Chapter 31
Implications of International Studies for National and Local Policy in Mathematics Education

John A. Dossey and Margaret L. Wu

Abstract This chapter examines large-scale comparative studies of mathematics education focussed on student achievement in an attempt to explain how such investigations influence the formation and implementation of policies affecting mathematics education. In doing so, we review the nature of comparative studies and policy research. Bennett's (1991) formulation of policy development and implementation is used in examining national reactions to the results of international studies. Focus is given to the degree to which mathematics educators and others have played major roles in determining related policy outcomes affecting curriculum and the development and interpretations of the assessment instruments and processes themselves.

International Studies of Mathematics Education

Writing in 1999, Martin Carnoy stated:

> The *quality* of national educational systems is increasingly being compared internationally. This has placed increased emphasis on mathematics and science curricula, English as a foreign language and communication skills. Testing and standards are part of a broader effort to increase accountability by *measuring* knowledge production and using such measures to assess education workers (teachers) and managers. Yet, the way testing is used to "improve quality" is heavily influenced by the *political* context and purposes of the evaluation system. Again, to develop effective policies for education improvement, the ideological–political content of a testing programme has to be clearly separated from its educational management content. (p. 16)

J. A. Dossey (✉)
Illinois State University, Normal, IL, USA
e-mail: jdossey@ilstu.edu

M. L. Wu
Victoria University, Melbourne, Victoria, Australia

M. A. (Ken) Clements et al. (Eds.), *Third International Handbook of Mathematics Education*, 1009
Springer International Handbooks of Education 27, DOI 10.1007/978-1-4614-4684-2_31,

Carnoy's insightful comments are as true today as they were in 1999. The appearance of large-scale international comparative studies of mathematics education, starting with those of the 1960s, have engendered three significant changes to the mathematics education landscape at national, state/provincial, and local levels. The first is an ever present reliance on large-scale sample survey data as a policy-making base for curricular and mathematics instruction decisions. The second is the heightened role occupied by nonmathematical organizations, governmental and nongovernmental, in the decision-making structure for what is important and what is not important in mathematics education programs, their design, and their implementation. The third is the use of comparative assessments as a lever for encouraging the convergence of curricular plans toward "national" or "global" models.

These changes have brought with them a reliance on values based in numerical indicators emanating from surveys, rather than from a body of mathematics education research based on a series of related research studies, be they quantitative or qualitative. Although expert panels who assisted in the design of these international assessments have involved knowledgeable mathematicians and mathematics educators, the resulting structure of the questionnaires and assessments used in the final collection of data has often then been modified to fit time, legal, or policy-based constraints which have distanced, in many cases, the data from classroom practice. Notwithstanding these disconnects, the global outcomes of international comparative studies of school mathematics have emerged as powerful arbiters of educational policy discussions of student competence, teacher quality, the path to school improvement, and the structure of schooling itself. At national, state/provincial, and local levels, assessment systems similar to the international assessments have been instituted by legislative acts as primary monitors of trends. Decision making and the institution of "educational crises" have become major media events stemming from the release of participant rankings in league-like tables of student achievement results or teacher qualifications.

The first major international comparative work in mathematics education was initiated by the International Commission on the Teaching of Mathematics during the 4 years following the organization's founding at the Fourth International Congress of Mathematicians in Rome in 1908. The study was created with the expressed purpose of conducting a comparative study on the methods and plans of teaching mathematics at the secondary and other levels of schooling. The study, spanning the years from 1908 through 1914, produced 187 volumes, containing 310 reports from eighteen countries (ICMI, 2011a). Excerpts based on data from the study can be found in the *Teaching of Arithmetic* and *Mathematics in the Elementary Schools of the United States* (Bidwell & Clason, 1970; Smith, 1909; United States Bureau of Education, 1911). In 1954, the Commissions' parent body, the International Mathematics Union, restructured and renamed the commission as the International Commission on Mathematics Instruction (ICMI). Along with the shift in the name, there was an implicit shift from the study of "the teaching of mathematics" to "mathematics instruction" in the activities of the organization (ICMI, 2011b).

In 1967, a *New York Times* article provided, in a manner similar to a sport's league standings table in a newspaper, the order of finish of national student achievement performances in the First International Mathematics Study (FIMS). This public release and the media's presentation focussing on standings signalled the emergence of a new way of examining and evaluating nation/state or provincial/local mathematics education programs. The influence of this approach to policy-building and blame-directing has only increased over time with the quadrennial release of data from the Trends in International Mathematics and Science Study (TIMSS) of the International Association for the Evaluation of Educational Achievement (IEA) and the triennial release of data from the Programme for International Student Assessment (PISA) developed by the Organisation for Economic Co-operation and Development (OECD). Mixed among the output from these massive studies are findings emanating from United Nations Educational, Scientific and Cultural Organisation (UNESCO), the World Bank, educational and economic think tanks, national assessments, doctoral dissertations, and education consortia and bureaus in individual countries.

These comparisons have grown over time to include comparisons of student achievement, teaching, teacher preparation, the context of mathematics education, and specialized topics included within or related to the mathematics curriculum such as problem solving, modelling, statistics, textbook contents, and information technology. A full discussion of all of these findings and their policy implications is beyond the scope of this chapter, which will focus on large-scale international studies of student achievement and the impact that they have had and continue to have on education policy in mathematics education.

International Association for the Evaluation of Educational Achievement (IEA)

The IEA was conceived in 1958 at an UNESCO meeting of sociologists, educational psychologists, and psychometricians. The IEA today, consists of a linked body of ministries of education and similar nationally-representative structures. Mathematically, the IEA became an important entity with the release of the findings of the First International Mathematics Study (FIMS) in 1967. This 12-nation study, based on data collected in 1964, focussed on 13-year-olds and students in the pre-university year of schooling. Policy relevant constructs emerging from the study were the importance of student opportunity-to-learn and equity issues as they affected academic performance. Other issues focussed on particular national differences in the education of teachers (Husén, 1967; IEA, 2011).

Seventeen years later, in 1981–1982, the IEA returned to mathematics assessment with the 20-nation Second International Mathematics Study (SIMS). This assessment featured a sharpened design based on a mathematics content framework and substantially more input from the mathematics community. The SIMS design featured pre- and post-measures about student opportunities to learn and perform in mathematics for 13-year-olds and students in the final year of secondary school. This study aroused increased interest in students' opportunity-to-learn, while heightening the key roles of curriculum and number of topics students are exposed to in a given year of study. In-depth questionnaires were used to probe teachers' coverage of key topics in the teaching of prealgebra at the middle-school level and content in precalculus and calculus at the end of secondary school (Burstein, 1993; McKnight et al., 1987; Robitaille & Garden, 1989; Travers & Westbury, 1989).

In 1995, the IEA returned to mathematics with the Third International Mathematics and Science Study (TIMSS). This time, the study cohort contained 45 countries, and the focal populations included 9- and 13-year-olds, as well as students in the final year of secondary education. In addition to focussing on major in- and out-of-school determinants of educational outcomes of schooling, TIMSS also conducted a special substudy comparing the mathematics curricula in the countries participating. The careful design and implementation of the design for the TIMSS 1995 study has provided an anchor for the subsequent IEA cycle of trend studies in mathematics, science, and reading. These ongoing data collections also highlight the semi-permanent status of such studies (now reconceived and renamed under the same acronym: the Trends in International Mathematics and Science Studies). The first assessments in the new formulation of TIMSS were carried out in 1999, 2003, and 2007. In 2011, more than 60 countries and jurisdictions participated in TIMSS 2011. Results from these studies are available online at the TIMSS study centre (http://timss.bc.edu/) and in a series of research monographs (Robitaille & Beaton, 2002; Schmidt, McKnight, Cogan, Jakwerth, & Houang, 1999; Schmidt, McKnight, Valverde, Houange, & Wiley, 1997; Schmidt, McKnight, & Raizen, 1997; Schmidt et al., 1996, 2001; Valverde, Bianchi, Wolfe, Schmidt, & Houang, 2002).

Results from TIMSS increased interest in the teacher preparation policies and practices around the world. At writing, the IEA is involved in the Teacher Education and Development Study in Mathematics (TEDS-M). This IEA study is focussed on how teacher preparation policies, programs, and practices contribute to the capability of teachers to teach mathematics in elementary and lower secondary schools (Grades 4 and 8). The framework, data, and findings from this study are available at the TEDS-M study centre (http://teds.educ.msu.edu/).

Programme for International Student Assessment (PISA)

In 1997, the Organisation for Economic Co-operation and Development, a group of democratic countries sharing economic-related information, decided to initiate a program of literacy assessments for 15-year-olds in the domains of mathematics literacy, science literacy, and reading in the mother tongue. PISA conducted its first survey in 2000, with subsequent surveys following in a triennial cyclic pattern, with the three domains rotating in their degree of overall emphasis within each passing assessment cycle. As a result, mathematics was the major focus of the assessment in 2003, in 2012, and will again be slated for 2021. In the intervening assessment cycles, mathematics is assessed only for trend reporting, with one of the other domains taking the role of primary focus. PISA also includes measures of general or cross-curricular competencies such as problem solving, measured in 2003 and 2012, and financial literacy measured in 2012.

Unlike TIMSS's focus on curricular-based knowledge, PISA focusses on measuring students' mathematical literacy, envisioned as students' ability to apply mathematical knowledge and skills and their developed capabilities to analyze, reason, and communicate effectively as they examine, interpret, and solve problems in contextualized settings. In PISA 2009, 34 OECD member countries and 41 partner countries participated. PISA is the only international education survey to measure the knowledge and skills of 15-year-olds, an age at which students in most countries are nearing the end of their compulsory time in school. Although PISA's results provide a picture of students' capabilities, they provide less direct relationships to the schooling students have received. At the same time, they may provide a better picture of the future capabilities of nations' students to cope with everyday applications of mathematics and science. These results allow countries and economies to compare best practices and to further develop their own improvements—ones appropriate for their school systems (McGaw, 2008; OECD, 2004a, 2004b).

In addition to this difference in aim, the PISA governing board is made up of representatives of national governments or members of their national ministries of education. Although some of these individuals are researchers, many are policy and legislative leaders with responsibilities related to reporting on the output of their nation's schools and status of the implementation of the approved curricula for mathematics.

Both the TIMSS and PISA assessments have had their share of proponents and detractors from within and outside of the educational world. From the foci of the assessments' content and the publication of the assessment frameworks to what students are expected to do in responding to the items and finally to the statistical analysis and reporting of the data, the studies have created a great amount of interest in student learning, performance, and achievement (Hopmann & Brinek, 2007; Kang, 2009; Kilpatrick, 2009; Murphy, 2010; Prais, 2003; Sjøberg, 2007). Supporting this interest, countries and professional societies have released special national studies, and the contractors carrying out the assessments have provided released items and other sample materials available along with other supporting documentation (Kilpatrick, 2009; materials on OECD/PISA Web site: http://www.pisa.oecd.org and on the TIMSS Web site: http://timss.bc.edu).

