjects all during the same assessment time-frame.

turn their attention to a full scale, more complete study. The underlying sampling issues that were exposed during the Pilot were addressed as the IEA specifically contracted sampling experts in order to ensure the validity of their data. The First International Mathematics Study (FIMS) was developed in 1962 not to provide causal data related to academic achievement but instead to provide insight into how input factors (such as home environment, school procedures) relate to output measures (achievement); there was an underlying understanding among the IEA researchers that education was a social and political function, and FIMS sought to research how education responds to specific societal differences. In the development stages of FIMS, the researchers recognized and actively worked to address the limitations of the Pilot, which not only included the sampling errors and inconsistency with data collection but also included the need for a wider range of international actors to develop a mathematics assessment (Husén, 1967b, 1979; Schwille, 2011).

During this time, IEA remained a loose collaboration of researchers rather than a formal organization and it relied on the participating researchers and countries to raise the necessary funds for the studies, to identify hypotheses to study, to agree on and narrow these hypotheses into a manageable set of goals, and to collect and disseminate the data within their respective countries (Gustafsson & Rosen, 2014; Husén, 1979; Husén & Postlethwaite, 1996). In the era of FIMS, the purpose and studies of the IEA continued along a similar trajectory that revealed an emerging pattern of interest: While attempting to take account of how teaching and learning are influenced by developments in society, there was an interest in conducting longitudinal studies that were not feasible due to constraints of both time and resources.

The study was developed to be a scientific research project, rather than a simple statement of data; thus, even with limitations that prevented the IEA from developing a longitudinal study, the founding fathers improvised ways to maintain the integrity of the scientific study in order to produce data that would be an acceptable alternative. Thus, the researchers made the decision to test two different groups of students in four populations at a single point in time rather than developing a longitudinal study – an implicit cohort-longitudinal design.

The decision was made to focus on two terminal points in each educational system: the point at which nearly 100 percent of students of a particular age were still present in schools and the point immediately prior to university. The first terminal point group was associated with thirteen-year-olds, yet differences across the countries committed to participating in