Similarity and Difference in International Comparative Research in Mathematics Education

David Clarke
University of Melbourne
<d.clarke@unimelb.edu.au>

This paper reports the key findings of an extensive review and critique of international comparative research in mathematics education. A feature of the reports of such research is the interweaving of similarities and differences, and it is proposed that research should address more explicitly the interconnectedness of these similarities and differences. Survey-style and case study approaches are examined in this regard. Issues raised include representation, appropriation, and exploitation, and the cultural authorship of international comparative research.

The Assumptions of International Comparative Research

International comparative research is certainly alive and well, judging by the volume of studies reported in the literature. In Raby's bibliography of comparative and international education studies published in 1999 alone, 937 papers from 110 journals are divided into twenty-four categories (Raby, 2000). Mathematics Education was not one of those categories. Categories such as "Gender" and "Higher Education" suggest the dimensions across which it is assumed to be legitimate to undertake international comparative research.

Keitel and Kilpatrick (1999) have problematised the assumptions on which international comparative studies of school mathematics are predicated. In particular, they question the treatment of the mathematics curriculum as unproblematic and the associated assumption that a single test can give comparable measures of curriculum effects across countries. They further suggest that the spectre of an "idealized international curriculum" lies behind even the most sophisticated research designs, including text and document analyses and the use of video to study classroom practice.

> A pseudo-consensus has been imposed (primarily by the English-speaking world) across systems so that curriculum can be taken as a constant rather than a variable, and so that the operation of other variables can be examined (Keitel & Kilpatrick, 1999, p. 253).

Thorsten (2000) makes a similar but more general point about the narratives by which education is related to economic strength and the cultural specificity of these narratives. That is, while the conduct of much of international comparative research in mathematics education seems predicated on assumptions derived from conceptions of a global mathematics curriculum, the interpretation of the results of such research will always be framed by politicians, policy makers and curriculum developers in terms of national rather than international aspirations, values, needs and conditions. The situatedness of such interpretations is not only inevitable, it is eminently sensible – even though the actual interpretations may be driven by political rather than educational motives, by a spirit of competition rather than cooperation, and by principles of elitism rather than equity.

We need to challenge the assumption that international comparative studies in education are necessarily evaluative, much less competitive. The potential utility of international comparative research should not be discounted because of the current

preoccupation with competition. The locally produced narratives that are fueled by international comparative research could be about the mutual benefits of sharing good practice and about the adaptive potential of the policies and practices of other educational systems to our own. Those who cite the absurdity of comparing apples with oranges need to consider the benefit of grafting new varieties onto sturdy old stock. The fruitfulness of the product will depend significantly on the (cultural) compatibility of the grafted material and the original stock.

Central to the conduct and use of international comparative research is the position adopted with respect to variation. Depending on the researcher's affiliations, variation is seen either as something that international research aspires to minimize or remove, or as an insurmountable barrier to the utilization of international research. The accommodation of variation in international research will not be achieved by either the identification of universally applicable "big theories" or "best practice" or by the pessimistic assumption of the incommensurability of international differences in educational policy and practice. International comparative research should document and report variation in educational policy and practice in a manner that anticipates further variation in the adaptation and application of such research. This goal is best achieved if our research designs maximize and optimize the contribution of those voices in which our research is constituted: Not only the voices of the participants (teachers and students in classroom settings, for example), but also to the voices of the interpreting researchers, whose cultural affiliations inevitably contribute to the form of their analyses.

Meaningful International Comparisons

One of the most widely reported results from studies of international assessment of student achievement such as the Third International Mathematics and Science Study (TIMSS) (Beaton & Robitaille, 1999) has been the high national mean scores for students from "Asian" countries. This appears to have triggered the following (naïve) line of reasoning: If Asian countries are consistently successful on international measures of mathematics performance, then less-successful non-Asian countries would do well to adapt for their use the instructional practices of Asian classrooms. Such a line of reasoning is grounded in four key assumptions: (i) that the term "Asian" identifies a coherent body of practice; (ii) that the performances valued in international tests constitute an adequate model of mathematics, appropriate to the needs of the less-successful country; (iii) that differences in mathematical performance are attributable to differences in instructional practice (and not to other differences in culture, societal affluence or aspiration, or curriculum); and (iv) that the distinctive instructional practices of more-successful countries (should these exist) can be meaningfully adapted for use by less-successful countries. Each of these key assumptions can be problematised on a variety of grounds (eg Clarke, 2003; Westbury, 1992).

In his re-analysis of data from the Second International Mathematics Study (SIMS), Bracey (1997) suggested that the differences in mathematics performance found at an international level were replicated in a partitioning of the USA sample along cultural or ethnic lines. As a simple illustration of this point: Asian-American students, participating in a school system that has been substantially maligned in the USA popular press, perform

at a level comparable with their high-performing counterparts in schools in Asian countries. This single illustration suggests that differences on particular measures of mathematical performance are at least as attributable to the cultural affiliation of the students as to the particular school system attended. The significance of such internal cultural variation is lost in the aggregation of performance data for countries as culturally plural as the USA, Australia, or Canada. Such analyses also have implications for societies with a small number of substantial ethnically-distinct communities, such as Malaysia and South Africa.

Berliner reiterated this point in an article in the Washington Post (Sunday, January 28, 2001, p. B3).

> Which America are we talking about? … the scores of white students in the United States were exceeded by only three other nations. But black American school children were beaten by every single nation, and Hispanic kids were beaten by all but two nations. … The true message of the TIMSS-R and other international assessments is that the United States will not improve in international standings until our terrible inequalities are fixed (Berliner, 2001, B3).

That is, rather than serving an agenda of international competitive comparison, the results of international achievement testing can be analysed to identify members of a nation who are less well served by the school system than others.