Other International Assessments of Mathematics Education

The Educational Testing Service (Lapointe, Mead, & Askew, 1992; Lapointe, Mead, & Phillips, 1988) conducted the International Assessment of Educational Progress (IAEP) with 13-year-olds in 1988, and 9- and 13-year-olds in 1991 with an expressed purpose of comparing participating countries' performances with that of US states through a statistical linking of the National Assessment of Educational Progress (NAEP) items common to NAEP and IAEP. This analysis showed wide variation in the performance of US states, with some performing statistically as well as the Asian nations, whereas others performed at the level of developing countries (Pashley & Phillips, 1993). Similar results were found in a special follow-up study to the TIMSS test conducted in 1998 (Kimmelman et al., 1999; Mullis et al., 2001).

Another international comparative education project was the Kassel project. It was initiated in 1993 by England, Germany, and Scotland, and later joined by Australia, Brazil, the Czech Republic, Finland, Greece, Holland, Hungary, Japan, Norway, Poland, Russia, Singapore, Thailand, Ukraine, and the USA. This project is focussed on collecting longitudinal samples of pupil work from the participating countries. As such, it differs from the preceding studies in that it focusses on individual student work over time rather than cross-sectional samples of student work. The analysis of the growth trajectories in these students is then used to ferret out key factors that lead to successful progress in mathematics within the participating countries (Blum & Kaiser, 2004; Burghes, Kaur, & Thompson, 2004).

Three other international comparative studies of note are the International Project on Mathematical Attainment (IPMA) study, the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) studies, and the First International Comparative Study of Language and Mathematics in Latin America.

The IPMA study focussed on student progress from the first year of primary school through secondary school, with data collected concerning student achievement, methods of teaching, resources available to teachers and students, and the nature of the curriculum studied. Countries participating for all or part of the study were Brazil, China, the Czech Republic, England, Estonia, Finland, Greece, Hungary, Ireland, Japan, Poland, Russian, Singapore, South Africa, Ukraine, the USA, and Vietnam. Reports from the study are available at (http://www.cimt.plymouth.ac.uk) and through a summary volume (Burghes, Geach, & Roddick, 2004; IPMA, 2011).

The SACMEQ series of studies report on student performance in reading and mathematics. The sponsoring organization consists of a consortium of the ministries of education from Botswana, Kenya, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Tanzania (mainland and Zanzibar), Uganda, Zambia, and Zimbabwe. Starting in 1995, there have been three cycles of assessment, with individual nation reports of recommendations derived from an overall data set representative of the member nations. The SACMEQ results report on the achievement of Standard 6 students (12–14 years of age). Cycle 1 reports were released in 2001, Cycle II reports in 2005, and Cycle III reports in 2010. These reports are available at the consortium Web site at (http://www.sacmeq.org/index.htm). SACMEQ began with support from UNESCO and has grown into a self-sufficient organization through joint support and the development of internal capacity (Greaney & Kelleghan, 2008; SACMEQ, 2011).

The First International Comparative Study of Language and Mathematics in Latin America was a project of the Latin American Laboratory for Assessment of the Quality of Education (LLECE) and involved a consortium of nations consisting of Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, The Dominican Republic, Honduras, Mexico, Paraguay, Peru, and Venezuela. The ministries of education of these Latin American and Caribbean nations were brought together in 1994 through the coordinating efforts of the UNESCO Regional Office for Latin

America and the Caribbean to develop a study focussed on information on students' achievements and associated factors that would be useful in establishing and implementing education policies within countries. The OECD has assisted UNESCO in the actual collection, analysis, interpretation, and reporting of the LLECE data. The focal content areas for the assessment were the language and mathematics knowledge and skills of third and fourth graders in the participating countries. In addition, information on a significant number of background and contextual variables was obtained from the schools and students (Casassus, Froemel, Palafox, & Cusato, 1998; LLECE, 2001). Information on the study and reports can be found at the consortium's Web site at http://www.llece.org/public/content/view/8/3/lang,en.

International Studies and Educational Policy

Reach of Educational Policy

Comparative international educational research in its purest form involves empirical work aimed at the revision of existing theories of the relationships within or between educational systems or between variables describing educational systems and economic indices or demographic data (Carnoy, 2006). If this is the case, how do the international comparative assessments of mathematics education fit this model? One might argue that their purpose is to describe student achievement at national levels. However, such a response would be short sighted. In reality, their purpose appears to be the creation of a platform for illustrating and relating students' achievement to salient policy variables such as distribution of achievement across racial and cultural groups, the relative performance of different genders in mathematical situations, the distribution of resources and teachers across geographical units, the relationships between the flow of students through the academic mathematics pipeline, and the relationship of various levels of output to national needs and labor projections. Within education, the output of such studies is of direct interest to curriculum experts, teacher educators, and those involved in professional development programs, and textbook writers and publishers of mathematical learning materials, as in Kilpatrick, Mesa, and Sloane (2007). Other interested consumers include governmental and policy experts, parents, and the public in general. As such, the results of national and educational comparative studies is a huge lever for those involved in educational policy, especially those interested in educational reform (Kellaghan, Greaney, & Murray, 2009).

TIMSS and PISA Assessment Frameworks

Given the role that the IEA and PISA results play in serving as levers in international and national discussions of educational policy, one might examine their geneses and stated purposes. IEA studies result from a cooperative group of research bodies, some of which are governmental and some not. In either case, the bodies are research oriented first and policy oriented second. The TIMSS studies are closely linked to instructional processes in classrooms and the curricula of the participating nations. The mission of the TIMSS assessments is

> to provide high quality information on student achievement outcomes and on the educational contexts in which students achieve. … [In doing so, TIMSS is dedicated] to providing countries with information to improve teaching and learning in these curriculum areas. Conducted every four years on a regular cycle, TIMSS assesses achievement in mathematics and science at the fourth and eighth grades. The achievement data are collected together with extensive background information about the availability of school resources and the quality of curriculum and instruction. (Mullis et al., 2009, pp. 2, 7)

The mission statement for TIMSS places learning outcomes, teaching, and learning contexts at the forefront, with an implied goal of linking achievement to curricula and instructional practices.

PISA studies, on the other hand, assess how well 15-year-old students are prepared to deal with contextualized situations where mathematics might provide assistance in finding resolutions. PISA refers to this capability as *mathematical literacy* and defines it as follows:

> An individual's capacity to formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts, and tools to describe, explain, and predict phenomena. It assists individuals to recognise the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens. (OECD, 2010c, p. 4)

Although PISA does not reject curricular links in developing students' literacy, the assessment's primary purpose is the determination and description of students' capabilities to formulate, implement, and solve mathematical problems.

The linking of TIMSS to teaching and learning and PISA to literacy does not say that they ignore the other's main focus. To do so would be a denial of the intrinsic link between the two goals and the huge overlap of outcomes that are examined by both programs. Many of the curricular and instructional research objectives in TIMSS are driven by policy considerations, and many policy objectives in PISA result in research themes linking PISA findings directly to school programs. In fact, over the past decade, the two large-scale assessment programs have moved toward one another in goals and in the nature of the items used in their assessments. In addition, their role in policy decisions also increased as nations, states and provinces, and local school districts have looked for guidance in forming curricular plans and selecting instructional approaches and materials.

Role of International Studies in Shaping Policy

The IEA international mathematics assessments came of age in the 1980s just in time to fill the increased desire within UNESCO and, later, within OECD for a set of indicators of student performance. Indicators, viewed as variables taking on values which describe inputs, processes, or outputs from the educational enterprise of some defined country or defined grouping of countries provided a way of quantifying education. Over time, such indicators became the source of policy, and at the same time, their values provided another lever for policy change. This recasting of indicators in quantifiable form further spread the influence of indicator systems, especially those that had linked assessments, as a source of policy initiatives.

Reports portraying indicators from such studies have fuelled governmental and nongovernmental agency reports on educational outcomes for the past 50 years. The OECD indicators had their birth in the OECD's International Indicators of Education Systems project (INES) in the late 1980s, a movement that coincided with a shift from research-based assessments of student performance within a nation to national assessments of educational progress. This shift was very evident with the maturing National Assessment of Educational Progress (NAEP) in the USA and the initiation of similar, but newer, assessment programs in New Zealand, Portugal, Spain, Sweden, and the UK. Central to the growth for the demand for data on education was the UNESCO (1990) World Declaration on Education for All recognizing education as a human right and relating it to the physical and economic health of nations. Its foci on learning, equity, and supportive environments and resources for education promoted the need for more policy-based items as part of the background and demographic sections of national and international assessments (Moskowitz, & Stephens, 2004; Rutkowski, 2008: UNESCO, 2011).

In 1991, OECD began the publication of annual indices of indicators in its *Education at a Glance* series. This provided easy reference for policy analysts to countries' profiles, as well as their comparative performances relative to other countries. The indicators and supporting data exhibit a wide range of outcomes discussing student performance when parental education, social-economic status, and other factors are considered and when national performances are adjusted for national economic indicators (OECD, 2010c).

At the same time, the influence of PISA was growing within OECD nations. Several non-member OECD nations participated in the 2003, 2006, 2009, and 2012 PISA cycles. This participation multiplied the influence PISA indicators have had on just the member states by including another group of developed nations and an even larger group of developing nations. These indicators do not just inform the leaders of these countries, they assist in the framing of policies and the direction of reforms. Nóvoa and Yariv-Masal (2003) noted:

> Such researches produce a set of conclusions, definitions of "good" or "bad" educational systems, and required solutions. Moreover, the mass media are keen to diffuse the results of these studies, in such a manner that reinforces a need for urgent decisions, following lines of action that seem undisputed and uncontested, largely due to the fact that they have been internationally asserted. (p. 424)

Results from PISA 2000 and 2003 indicators supported the development of national goals for secondary-school curricula in Flemish-speaking regions of Belgium, strengthening mathematics program implementation by increasing the numbers of secondary-school mathematics advisors in New Zealand, and allocating more resources for the mathematical and science education of prospective primary school teachers in Sweden (Owen, Stephens, Moskowitz, & Gil, 2004). When the favourable results of PISA were announced in Finland, rather than reflect positively on them, the Finnish government moved to harmonize the system and to allocate more time to core subjects, potentially removing some of the advantage Finnish schools might be providing their students (Välijärvi, Linnakylä, Kupari, Reinikainen, & Arffman, 2002). But, perhaps the most notable effect was the reception of the PISA 2000 results in Germany. The below median performance in overall literacy ranking stunned the nation, and German educational authorities called for urgent reform measures to right the ship and get Germany back on course. This outcry for reform was based on the argument that the PISA test measures outcomes associated with the emerging world and its workplace requirements (Ertl, 2006). Again, a governmental reaction, in this case was based partially on economics and partially in shock.

Parallel to the growth of indicators and assessments, national educational ministry personnel worldwide were becoming active in the administration, development, analysis, and reporting of findings of their own national assessments and national reports of TIMSS and PISA. This shift of ministry officials into becoming players rather than consumers was aimed more at policy issues than research issues. In several countries, the shift was also accompanied by a shifting from curriculum questions and research-oriented issues to sharpening background assessments to answer other less-curricular-centred issues that were more pressing nationally from a policy standpoint. This movement, in some cases, weakened the focus on content alone in favour of more general policy-centred questions within assessments. Within Europe, first, and then on a broader stage, concern began to arise about the shifting use of indicators from being a focal point for understanding the educational enterprise to potentially being used to shape and control the educational systems of nations from a normative standpoint (Grek et al., 2011; Lester, Dossey, & Lindquist, 2007).

With the advent of multinational industries and the easy international exchange of knowledge and data, there has arisen a demand for evidence-based research findings to quantify the adequacy of state or provincial and national educational systems. This demand is a side product of the advent of reform programs in mathematics based on standards (Chatterji, 2002; Steiner-Khamsi, 2006, 2007). Further, the demand for information in the form of comparative data has grown to the point where data resulting from the output of large-scale comparative studies can be viewed as forming international knowledge banks which enter into the processes of borrowing and lending. This view of the outcome data serving as a knowledge bank was first broached at an educational meeting at the World Bank in 1996 (Eaton & Kortum, 1996; Jones, 2004; Jones & Coleman, 2005).

The rise of such international knowledge banks has brought with it two new major policy influences. The establishment of international means and indices for outcomes has resulted in funder/donor pressures to move country means above international averages on targeted indices. This has been especially true in developing countries. In many cases, such targeted indices are only marginally related to national or local goals or to the culture of a given country. Such pressures are most prevalent in developing countries, where ministries themselves find continued funding associated with a "harmonization" of their programs and expectations to share the funders' knowledge and approaches and to work together with partner countries to "converge" and improve their programs (World Bank, 2005). With this pressure from large donors, financial lenders, or policy groups such as OECD comes a change in their behaviour. They begin to make a shift from being a lender of capital to becoming a lender of educational policy.

The second shift related to policy is directly tied to the emergence of the large knowledge banks of studies such as those associated with the IEA and OECD studies. This shift comes from within the affected countries themselves. As particular indices are seen as being associated with positive movement and successful transformation of curricula, national and local politicians and policy makers use the indices of the knowledge banks as fulcrums for change. Politicians and policy makers turn to the existence of such study-based indices as external justification for the policy points they are promoting as needed changes in their national or local programs (Cussó & D'Amico, 2005; Grek, 2009; Peters, 2002; Phillips & Ochs, 2003).

Luhmann (1990) and Schriewer (1990) argued that the very existence of rankings provided by assessments such as the IEA and OECD studies provide a perceived base of scientific rationality for policy proposals and their public explanation. In fact, it has been argued that this very perception of the large-scale studies answering questions about curriculum has led to the lack of other research on curriculum reforms (Vithal, Adler, & Keitel, 2005). The use of indices as a basis for monitoring and leveraging change in countries, especially lower-performing nations, often leads to the declaration of crises and the increase in educational policy "borrowing" from league-leading nations. Such adoption of other countries' policies is made without careful consideration of the internal system supports which have made the policy successful, the cultural differences between the programs of the lending and borrowing nations' educational programs, and the impact such changes will make on the internal coherence of the curriculum of the "borrowing" nation (Nguyen, Elliott, Terlouw, & Pilot, 2009; Phillips & Ochs, 2003; Ripley, 2011; Thomas, 2001).

Policy Convergence

The concept of policy convergence was first introduced by Kerr (1983) as "the tendency of societies to grow, more alike, to develop similarities in structures, processes and performance" (p. 3). Over time, many have noted this tendency and attributed it to a number of causes. Bennett (1991) examined the topic at a level less general than "societal convergence" in his examination of policy convergence. He claimed that policy convergence should be examined as a movement from varied positions to a common point over time. In examining the forces that lead to convergence, Bennett (1991) posited a taxonomy of four processes that result in policy convergence in times of change:

1. *Emulation.* This approach to convergence involves the utilization of evidence about another's programs to modify one's own programs. As it is the adoption of a blueprint, emulation can explain some policy changes, but not outcomes themselves.
2. *Elite networking and policy communities.* This form of policy convergence is based in the actions of a transnational group of policy makers sharing a common focus on a policy issue. Unlike emulation, there is shared engagement in working on and adopting similar policies. There may even be a group charged with discussing a set of issues around the topic central to an emergent policy; in other cases, such groups may be self-appointed.
3. *Harmonization.* This approach to convergence of policy involves interdependence of the policy-making bodies and the existence of a super-body responsible for shaping and monitoring the common policy. However, the harmonization provides a movement together without the need for external controls or oversight. The European Community (EC) and OECD were held up as examples of such linked policy-making bodies. Harmonization requires a balance of relinquished autonomy with a hope for a gain in unproductive diversity in cross-national policies.
4. *Penetration.* Convergence through penetration occurs when the policy-adopting bodies are forced to implement an externally developed policy. Examples of penetration exist when nations are forced to implement an international standard or be closed out of a market. In some cases, this may be the result of harmonization strengthened to a regulatory system that defines who can participate or benefit in a given market of human activity—telecommunications, intercontinental aviation, and measurement standards.

It is the latter, and more coercive, types of convergence that are causing concern among educational policy experts. External loan institutions (e.g., UNESCO, World Bank, International Monetary Fund) have the leverage of expected improved outcomes for continued funding. Internal politicians and policy makers have the leverage of the public press to achieve convergence through public opinion and political power. Such uses of IEA and OECD data are being questioned in many quarters of the comparative education and mathematics education communities (Alexiadou, 2007; Carnoy, 2006; Grek, 2009). The results of research have also raised doubts about whether curricular convergence-focussed activities result in increased outputs (Grier & Grier, 2007; Mayer-Foulkes, 2010). Studies of interventions aimed at convergence have often shown that although convergence to the mean occurred with several variables, significant, unwanted, and unexpected increases occurred in the variance of both the focus and related variables. Such patterns could be very counterproductive in educational settings struggling to improve across the board.

Impact of Educational Policy in Different Countries

An examination of national reactions to the release of IEA or OECD data presents an opportunity to study the actual impact of large-scale international comparative studies on mathematics education programs at a national level. Reactions to the release need to be monitored from public, media, and policy levels, as the degree of knowledge and potential leverage differ greatly among these bodies as one moves from nation to nation.

In high-performing countries, at least as characterized by the study league-tables, the reaction has often been one of satisfaction, raised even to the point of self-congratulation. This reaction is often accompanied by reference to performance on particular indices comparing their performance with that of other countries. In low-performing countries, there are public calls for reform, which often take one of two forms. One is a call for a return to the basics; the other is a call for changes leading to harmonizing the national program with that of other countries having higher league-values in indices of comparison that are viewed as desirable. In other instances, the release in lower-performing countries is a governmental one calling for change in policy with specific reference to a greater federal role. Sometimes, action has been called for to fold the perceived needs into supporting an even broader political agenda involving governmental roles and the roles of public–private education within the country. A third reaction is one of indifference, suggesting that the results are just one way in which one could evaluate the outcomes of the national system. Such reactions might result in no action, the institution of a study to look into the results more deeply, or starting a small-scale study or group of projects examining alternatives without a great deal of fanfare.

Mini-Case Studies of Policy Influence of International Studies on Mathematics Education

The following mini-case studies of national performance at the Grade 8 (13-year-old) levels of the IEA studies and the PISA 15-year-old literacy studies present brief histories of the reactions and policy decisions surrounding the release of results from these large-scale international comparative studies. Occasional comments will be made concerning issues tied to either Grade 4 or 12 aspects of the IEA program. The first country examined is the USA, a country where changes have traversed the full span of Bennett's levels of convergence because of reactions to performance in international studies and recommendations from governmental studies and professional organizations. Next, Germany and Finland are examined for the differences they experienced in student results and reactions to public opinions. Finally, some comments are made concerning Singapore and past and present movements in mathematics education there.

USA

Perhaps the region having the most public reactions to study releases has been the USA, where each national and international assessment has had the same level of reaction in the media as a major sporting event. These reactions have even triggered national commissions whose reports have also created waves of interest and policy-related actions.

The USA has participated in every IEA mathematics study at some level. Our study of reactions to large-scale international assessments in the USA is presented in three phases: reaction to the 1967 release of FIMS, reaction to the 1987 release of SIMS, and reaction to the 1995 release of TIMSS and successive releases of IEA Trend studies and to OECD PISA studies. The USA performed significantly below the IEA average in FIMS and SIMS, at the international average in 1995, and above the international average in 1999, 2003, and 2007 (Beaton et al., 1996; Husén, 1967; Mullis et al., 2000, 2004, 2008; Robitaille & Garden, 1989).

In the OECD PISA studies of 15-year-olds' performances, the USA performed no differently from the PISA mean in the 2000 assessment and then significantly below the international mean in the 2003, 2006, and 2009 assessments (OECD, 2001, 2004a, 2007, 2010a).

Phase 1. Reactions to the 1967 release of the first IEA mathematics study (FIMS) began with the *New York Times* coverage. Although the notion of international comparative studies was new to the public, curricular and instructional scholars, and policy practitioners, each group immediately saw the data and findings as potential policy levers.

The USA was, at this time, about 10 years into the development of new curricular programs, which originated in the mid-1950s because of a perceived lack of skills of students entering scientific, technology, engineering, and mathematics study at the collegiate level. The most notable of these was the School Mathematics Study Group (SMSG). Although the initial impetus for these new curricular programs had been the unpreparedness of entering university students in the broad sciences, the 1957 launching of the Sputnik satellite by the Soviet Union was quickly given the credit for their creation. Financial support for these programs, and others, was provided by the National Science Foundation (NSF) and materials were quickly brought to field tests in the schools of the nation (NCTM, 1961, 1964).

These programs, which were backed by many in mathematics education and the mathematics community, focussed on the development of new textbook series and supporting materials. School mathematics was to have a greater focus on its underlying structure and the relationship between this structure and the algorithms that had dominated the content of the traditional programs. Instructionally, there was a shift from teacher presentation to an approach making greater use of guided discovery and manipulative materials to illustrate and motivate mathematics learning. Paralleling this work, the projects, universities, and school districts instituted a number of professional development projects for teachers aimed at strengthening their understanding of and capabilities to teach the newer curricula. Parents were also factored into the change equation with workshops held in conjunction with school parent–teacher organizations. The press labelled the entire reform effort the *new math*.

National reports were issued by the College Entrance Examination Board (CEEB—later to become the College Board) and a group of mathematicians and mathematics educators looking into the future. The CEEB (1959) Report of the Commission on Mathematics presented a review of secondary mathematics programs and made a call for mathematics for all students before turning to its main point—revising the secondary program for college-capable students. In particular, the report provided a call for a balanced treatment of concepts and skills with a stress on deductive reasoning throughout the secondary-school program. It also suggested attention be given to structure, use of sets and functions as a unifying feature, combined with a functional approach to trigonometry (Jones & Coxford, 1970).

Once the overall program of reform was well underway, the movement was not without its critics. Foremost among these was Morris Kline (1961) of New York University and a list of other mathematicians ("On the mathematics curriculum of the high school," 1962). This group of mathematicians was concerned about the undue emphasis on mathematical structure in the reforms, the lack of ties to the real world, and the lack of reasons for studying mathematics beyond mathematics itself. Another line of attack came from the Executive Director of the National Association of Secondary School Principals (P. Elicker, personal communication, January 23, 1962), calling for cutting off of federal funds for SMSG, as it was creating a national curriculum which would usurp the state and local rights to educational policy. This trickle of dissatisfaction from some vocal voices in the mathematics community, coupled with the voices of teachers stressed by dealing with new curricula, and joined by the dissatisfaction of parents unable to assist their children with homework viewed as unfamiliar and abstract, set the stage for change.

The release of the results from IEA FIMS was accompanied by a headline on the first page of the March 12, 1967, *New York Times* which read: "United States Gets Low Marks in Math." This headline, and the accompanying report that the USA had finished 11th out of 12 at the 13-year-old level and 13th out of 13 at the final year secondary-school level, cast a significant blow to the reform movement in school mathematics in the USA. The trickle of dissatisfaction turned to a torrent, with critics pointing to a downplaying of the "basics" or arithmetic facts and computational algorithm proficiency as the culprit. A crisis was proclaimed, the FIMS results served as the lever, and the result was a backlash against curricular reform in school mathematics.

Jeremy Kilpatrick (1971) presented a thoughtful analysis of the FIMS study, noting especially the tradeoffs that a researcher makes in moving from very small samples to a large sample where the notion of the context of student learning is lost. Students' opportunity-to-learn stood out as a salient, researchable topic to pursue in secondary and follow-up studies. It was clear that US students had far less exposure to advanced topics, and more review of previously studied topics, at both the 13-year-old level and at the final year of the secondary-school level. Further, student performance on advanced topics indicated that they were potentially teachable and learnable at the levels where assessments were given.

The following decade brought work on redefining the basics, based in many cases on curricula from other countries. This work resulted in a gradual expectation for a greater focus on algebraic and geometric content in the middle school. More importantly, it led to the NCTM developing, with wide feedback, its 1980 *An Agenda for Action* which laid out a new broad conception of the basics in school mathematics and moved problem solving to a pre-eminent position in the curriculum. The *Agenda* endorsed appropriate uses of technology in the curriculum and recommended that assessment of students be expanded beyond the traditional algorithmic-based approaches. Further, the *Agenda* called on teachers to exhibit greater levels of efficiency and effectiveness in their instruction.

Employing Bennett's (1991) model for convergence, one might indicate that the period from 1967 to 1980 was a period of reflection and emulation. Although there were smaller cycles of focus on manipulatives and the appearance of hand calculators during the interval, the policy focus was on defining a new way forward in school mathematics based on looking at others, learning from the first findings of the fledgling National Assessment of Educational Progress which released its first findings in the early 1970s, and developing a more policy-oriented outlook in the mathematics and mathematics education community. As there was no national department of education in the federal government at this point, the focus was on opening a conversation and providing a model, the *Agenda*, that the profession could examine and debate. The emergence of the professional community and its contacts at the first international mathematics education congresses with leaders from other nations and the emergence of the research community in mathematics education during this same period began to lead toward the formation of elite networks of policy-minded individuals in the mathematics and mathematics education communities.

Phase 2. The *Agenda* ushered in a decade of work which ultimately resulted in the development and release of NCTM's *Curriculum and Evaluation Standards for School Mathematics* in the spring of 1989 (McLeod, Stake, Schappelle, Mellissinos, & Gierl, 1996). Across the 1980s, prior to and immediately following the release of the influential *Nation at Risk* report (National Commission on Excellence in Education, 1983) calling for reform in US education, the NCTM, along with other major mathematical groups, had been moving toward drafting a statement of what students should know and be able to do as a result of their mathematics education. This process was guided by an emerging group that Bennett (1991) would term an *elite*. Formed by educationally-oriented members of the mathematics community and leaders of the National Council of Teachers of Mathematics, this group worked to form a community of teachers, researchers, and scholars fuelled by the notion of improving school mathematics and making the reform stick.

The report of the 1986 NAEP, *The Mathematics Report Card: Are We Measuring Up?* (Dossey, Mullis, Lindquist, & Chambers, 1988), noted growth in students' mathematics achievement since previous NAEP assessments. But the report also noted that students were frequently unable to work straightforward problems involving concepts of which they should have full command at their grade level. Since this was the first report of US student achievement after the *Nation at Risk* report and the release of the results of the 19-nation IEA's Second International Mathematics Study (SIMS) in the January 1987 publication of *The Underachieving Curriculum* (McKnight et al., 1987), the nation could have shifted immediately into a crisis mode (K. J. Travers, personal communication, July 4, 2011).

However, the US mathematics community and mathematics education community, in conjunction with the National Research Council, had formed the Mathematical Sciences Education Board (MSEB) in 1985 to coordinate the nation's response to the underperformance in mathematics education. The MSEB was structured to be broadly representative of the mathematics community from elementary school teachers to distinguished university professors, representatives of state and local boards of education, employers from the scientific and technological sectors, and representatives of teacher, parent, and policy groups. In January 1989, the MSEB released *Everybody Counts,* which set the stage for what US mathematics education programs needed to do, based on research and comparative studies, to reach the goal of mathematics for all and the goal of an increased flow of qualified students at all levels along the mathematics pipeline. This document served as a policy precursor for the release of the NCTM *Curriculum and Evaluation Standards for School Mathematics* in March 1989. This release was met with positive comments and a lack of crisis focus. States signed on to the standards, and within three years all but a few states had changed their curricular frameworks to parallel the recommendations of the NCTM standards.

This was a tremendous step forward for mathematics education policy in the USA. Although the nation now had a Department of Education, direction of schools was still vested in the state departments of education, which, to a large degree, abrogated their responsibilities for curriculum to the leaders of over 15,000 separate school systems spread across the country. This vast and dispersed responsibility for mathematics education at the local level has been a major and defining feature of US mathematics education. The appearance of the *Curriculum and Evaluation Standards* (NCTM, 1989), and the year-long public vetting of the draft with special attention paid to state departments of education, led to 46 states setting or modifying, within three years, their written state curricula to parallel the recommendations of the final standards document. Further, professional development materials and training sessions were provided to educate leaders to talk about the standards and work with state and local school districts in implementing the standards at the local level. With this effort and the formulation of the MSEB and its work with NCTM, the policy community focussed on convergence of the mathematics education curriculum and attempts at convergence moved to the harmonisation level of Bennett's (1991) taxonomy. Leaders of the mathematics, mathematics education, and policy communities met regularly to shape and monitor activities aimed at strengthening US school mathematics. Although not everyone supported the standards-based movement, there was focussed change afoot.

Phase 3. Subsequent releases of the IEA documents from TIMSS and from the IEA Trends studies in 1999, 2003, and 2007 were viewed as signals of distances to go, but not as imminent crises. The same could be said for the release of the OECD

PISA studies of 2000, 2003, 2006, and 2009. The scores show consistent underperformance at the OECD level. Although each release was met with media proclamations which spoke of doom and despair over the state of education, especially in reading, mathematics, or science, little direct action was taken at the local level. Teachers were involved in professional development, and the updating of the NCTM standards with recommendations shaped ever closer to grade/age-level expectations appeared as the *Principles and Standards for School Mathematics* (NCTM, 2000).

In Washington, DC, the situation was different. With the change of administrations in 2000, President George Bush pushed for and won legislative approval for his *No Child Left Behind* (NCLB) law that created a mandatory national testing program which held schools accountable for achieving specific and increasing levels of performance. Those levels were keyed to a new NAEP framework for mathematics that called for increased focus on algorithmic skills and a lessening of attention to measurement, geometry, and probability as targets for the NAEP assessments. In addition, the legislation moved the NAEP to an annual testing program for all students in Grades 3 through 8 and at one level in secondary school. The law further instituted a requirement that all states ensure that the schools under their aegis bring their students up to the "proficient" level of performance by 2014 (Olson, 2004). Intervals defining *below basic, basic, proficient,* and *advanced* levels of performance were defined psychometrically via achievement-level-setting procedures working with the individual NAEP items, student percentages, and Item Response Theory (IRT) parameter information (Pellegrino, Jones, & Mitchell, 1999). This focus on accountability by achievement levels had been growing across the 1990s parallel to the implementation of the NCTM standards, but NCLB brought it front and centre.

With the institution of the NCLB law, the federal NAEP testing program, and its framework, one had the essence, at least, of penetration in the policy community. However, at the time of this writing, the lasting impact of this legislation and its punitive aspects for schools that fail to achieve raising their students to the *proficient* level by 2014 is uncertain, as legislative forces are afoot to change NCLB. The path to convergence that had its roots in the IEA release of the FIMS data, the growth of the policy community within the mathematics and mathematics education community through NCTM, MSEB, and the many state mathematics teacher groups, and the success of the standards showed a pattern of harmonization. However, the impact of the NCLB, the insertion of a NAEP assessment system not harmonized with the NCTM (1989, 2000) standards, but having punitive outcomes for noncompliance illustrates the power of the existence of policy groups which have the ability to force convergence through penetration.

At the time of this writing, US schools are working through another policy-induced change to the mathematics curriculum and state level assessment systems. In 2010, the Council of Chief State School Officers (CCSSO) and Achieve, an organization formed in 1996 by the state governors and corporate leaders and focussed on educational reform, released their *Common Core State Standards for School Mathematics* (CCSSM) (see CCSSO & NGA, 2010; Porter, McMaken, Hwang, &

Yang, 2011). This set of recommendations was immediately adopted by 40 states as their state-level standards for school mathematics for K-12 public schools. As such, the CCSSM provides the framework for expected mathematics outcomes and becomes the state-level proxies for meeting the NCLB goals for student progress toward proficiency. Although it is too early to judge the impact of this rapid insertion of new materials into the mathematics reform and policy mix in the USA, it clearly shows that convergence by insertion is the order of the day, with the impetus for structural change originating outside the professional mathematics education community. Time will tell the outcome of the NCTM (1989, 2000)-standards-led move to convergence of the K-12 mathematics education curriculum in the schools of the USA and the influence of the CCSSM movement on the trajectory the NCTM standards engendered.

Germany and Finland

Although Germany and Finland are close geographically, their experiences with the PISA assessments and policy reactions are quite dissimilar. Both countries place a high value on public education but toward different ends. Neither had been consistent participants in the TIMSS 13-year-old (eighth grade) level assessments from 1995 forward. Germany had an eighth-grade ranking of 23rd out of 41 countries in 1995, whereas Finland had an eighth-grade ranking of 14th out of 38 countries in 1999. In PISA, both countries participated in each assessment from 2000 forward. The countries' performances can be viewed in terms of place ranking out of the number of participating countries or by their PISA mathematical literacy score. Using this notation (ranking, literacy score), Germany's results for the four assessments were as follows: 2000 (21/41, 490), 2003 (19/40, 503), 2006 (17/48, 504), and 2009 (16/65, 513). Finland's results were: 2000 (5/41, 536), 2003 (2/40, 544), 2006 (1/48, 540), and 2009 (6/65, 565).

Germany. Germany's students' performances in 2000 through 2006 were met with public outcries, and the nation was caught up in rethinking its educational structure, what other factors might have influenced the scores, and a myriad of other possibilities (Miserable Noten fűr Dekutsche schűler, 2001; OECD, 2002; Stanat et al., 2002). Finland, on the other hand, had high performances, and its citizens were hardly aware of the PISA assessment program or their students' achievements. The different reactions are reflective of the countries' cultures and their approaches to educating their children. However, the policy reactions are somewhat surprising.

Prior to this time, Germany's education expectations were organized at the state level, with each state developing and monitoring school outcomes within their own *Länder*. The reaction was swift to the 2000 and 2003 findings. By 2003, there was a report outlining recommended standards and assessments by which these expectations would be monitored (Klieme et al., 2003). This report was passed through the Standing Committee of *Länder* Ministers in December 2003 and became the law of the land for implementation with the 2004–2005 school year. This was change, and unanimous change, at an unprecedented pace for German education. Unlike other reforms, the trade unions and businesses and industry quickly endorsed the changes as well (Ertl, 2006). As a result, new curricular guidelines and texts had to be developed and teachers provided with professional development relative to the implementation of the new goals. Individual states in Germany still had the authority to react to the strictures of the new standards in their own fashion. Educators in Germany felt that the changes within the mathematics curricular recommendations moved the curriculum closer to an empirical and practice-focussed conception than to the more didactical–cultural conception that had defined German education (Bohl, 2004). The conception of OECD literacy as an outcome was not central to German schooling prior to the reactions to the national PISA outcomes. This, combined with the notion of developing the competencies associated with the individual disciplines sampled by PISA, has furthered the stress in moving from traditional approaches to schooling (Sloane & Dilger, 2005).

The process and changes that resulted in the convergence observed in Germany was a significantly compressed version of that observed in the USA. In the USA, the transitions occurred over a period of 40–50 years in moving from the uncoordinated curricula of the early 1950s to the adoption of standards-based outcomes by the states in the late 1990s. In Germany, these transitions were compressed into little over a 4-year span. Given that many of the mathematics educators in Germany were well linked to others in the international mathematics education community and that the notions of *competencies* defining outcomes were part of the experience in Germany's neighbouring country of Denmark (Niss, 1999; OECD, 2003), clearly, communication was already in place between the leaders of the curricular areas in German education and other international policy players at the start of the period of reaction to the PISA results. However, the reflective convergence that usually accompanies change resulting from harmonization was sharply curtailed by the quick institution of new standards by the ministers of culture and education in 2003. Germany is a case where the *Länder* ministries and educational administrators were handed the new standards almost as a fait accompli to be inserted into a new nationwide mathematics curricular structure.

Not all sectors of the education establishment were happy with the decisions made by the ministers and the move to standards-based outcomes. Ertl (2006) noted that the

> Federal Ministry's post-PISA agenda seems to be firmly focused on raising national educational standards by pursuing measures that will improve Germany's low ranking in the PISA league table. It places less emphasis on the solution of the other major problems identified by PISA, the strong connections between the socio-economic background of students and their education achievement. (p. 630)

Finland. The situation in Finland, contrasted with that in Germany, shows another country where education was valued, but the philosophical view of the process was different. Finland did not participate in TIMSS 1995 but did participate in 1999, where their Grade 8 equivalent students performed significantly higher than the IEA average performance in mathematics for this level (Mullis et al., 2000).

Finland's performances in the OECD PISA assessments have been stellar, with its students attaining the highest non-Asian country performance in each of the PISA assessments from 2003 through 2009 (OECD, 2001, 2004a, 2007, 2010b).

Finland's student achievements in the TIMSS and OECD assessments have garnered considerable kudos in the international education community and the public press. This focus has brought attention to the differences in both curricular programs and quality of instructional staff found in Finnish schools. Many have asked what factors led to their consistently high achievement on the PISA mathematical literacy assessments. To answer that question, one can start with the fact that Finland has a National Board of Education (FNBE) which oversees the educational enterprise of the nation. Starting with the 1985 mathematical curricular framework, the FNBE started a movement away from the comprehensive school with a strong core curriculum in mathematics for Grades 1–9. Although the board still provided a framework with four mathematical strands (number concepts, expressions and equations, geometry, and applied mathematics), the focus shifted from an emphasis on basic concepts and structure to one emphasizing problem solving, applications, and everyday uses of mathematics. This change was accompanied by professional development for teachers on teaching through problem solving and the use of projects to involve students in using their mathematics to solve problems from everyday settings. Follow-up research indicated that this movement was a partial success, but it succeeded in moving teachers to teaching only about problem solving, not through problem solving.

To further aid teachers in the transition, the FNBE and the municipalities provided teachers with more professional development, publishers produced problem booklets keyed to grade levels, and special emphasis was given to Japanese-style "open-problems." This change moved the agenda on problem solving and realistic applications of mathematics further. The biggest change which might have affected the PISA results, was the release of a revised framework for mathematics by the FNBE in 1994. This action decentralized the curricular oversight by removing the listing of specific content and turned the task of developing the mathematics curriculum over to the local schools' teachers. The FNBE did provide guidance that teachers should still examine the traditional content critically and thin the curriculum of material that did not have any use in the further development of mathematics. The FNBE also stated that Grades 1–6 should master the basic concepts and be capable of performing calculations on paper, mentally, and through the use of a hand-calculator (Kupiainen & Pehkonen, 2008).

In 1999, Finnish education officials provided schools with a marking guide scaled from 4 (reject) to 10 (excellent) with advice to move students to at least the 8 (good) level. Although there is no national assessment used to place each student in an achievement level bracket for mathematics, Finland does have an assessment given to a representative sample of ninth graders. These papers are analyzed, published, and discussed. Further, individual schools can buy copies of these tests to be given locally and then compare their results, and marks, with those given on the national sample of tests. This information helps provide a degree of uniformity to outcomes at a national level. There is also an assessment given to a sample of sixth-graders every fifth year.

When Finnish educators reflect on what has enabled their system to perform so well, they cite the following factors: their comprehensive educational structure with heterogeneous grouping of students, the societal focus of the schools with free healthcare and cohesive group-focussed structure, the use of specialist teachers of mathematics at lower grades in many schools, the focus on equity and Co-operation rather than competition, the focus on problems and the use of mathematics, widespread student belief that they can solve problems, and the strong and supported corps of teachers (Kupiainen & Pehkonen, 2008; Malaty, 2006; Rautalin & Alasuutari, 2009; Sahlberg, 2010).

Teachers in Finland had a more advanced education than their peers in most countries, and this education is balanced between content knowledge and content-based pedagogical knowledge. This advanced preparation for their teaching and for the professional ways in which they approach the tasks confronting them has resulted in teaching being one of the most respected careers in Finland. This confidence in teachers as a whole has allowed them to plan and implement curricula and assessment programs fitting to their individual schools.

Other nations might note the heavy focus on equity and Co-operation—not choice and competition—in Finnish schools. Also, when teachers are provided government-paid educational preparation and are given significant recognition and public backing for their work, teaching becomes a desired profession by well-qualified individuals. Although Finland is reticent to say "Do this and you, too, can have high scores," their Ministry of Education has reflected on the differences in Finnish education and tried to provide some background that might explain the cultural differences and practices as reasons for their performances (Hautamäki et al., 2008).

As in other countries, there is some concern about the high PISA scores from the mathematics community in Finland. Citing students' recent low performance on graduation tests, members of university faculty argue that PISA provides a view of everyday mathematics and note the value of such knowledge, but also argue that such knowledge does not include advanced concepts and skills in algebra and other core subjects necessary for study and gainful employment after secondary school (Astala et al., 2005).

That said, there is still concern about the influence of outside forces on Finnish education (Grek et al., 2011; Rautalin & Alasuutari, 2009). The development of the Finnish system of education and the changes made between 1985 and 2000 were based on within-country self-study and the selective importation and emulation of practices seen to work in other countries. These imports were carefully woven into the curricular and professional development work provided for teachers. The OECD PISA results are seen with some distrust, as they come with a cloak of data and information, but bear the impact of scientific truth. Researchers notice that statistical comparisons can often lead to the emulation of some practice of a country placed above the average of other countries (Rautalin & Alasuutari, 2009). Such comparisons and interpretations then become levers for change. In fact, in Finland, the outcome that Finnish student achievement levels had the least variance as a system in the PISA assessments was read as suggesting that perhaps there should be more attention paid to the top students, perhaps they could achieve even more. Although the Finnish take the homogeneity as one of their strengths, the numerical interpretation can be used to suggest a failing. In reaction to the 2000 results, Finnish officials decided to add more emphasis to the curriculum and instituted a call for more core subject influence (Välijärvi et al., 2002). Hence, the numerical results suggested a possible weakness, and hence, even in the face of superior achievement, changes antithetical to the historical culture of Finnish education were made.

In Finland, we again see a country growing out of its own educational history through emulation to develop a system drawing on the best practices of other counties and schools within its own borders. The FNBE directives on curriculum made changes across the 1980s and 1990s consistent with programmatic changes in other countries that seemed to fit Finnish schools, but did so by modifying those practices to the culture of Finnish education. Although this approach led to harmonization through curricular guidance, the Finnish Ministry in 1985 backed off a bit in decentralizing the education system to provide more local control of curriculum within broader guidelines. It was only with the numerical results dealing with the homogeneity of results across the Finnish student body in the PISA assessment that one saw outside influence reach Finland in the form of indicator influence inducing local policy. Although the influence did not have the impact of penetration noted in Germany, this is an instance of convergence of education structure as a result of international assessment and indicator results.

Singapore

Asian student performance has dominated the achievement charts as their countries have held the majority of top rankings in the international large-scale assessments of mathematics performance since their inceptions in the middle 1980s. Asian students' stellar performances have originated from Hong Kong, Japan, Korea, Macau, Singapore, Taipei, and most recently Shanghai. Despite their high rankings in international assessments, Asian countries have not been complacent with their current education systems.

Singapore did not participate in the FIMS or SIMS studies; rather, it made its entry with the TIMSS 1995 study. In 1995 through 2007, Singapore's students performed in the top group of countries and had the highest means, with the exception of the 2007 study, when Chinese Taipei had the highest numerical position but not significantly higher than that of Singapore (Beaton et al., 1996; Mullis et al., 2000, 2004, 2008). In the OECD studies, Singapore, a non-OECD country, has participated in only the 2009 assessment. The Singapore students finished second numerically but not significantly lower than the students of Shanghai-China (OECD, 2010a).

In Singapore, the gap between the intended and the implemented and achieved curricula is small. This alignment results from a close monitoring of teacher progress and student achievement. There is a strong and articulated program of professional development that parallels the curriculum, providing important, grade-specific suggestions in the same time frame where teachers can immediately implement them in their classrooms (Kaur, 2009). The Singapore Ministry of Education noted three problems emanating from the TIMSS findings. The first was that students did not perform well on mathematics that they had not specifically learned and practised. The second was student difficulty in transferring learned knowledge to different contexts. The third dealt with comprehension problems rooted in language issues which arose when unfamiliar words appeared in problems. All three of these issues have found their ways into curricular reform for mathematics in Singapore.

Singapore leaders feel that the results of the studies can provide fresh perspectives and benchmark their performance relative to other countries. However, there is some fear that the high performance levels may lead to feelings of complacency relative to local standards. Singapore feels that such participation provides opportunities to participate in other international comparative projects which have the possibilities of enriching their programs. In particular, they have participated in the Kassell project, the multinational IPMA, and a bilateral project with Brunei Darussalam. Singapore mathematics educators also participate in study tours to other countries and attend conferences of professional mathematics groups internationally. All of these efforts are viewed as adding new vistas to their program's possibilities (Wong, Lee, Kaur, Yee, & Fong, 2009).

Unlike other top-performing Asian countries, Singapore students not only performed well in mathematics, they also displayed a positive attitude towards learning mathematics. The high performance of Singapore students attracted the attention of many Western mathematics educators. The Singapore mathematics curriculum and textbooks have been the focus of a number of studies aimed at identifying factors contributing to the high performance of Singapore students (American Institutes for Research, 2005). Such focussed cross-cultural studies are examples of many small-scale international comparative studies initiated as a result of TIMSS and PISA findings. These again are illustrative of attempted emulation and harmonization processes under Bennett's (1991) model of policy convergence. But this time, other countries want to learn from Singapore's success story.

Concerns Regarding the Impact of International Studies

From the case studies, it is clear that international studies have had, and continue to have, a strong impact on policies for a number of countries. Although such an impact may lead to positive outcomes for mathematics education, there could also be consequences from international studies that are damaging to mathematics education. A critical review of the impact of international studies is essential. The following presents a discussion of concerns regarding the possible impact of international studies.

Concerns Regarding Statistical Precision of the Results

It is not unusual for policy makers to draw quick conclusions by looking at the change in country rankings from one assessment cycle to another. For example, if the ranking (or the country mean score) is worse than for the previous cycle, there may be an immediate outcry about the decline of mathematics standards in the country. This outcry in turn could lead to policy changes. What the policy makers have often missed is that there is always a margin of error in any reported measure. Although those who conduct international studies take great pains in articulating the confidence level surrounding performance measures, these margins of error are often ignored. A policy change may be totally unwarranted, as the change in country ranking could simply be the result of random fluctuation due to the sampling of students (Wu, 2010a, 2010b).

Concerns regarding inferences on causal relationships. International studies such as PISA and TIMSS are cross-sectional sample surveys. Such survey designs are not powerful in establishing causal relationships. Even though student and school background characteristics are captured and correlated with achievement measures, positive correlations do not establish causal relationships. A positive correlation between students' interest in mathematics and test scores in mathematics may be expected. But it is difficult to conclude whether higher interest in mathematics leads to higher achievement, or in fact, higher achievement raises interest. Similarly, better school resources could be positively correlated with higher achievement. But there could be mediating variables such as student socio-economic status (SES) that explain both student achievement and school resources. For example, private schools may have better resources and higher achievement scores, but both could be due to the higher SES of students in private schools. In general, translating survey results into policy measures relies on many assumptions and hypotheses. Some policy changes in response to international study results may be completely off the track.

Concerns regarding using mean scores only. Often the main focus on results of international studies is the country mean score. Although the mean score summarizes overall performance, it could be the case that a country has a large group of low achievers because of geographical remoteness or immigrant composition. That is, the lower mean score could be the result of specific factors rather than an inefficient education system across the board. Policy changes need to take into consideration a myriad of indicators and not just the ranking and mean score of a country. The emphasis on ranking and mean scores, often fuelled by the media, could lead to inappropriate policy changes (Hutchison & Schagen, 2007).

Concerns regarding policy convergence. Although there is a great deal of benefit arising from collaboration, whether internationally or between local communities, there are also a number of concerns in "borrowing" from other education systems, be it the curriculum, assessment, or a management approach. In Bennett's (1991) model, policy convergence in the form of *emulation* appears to be the most flexible, and *penetration* appears to be the most rigid. An authoritarian approach to enforcing standards may work well, provided the standards are sound. The mini-case studies in this paper show that there are significant differences between education systems across the world, and that there are different success models. Finland has clearly showed that a decentralized system with little emphasis on standardized testing can lead to high education attainments, whereas East Asian countries with highly centralized and examination-based education systems are also top performers. What works for one country may not work for another country because of cultural differences and local conditions. This variability is also the case for policy convergence within a country. A national curriculum brings uniformity across states or provinces but stifles diversity and innovation. If education systems are regarded as business models, then the importance of diversity and competition cannot be ignored, as educators have learned from the political and economic arena. When a borrowed system does not fit well within an education community, the consequence could range from a waste of resources to serious damages to the education system (Vithal et al., 2005).

Concerns regarding the use of assessment to drive teaching and learning. Assessments of students should be undertaken as an evaluation of the outcomes of education. Assessments should be designed around teaching and learning, and not the other way round, where teaching and learning are designed around assessment. This direction of design is important as there are important differences between teaching and assessment. One may design an authentic task in assessment where multiple skills are required to solve a problem, but to teach those skills, basic building blocks of skills need to be taught separately, and often in a context-free mode. Only when students have mastered individual skills can they combine the skills and apply them. That is, the way mathematics is taught may be at variance with the way mathematics is assessed. As international studies like PISA and TIMSS are assessments, the adaptation or adoption of the PISA and TIMSS assessment frameworks as curriculum frameworks may not be desirable. For example, as PISA focusses only on problem solving and application in everyday settings, it would be an error for curriculum designers not to include skills involving abstract mathematics as well as basic foundations of mathematics which are often context-free. There is a particular concern when, in order to improve a country's international test scores, the curriculum is changed to match the assessment frameworks of the international assessments.

Additional comments relative to design, interpretation, difficulties in conducting cross-cultural studies, and the drawing of inferences were the focus of a symposium held by the Board on Comparative Studies in Education at the National Research Council in Washington, DC, in 2000 (Porter & Gamoran, 2002).

Retrospective

It is generally acknowledged that international studies such as TIMSS and PISA have an enormous impact on educational policy debates, if not on the policies themselves (Figazzolo, 2009). However, it is not always straightforward to identify the impact of international studies on policies since many policy changes are influenced by international assessments in subtle and indirect ways. Sometimes policy changes evolve over a long period of time, moving slowly and thoughtfully through each of the steps to lasting educational reform. In such cases, it is difficult to attribute a specific lever that triggers a policy implementation. In other cases, media-induced crises lead to rapid, and often thoughtless, reforms lacking foundations in either research or practice. In this chapter, we have reviewed national and local reactions to international studies that are quite public, as well as political and economic processes whose implications are less overt but nevertheless important in influencing policies. Below, we provide a summary of different kinds of policy implications of international studies.

First, results of international studies have been used as policy levers. This trend has been increasing with recent data releases, most notably in the USA, Germany, and Japan. In some cases, the results are used simply as an opportunistic justification for some policies that have already been rolled out. In other cases, new policies have been devised in direct response to the poor performance of students. The policy changes range from changing curriculum content to providing resources to schools.

Second, international studies have been used as performance measures to gauge the success or otherwise of a policy. For example, a policy might be linked to an international study through the setting of a target level of a country's performance in the study. More recently, national achievement measures have been used as economic incentives or indicators by international funding organizations working with developing countries.

Third, international studies have provided a wealth of data and, with that, opportunities for mathematics education researchers to carry out in-depth analyses ranging from classroom climate, gender equity, to curriculum design. Many of these studies are funded by policy bodies with a view that these analyses may influence policies down the line, even if there may be no immediate policy changes based on the research findings. The authors of a number of chapters in this handbook have discussed the link between mathematics education research and policy implementation.

Fourth, international studies such as TIMSS and PISA have led to further transnational dialogs between researchers in assessment, curriculum, and instruction. In Bennett's (1991) model of policy convergence, these are examples of emulation and harmonization.

Fifth, international studies have increased an awareness of the use of student performance measures, and, in some cases, led to the establishment of national sample-based or full-cohort standardized tests. Such enforced tests are examples of Penetration under Bennett's (1991) model where, by law, achievement targets from the tests are set.

International studies have had both positive and negative impacts. On the one hand, it is encouraging to see increased discussion and debate on curriculum content in mathematics, teaching strategies in the classrooms, assessment methodologies, and a rethinking of the values and goals of education more generally. The discussions have certainly stimulated a great deal of reflection, evaluation and constructive criticisms. These have been positive outcomes from international studies. On the other hand, there have also been hasty reactions to study results and rash policy decisions based on unfounded inferences. In particular, the media and some policy makers have been prone to brush aside caveats clearly stated in the study reports, to ignore the degree of confidence one can have in the measures, and to launch into actions that have been typically politically motivated. There are often policy measures that are quick fixes to improve test scores rather than for long-term investment for a better education. These are examples of the negative impact of international studies.

Hopefully, the outcomes of international studies have fostered curricular considerations and productive changes, a careful reflection on cross-cultural comparative methodology, and steps to the improvement of student learning of mathematics worldwide. Researchers and policy-inclined individuals in the mathematics education community need to ask what should be and what are the policy ramifications associated with the TIMSS and PISA assessments, as well as those associated with other international and national assessments of mathematics education. What are the benefits that can be obtained from a careful analysis of the tests, curricula, instructional patterns, opportunity-to-learn, instructional materials and other resources, teacher preparation and professional development and support programs, and related research findings? What are the positive and negative effects resulting from borrowing and promoting the TIMSS and PISA frameworks for developing countries and inducing the insertion of these frameworks into national curricular framework discussions? These questions shape an agenda for mathematics education and policy researchers to examine in the coming decade, as the role of international assessments will surely continue to grow in the number and range of nations participating and in the sources of important indicators chosen (Jones, 2005). For mathematics educators to dismiss the powerful force such assessment programs have on educational policy decisions worldwide would be a dangerous mistake from cultural, mathematical, and educational perspectives.

References

Alexiadou, N. (2007). The Europeanisation of education policy: Researching a changing governance and "new" modes of coordination. *Research in Comparative and International Education, 2*(2), 102–116.

American Institutes for Research. (2005). *What the United States can learn from Singapore's world-class mathematics system (and what Singapore can learn from the United States): An exploratory study*. Washington, DC: AIR. Retrieved from http://www.air.org/reports-products/index.cfm?fa=viewContent&content_id=598

Astala, K., Kivelä, S. K., Koskela, P., Martio, O., Näätänen, M., & Tarvainen, K. (2005, December). The PISA survey tells only a partial truth of Finnish children's mathematical skills. *Matilde, 29*, 9.

Beaton, A. O., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Bennett, C. J. (1991). What is policy convergence and what causes it? *British Journal of Political Science, 21*(2), 215–233.

Bidwell, J. K., & Clason, R. G. (Eds.). (1970). *Readings in the history of mathematics education*. Washington, DC: National Council of Teachers of Mathematics.

Blum, W., & Kaiser, G. (2004). *Kassel Project in Germany.* Budapest, Hungary: Wolters−Kluwer.

Bohl, T. (2004). Empirische Unterrichtsforschung und Allgemeine Didaktik. Entstehung, Situation und Konsequenzen eines prekären Spannungsverhältnisses im Kontext der PISA-Studie [Empirical research on teaching and general didactics: A precarious tension and consequences in the context of the PISA study]. *Die Deutsche Schule*, *96*(4), 414–425.

Burghes, D., Geach, R., & Roddick, M. (Eds.). (2004). *IPMA*. Budapest, Hungary: Wolters−Kluwer.

Burghes, D., Kaur, B., & Thompson, D.R. (Eds.). (2004). *Kassel Project—Final Report*. Budapest, Hungary: Wolters−Kluwer.

Burstein, L. (Ed.). (1993). *The IEA Study of Mathematics III: Student growth and classroom processes*. Oxford, UK: Pergamon.

Carnoy, M. (1999). *Globalization and educational reform: What planners need to know*. Paris, France: United Nations Education, Scientific, and Cultural Organisation.

Carnoy, M. (2006). Rethinking the comparative and the international. *Comparative Education Review, 50*(4), 551–570.

Casassus, J., Froemel, J. E., Palafox, J. C., & Cusato, S. (1998). *First international comparative study of language, mathematics, and associated factors in third and fourth grades*. Santiago, Chile: Latin American Laboratory for Evaluation of the Quality of Education.

Chatterji, M. (2002). Models and methods for examining standards-based reforms and accountability initiatives: Have the tools of inquiry answered pressing questions on improving schools? *Review of Educational Research, 72*(3), 345–386.

College Entrance Examination Board. (1959). *Program for college preparatory mathematics*. New York, NY: Author

Council of Chief State School Officers & National Governors Association. (2010). Common core state standards for mathematics. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf.

Cussó, R., & D'Amico, S. (2005). From development comparativism to globalization comparativism: Towards more normative international education statistics. *Comparative Education, 41*(1), 1–16.

Dossey, J. A., Mullis, I. V. S., Lindquist, M. M., & Chambers, D. L. (1988). *The mathematics report card: Are we measuring up?* Princeton, NJ: Educational Testing Service.

Eaton, J., & Kortum, S. (1996). Trade in ideas: Patenting and productivity in the OECD. *Journal of International Economics, 40*(3–4), 251–278.

Ertl, H. (2006). Educational standards and the changing discourse on education: The reception and consequences of the PISA study in Germany. *Oxford Review of Education, 32*(5), 619–634.

Figazzolo, L. (2009). *Impact of PISA 2006 on the education policy debate*. Education International. Retrieved from http://www.ei-ie.org/research/en/documentation.php.

Greaney, V., & Kellaghan, T. (2008). *Assessing national achievement levels in education*. Washington, DC: International Bank for Reconstruction and Development/The World Bank.

Grek, S. (2009). Governing by numbers: The PISA "effect" in Europe. *Journal of Educational Policy, 24*(1), 23–37.

Grek, S., Lawn, M., Lingard, B., Rinne, R., Segerholm, C., & Simola, H. (2011). National policy brokering and the construction of the European Education Space in England, Sweden, Finland and Scotland. In J. Ozga, P. Dahler-Larsen, C. Segerholm, & H. Simola (Eds.), *Fabricating quality in education: Data and governance in Europe* (pp. 47–65). London, UK: Routledge.

Grier, K., & Grier, R. (2007). Only income diverges: A neoclassical anomaly. *Journal of Developmental Economics, 84*, 25–45.

Hautamäki, J., Harjunen, E., Hautamäki, A., Karjalainen, T., Kupiainen, S., Laaksonen, S., Lavonen, J., Pehkonen, E., Rantanen, P., Scheinin, P., Halinen, I., & Jakku-Sihvonen, R. (2008). *PISA 2006 Finland: Analyses, reflections and explanations*. Helsinki, Finland: Ministry of Education.

Hopmann, S. T., & Brinek, G. (2007). Introduction: PISA according to PISA—Does PISA keep what it promises? In S. T. Hopmann, G. Brinek, & M. Retzel (Eds.), *PISA according to PISA—Does PISA keep what it promises* (pp. 9–19). Vienna, Austria: LIT-Verlag.

Husén, T. (Ed.). (1967). *International Study of Achievement in Mathematics: A comparison of twelve countries* (2 vols.). Stockholm, Sweden: Almqvist & Wiksell.

Hutchison, D., & Schagen, I. (2007). Comparisons between PISA and TIMSS—Are we the man with two watches? In T. Loveless (Ed.), *Lessons learned—What international assessments tell us about math achievement* (pp. 227–261). Washington, DC: Brookings Institute Press.

International Association for the Evaluation of Educational Achievement. (2011). *Brief history of the IEA*. Retrieved from http://www.iea.nl/brief_history_of_iea.html.

International Commission on Mathematical Instruction. (2011a). *A historical sketch of ICMI*. Retrieved from http://www.mathunion.org/icmi/about-icmi/a-historical-sketch-of-icmi/.

International Commission on Mathematical Instruction. (2011b). *The first century of the International Commission on Mathematical Instruction (1908–2008)*. Retrieved from http://www.icmihistory.unito.it/timeline.php.

International Project on Mathematical Attainment. (2011). *CIMT research projects and publications.* Retrieved from http://www.cimt.plymouth.ac.uk/menus/research.htm.

Jones, P. W. (2004). Taking the credit: Financing and policy linkages in the educational portfolio of the World Bank. In G. Steiner-Khamsi (Ed.), *The global politics of educational borrowing and lending* (pp. 188–200). New York, NY: Teachers College Press.

Jones, G. A. (2005). What do studies like PISA mean to the mathematics education community? In H. L. Chick & J. L. Vincent (Eds.), *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 71–74). Melbourne, Australia: PME.

Jones, P. W., & Coleman, D. (2005). *The United Nations and education: Multilateralism, development and globalization*. London, UK: Routledge/Falmer.

Jones, P. S., & Coxford, A. F., Jr. (Eds.). (1970). *A history of mathematics education in the United States and Canada* (Thirty-second Yearbook of the National Council of Teachers of Mathematics). Washington, DC: National Council of Teachers of Mathematics.

Kang, H. J. (2009). A cross-cultural curriculum study on U.S. elementary mathematics textbooks. In S. L. Swars, D. W. Stinson, & S. Lemons-Smith (Eds.), *Proceedings of the 31st Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 5, pp. 379–386). Atlanta, GA: Georgia State University.

Kaur, B. (2009). Performance of Singapore students in Trends in International Mathematics and Science Studies (TIMSS). In K. Y. Wong, P. P. Lee, B. Kaur, F. P. Yee, & N. S. Fong (Eds.), *Mathematics education: The Singapore journey* (pp. 439–463). Singapore: World Scientific.

Kellaghan, T., Greaney, V., & Murray, T. S. (2009). *Using the results of a national assessment of educational achievement*. Washington, DC: World Bank.

Kerr, C. (1983). *The future of industrial societies: Convergence or continuing diversity?* Cambridge, MA: Harvard University Press.

Kilpatrick, J. (1971). Some implications of the International Study of Achievement in Mathematics for mathematics educators. *Journal for Research in Mathematics Education, 2*(2), 164–171.

Kilpatrick, J. (2009, February–March). TIMSS 2007 mathematics: Where are we? *MAA Focus*, 4–7.

Kilpatrick, J., Mesa, V., & Sloane, F. (2007). U.S. algebra performance in an international context. In T. Loveless (Ed.), *Lessons learned: What international assessments tell us about math achievement* (pp. 85–126). Washington, DC: Brookings Institution Press.

Kimmelman, P., Kroze, D., Schmidt, W., van der Ploef, A., McNeely, M., & Tan, A. (1999). *A first look at what we can learn from high-performing school districts: An analysis of TIMSS data from the First in the World Consortium*. Washington, DC: National Institute on Student Achievement, Curriculum, and Assessment.

Klieme, E., Avenarius, H, Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H., & Vollmer, H. (2003). *Zur Entwicklung nationaler Bildungsstandards* [Toward the development of national standards for education]. Berlin, Germany: Bundesministerium fűr Bildung und Forschung.

Kline, M. (1961, October). Math teaching assailed as peril to U.S. scientific progress. *New York University Alumni News*, n.p.

Kupiainen, S., & Pehkonen, E. (2008). PISA 2006 mathematical literacy assessment. In J. Hautamäki, E. Harjunen, A. Hautamäki, T. Karjalainen, S. Kupiainen, S. Laaksonen, J. Lavonen, E. Pehkonen, P. Rantanen, P. Scheinin, I. Halinen, & R. Jakku-Sihvonen (Eds.), *PISA 2006 Finland: Analyses, reflections and explanations* (pp. 117–143). Helsinki, Finland: Ministry of Education.

Lapointe, A. E., Mead, N. A., & Askew, J. M. (1992). *Learning mathematics*. Princeton, NJ: Educational Testing Service.

Lapointe, A. E., Mead, N. A., & Phillips, G. W. (1988). *A world of differences: An international assessment of mathematics and science*. Princeton, NJ: Educational Testing Service.

Latin American Laboratory for Assessment of the Quality of Education. (2001). *First international comparative study of language, mathematics, and associated factors for third and fourth grade primary school students*. Santiago, Chile: Latin American Laboratory for Assessment of the Quality of Education.

Lester, F., Jr Dossey, K. J. A., & Lindquist, M. M. (2007). Challenges and opportunities in the analysis of NAEP mathematics results. In P. Kloosterman & F. K. Lester Jr. (Eds.), *Results and interpretations of the 2003 mathematics assessment of the National Assessment of Educational Progress* (pp. 311–331). Reston, VA: National Council of Teachers of Mathematics.

Luhmann, N. (1990). *Essays on self-reference*. New York, NY: Columbia University Press.

Malaty, G. (2006, December). What are the reasons behind the success of Finland in PISA? *Matilde, 29*, 4–8.

Mathematical Sciences Education Board. (1989). *Everybody counts*. Washington, DC: National Academy Press.

Mayer-Foulkes, D. (2010). *Divergences and convergences in human development* (Human Development Research Paper 2010/20). New York, NY: United Nations Development Programme.

McGaw, B. (2008). The role of the OECD in international comparative studies of achievement. *Assessment in Education: Principles, Policy & Practice, 15*(3), 223–243.

McKnight, C. C., Crosswhite, F. J., Dossey, J. A., Kifer, E., Swafford, J. O., Travers, K. J., & Cooney, T. J. (1987). *The underachieving curriculum: Assessing U.S. school mathematics from an international perspective*. Champaign, IL: Stipes Publishing.

McLeod, D. B., Stake, R. E., Schappelle, B. P., Mellissinos, M., & Gierl, M. J. (1996). Setting the standards: NCTM's role in the reform of mathematics education. In S. A. Raizen & E. D. Britton (Eds.), *Bold ventures: Case studies of U.S. innovations in mathematics education* (Vol. 3, pp. 13–132). Dordrecht, The Netherlands: Kluwer.

Miserable Noten für deutsche Schüler [Abysmal marks for German students]. (2001, December 4). *Frankfurter Allgemeine*.

Moskowitz, J. H., & Stephens, M. (Eds.). (2004). *Comparing learning outcomes: International assessments and educational policy*. London, UK: Routledge/Falmer.

Mullis, I. V. S., Martin, M. O., Foy, P., Olson, J. F., Preuschoff, C., Erberber, E., Arora, A., & Galia, J. (2008). *TIMSS 2007 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., Chrostowski, S. J., & Smith, T. A. (2000). *TIMSS 1999 international mathematics report*. Chestnut Hill, MA: International Study Center, Lynch School of Education, Boston College.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., O'Connor, K. M., Chrostowski, S. J., Gregory, K. D., Garden, R. A., & Smith, T. A. (2001). *Mathematics benchmarking report: TIMSS*

1999–Eighth grade. Chestnut Hill, MA: TIMSS International Study Center, Lynch School of Education, Boston College.

Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2009). *TIMSS 2011 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Murphy, S. (2010). The pull of PISA: Uncertainty, influence, and ignorance. *Interamerican Journal of Education for Democracy, 3*(1), 28–44.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperatives of education reform*. Washington, DC: Government Printing Office.

National Council of Teachers of Mathematics. (1964). *The revolution in school mathematics*. Washington, DC: Author.

National Council of Teachers of Mathematics. (1961). *An analysis of new mathematics programs*. Washington, DC: Author.

National Council of Teachers of Mathematics. (1980). *An agenda for action*. Reston, VA: Author.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Nguyen, M., Elliott, J., Terlouw, C., & Pilot, A. (2009). Neocolonialism in education: Cooperative learning, Western pedagogy in an Asian context. *Comparative Education, 45*(1), 109–130.

Niss, M. A. (1999). Kompetencer og uddannelsesbeskrivelse [Competencies and description of education]. *Uddannelse, 9*, 21–29.

Nóvoa, A., & Yariv-Masal, T. (2003). Comparative research in education: A mode of governance or a historical journey? *Comparative Education, 39*(4), 423–438.

Olson, L. (2004, September 22). No Child Left Behind Act changes weighed. *Education Week, 24*(4), 31, 34.

On the mathematics curriculum of the high school. (1962). *American Mathematical Monthly, 69*(3), 189–193.

Organisation for Economic Co-operation and Development. (2001). *Knowledge and skills for life: First results from PISA 2000*. Paris, France: Directorate for Education, OECD.

Organisation for Economic Co-operation and Development. (2002). *PISA in the news in Germany: Dec 2001–Jan 2002*. Paris, France: Directorate for Education, OECD.

Organisation for Economic Co-operation and Development. (2003). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris, France: Directorate for Education, OECD.

Organisation for Economic Co-operation and Development. (2004a). *Learning for tomorrow's world: First results from PISA 2003*. Paris, France: Directorate for Education, OECD.

Organisation for Economic Co-operation and Development. (2004b). *What makes school systems perform? Seeing school systems through the prism of PISA*. Paris, France: Directorate for Education, OECD.

Organisation for Economic Co-operation and Development. (2007). *PISA 2006: Science competencies for tomorrow's world* (Vol. 1: Analysis). Paris, France: Directorate for Education, OECD.

Organisation for Economic Co-operation and Development. (2010a). *PISA 2009 results: What students know and can do–Student performance in reading, mathematics and science* (Vol. 1). Paris, France: Directorate for Education, OECD.

Organisation for Economic Co-operation and Development. (2010b). *Draft of PISA 2012 Mathematics Framework*. Paris, France: Directorate for Education, OECD.

Organisation for Economic Co-operation and Development. (2010c). *Education at a glance 2010: OECD indicators*. Paris, France: Directorate for Education, OECD.

Owen, E., Stephens, M., Moskowitz, J., & Gil, G. (2004). Toward education improvement: The future of international assessment. In J. H. Moskowitz & M. Stephens (Eds.), *Comparing learning outcomes: International assessments and educational policy* (pp. 3–23). London, UK: Routledge/Falmer.

Pashley, P. J., & Phillips, G. W. (1993). *Toward world class standards: A research study linking international and national assessments*. Princeton, NJ: Educational Testing Service.
Pellegrino, J. W., Jones, L. R., & Mitchell, K. J. (Eds.). (1999). *Grading the nation's report card: Evaluating NAEP and transforming the assessment of educational progress*. Washington, DC: National Academy Press.
Peters, M. (2002). Education policy research and the global knowledge economy. *Educational Philosophy and Theory, 34*(1), 91–102.
Phillips, D., & Ochs, K. (2003). Processes of policy borrowing in education: Some explanatory and analytic devices. *Comparative Education, 39*(4), 451–464.
Porter, A. C., & Gamoran, A. (Eds.). (2002). *Methodological advances in cross-national surveys of educational achievement*. Washington, DC: National Academy Press.
Porter, A., McMaken, J., Hwang, J., & Yang, R. (2011, April). Common core standards: The new U.S. intended curriculum. *Educational Researcher, 40*(3), 103–116.
Prais, S. J. (2003). Cautions on OECD's recent educational survey (PISA). *Oxford Review of Education, 29*(2), 139–163.
Rautalin, M., & Alasuutari, P. (2009). The uses of the national PISA results by Finnish officials in central government. *Journal of Education Policy, 24*(5), 539–556.
Ripley, A. (2011). The world's schoolmaster: How a German scientist is using test data to revolutionize global learning. *Atlantic Magazine, 308*(1), 109–110.
Robitaille, D. F., & Beaton, A. E. (Eds.). (2002). *Secondary analysis of the TIMSS data*. Dordrecht, The Netherlands: Kluwer.
Robitaille, D. F., & Garden, R. A. (Eds.). (1989). *The IEA Study of Mathematics II: Contexts and outcomes of school mathematics*. Oxford, UK: Pergamon.
Rutkowski, D. (2008). Towards an understanding of educational indicators. *Policy Futures in Education, 6*(4), 470–481.
Sahlberg, P. (2010, December 27). Learning from Finland: How one of the world's top educational performers turned around. *Boston Globe*, A9.
Schmidt, W. H., Jorde, D., Cogan, L. S., Barrier, E., Gonzalo, I., Moser, U., Shimizu, K., Sawada, T., Valverde, G., McKnight, D., Prawat, R., Wiley, D. E., Raizen, S., Britton, E. D., & Wolfe, R. G. (1996). *Characterizing pedagogical flow: An investigation of mathematics and science teaching in six countries*. Dordrecht, The Netherlands: Kluwer.
Schmidt, W. H., McKnight, C., Cogan, L. S., Jakwerth, P. M., & Houang, R. T. (1999). *Facing the consequences: Using TIMSS for a closer look at U.S. mathematics and science education*. Dordrecht, The Netherlands: Kluwer.
Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H. C., Wiley, D. E., Cogan, L. S., & Wolfe, R. G. (2001). *Why schools matter: A cross-national comparison of curriculum and schooling*. San Francisco, CA: Jossey-Bass.
Schmidt, W. H., McKnight, C., & Raizen, S. (1997). *A splintered vision: An investigation of U.S. science and mathematics education*. Dordrecht, The Netherlands: Kluwer.
Schmidt, W. H., McKnight, C., Valverde, G. A., Houange, R. T., & Wiley, D. E. (1997). *Many visions, many aims* (Vol. 1: A cross-national investigation of curricular intentions in school mathematics). Dordrecht, The Netherlands: Kluwer.
Schriewer, J. (1990). The method of comparison and the need for externalization: Methodological criteria and sociological concepts. In J. Schriewer & B. Holmes (Eds.), *Theories and methods in comparative education* (pp. 25–83). Frankfurt am Main, Germany: Lang.
Sjøberg, S. (2007). PISA and "real life challenges": Mission impossible? In S. T. Hopmann, G. Brinek, & M. Retzel (Eds.), *PISA according to PISA—Does PISA keep what it promises?* (pp. 203–224). Vienna, Austria: LIT-Verlag.
Sloane, P. F. E., & Dilger, B. (2005) The competence clash—Dilemata bei der űbertragung des "Konzepts der nationalen Bildungsstandards" auf die berufliche Bildung [The competence clash: Dilemmas in the transmission of the "concept of national education standards" to vocational training]. *Berufs- und Wirtschaftspädagogik*. Retrieved from http://www.bwpat.de.
Smith, D. E. (1909). *The teaching of arithmetic*. Boston, MA: Ginn.

Southern and Eastern Africa Consortium for Monitoring Educational Quality. (2011). SACMEQ: 1995–2010. Retrieved from http://www.sacmeq.org/.

Stanat, P., Artelt, C., Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Schümer, G., Tillmann, K., & Weiß, M. (2002). *PISA 2000: Overview of the study—Design, method and results*. Berlin, Germany: Max Planck Institute for Human Development.

Steiner-Khamsi, G. (2006). The economics of policy borrowing and lending: A study of late adopters. *Oxford Review of Education, 32*(5), 665–678.

Steiner-Khamsi, G. (2007). International knowledge banks and the production of educational crises. *European Educational Research Journal, 6*(3), 285–292.

Thomas, J. (2001). Globalization and the politics of mathematics education. In B. Atweh, B. Nebres, & H. Forgasz (Eds.), *Sociocultural research on mathematics education: An international perspective* (pp. 95–112). Mahwah, NJ: Erlbaum.

Travers, K. J., & Westbury, I. (Eds.). (1989). *The IEA Study of Mathematics II: The analysis of mathematics curricula*. Oxford, UK: Pergamon.

UNESCO. (1990). *World declaration on education for all.* Retrieved July 15, 2011 from http://www.unesco.org/education/efa/ed_for_all/background/jomtien_declaration.shtml.

UNESCO. (2011). *The hidden crisis: Armed conflict and education*. Paris, France: United Nations Educational, Scientific, and Cultural Organization.

United States Bureau of Education. (1911). *Mathematics in the elementary schools of the United States*. Washington, DC: Government Printing Office.

Välijärvi, J., Linnakylä, P., Kupari, P., Reinikainen, P., & Arffman, I. (2002). *The Finnish success in PISA—And some reasons behind it*. Jyväskylä, Finland: Finnish Institute for Educational Research, University of Jyväskylä.

Valverde, G. A., Bianchi, L. J., Wolfe, R. G., Schmidt, W. H., & Houang, R. T. (2002). *According to the book: Using TIMSS to investigate the translation of policy into practice through the world of textbooks*. Dordrecht, The Netherlands: Kluwer.

Vithal, R., Adler, J., & Keitel, C. (Eds.). (2005). *Researching mathematics education in South Africa: Perspectives, practices and possibilities*. Cape Town, South Africa: Human Sciences Research Press.

Wong, K. Y., Lee, P. P., Kaur, B., Yee, F. P., & Fong, N. S. (Eds.). (2009). *Mathematics education: The Singapore journey*. Singapore: World Scientific.

World Bank. (2005, October 12). *Education for all—Fast track initiative. Fact sheet: About aid effectiveness.* Washington, DC: World Bank.

Wu, M. (2010a). *Comparing the similarities and differences of PISA 2003 and TIMSS* (OECD Education Working Paper No. 32). Paris, France: Directorate for Education, OECD.

Wu, M. (2010b). Measurement, sampling, and equating errors in large-scale assessments. *Educational Measurement: Issues and Practice, 29*(4), 15–27.