A corollary to this line of reasoning is voiced by Wang (2001) who, in discussing technical concerns with TIMSS, cites Hu (2000, p. 8) as saying, "This study does not break down Americans by race, if they did, Asian Americans would likely score as high as Asians in their home countries, and Whites would rank near top of the European nations." There are several ways to interpret this observation. Berliner's approach seems the most rational and productive: From several perspectives the comparison of national means of student achievement is problematic. Comparisons between sectors of the community within a given country may be more fruitful, within a given state or school system even more so. Such comparisons may at least highlight community groups who are less equal in the benefits they accrue from a school system intended to benefit all students equally. Educational policy can then be framed to address any inequalities. Of course, this raises the question of the value of international comparative research.

The central problem of international comparative research in education can be summarized as: How are the educational systems of different countries most usefully compared if our goal is the improvement of those educational systems? The previous remarks are not intended to challenge the premise that school systems enact cultural values. However, they do challenge the simplistic identification of culture with nationality. Once the identification (confusion) of nation with culture has been problematised, then the utility of international comparative research can be considered with greater cultural sensitivity.

Similarity and Difference in International Comparative Research

Schmidt, McKnight, Valverde, Houang and Wiley (1997) investigated the mathematics curricula of the "almost 50" countries participating in the Third International Mathematics and Science Study (TIMSS). The documented differences in curricular organisation were extensive. Even within a single country differentiated curricular catered to communities perceived as having different needs. Countries differed in the extent of such differentiation, in the complexity or uniformity of their school systems, and in the distribution of educational decision-making responsibility within those school systems. Given such

diversity, the identification of any curricular similarity with regard to mathematics should be seen as significant. And there were significant similarities. There were similarities of topic, if not of curricular location; broad correspondences of grade level and content that became differences if you looked more closely; differences in the range of content addressed at a particular grade level, but which repeated particular developmental sequences where common content was addressed over several grade levels. In another international study of mathematics curricula, the OECD study of thirteen countries' innovative programs in mathematics, science and technology found that, "Virtually everywhere, the curriculum is becoming more practical" (Atkin & Black, 1997, p. 24). Yet, despite this common trend, the same study found significant differences in the reasons that prompted the new curricula (Atkin & Black, 1996). These interwoven similarities and differences are almost the signature of international comparative research.

Schmidt et al (1997) reported that differences in the characterization of mathematical activity were extreme at the Middle School level; from "representing" situations mathematically, "generalizing" and "justifying" to "recalling mathematical objects and properties" and "performing routine procedures". Despite the apparent diversity, it was the latter two expectations that were emphasised in the curricula studied. Given the documented diversity, it is the occurrence of similarity that requires explanation. Some curricular similarities may be the heritage of a colonial past. Others may be the result of more recent cultural imperialism or simply good international marketing.

In attempting to tease out the patterns of institutional structure and policy evident in international comparative research (particularly in the work of LeTendre, Baker, Akiba, Goesling, & Wiseman, 2001), Anderson-Levitt (2002) noted the "significant national differences in teacher gender, degree of specialization in math, amount of planning time, and duties outside class" (p. 19). But these differences co-exist with similarities in school organization, classroom organization, and curriculum content. Anderson-Levitt (2002, p. 20) juxtaposed the statement by LeTendre et al. that "Japanese, German and U.S. teachers all appear to be working from a very similar 'cultural script'" (2001, p. 9) with the conclusions of Stigler and Hiebert (1999) that U.S. and Japanese teachers use different cultural scripts for running lessons. The apparent conflict is usefully (if partially) resolved by noting with Anderson, Ryan and Shapiro (1989) that both U.S. and Japanese teachers draw on the same small repertoire of "whole-class, lecture-recitation and seatwork lessons conducted by one teacher with a group of children isolated in a classroom" (Anderson-Levitt, 2002, p. 21), but they utilise their options within this repertoire differently.

LeTendre et al (2001) claim that "Policy debates in the USA are increasingly informed by use of internationally generated, comparative data" (p. 3). LeTendre and his colleagues go on to argue that criticisms of international comparative research on the basis of "culture clash" ignore international isomorphisms at the level of institutions (particularly schools). LeTendre et al. report yet another interweaving of similarity and difference.

> We find some differences in how teachers' work is organised, but similarities in teachers' belief patterns. We find that core teaching practices and teacher beliefs show little national variation, but that other aspects of teachers' work (e.g., non-instructional duties) do show variation (p. 3).

These differences and the similarities are interconnected and interdependent and it is likely that policy and practice are best informed by research that examines the nature of the

interconnection of specific similarities and differences, rather than simply the frequency of their occurrence.

Issues of Authorship, Voice, and Purpose

In an international comparative study, any evaluative aspect is reflective of the cultural authorship of the study. If the researchers are to make judgements of merit, whether they are about student achievement or classroom practice, they can only do so from the position of the authoring culture. The efficacy of a practice can only be judged to the extent that it achieves a specified goal. The most obvious goal against which to assess the efficacy of a practice is the goal of the individual or school system engaged in the practice. For the purposes of international comparative research, however, it is legitimate for someone outside the system being studied to evaluate a practice relative to their own goals – provided that this distinction is made explicit. For example, a researcher evaluating a particular curriculum structure would draw different conclusions regarding the efficacy of the structure if the evaluating researcher were assessing its potential utility for a school system in which most students only completed tenth grade compared with one for which twelfth-grade completion were the norm. This is only to say that the report of an international comparative study need not be evaluative, but the readership of such a report may engage in evaluation of the report in relation to their own goals, school system, and culture.

The other aspect of cultural authorship relevant here is the issue of representation and voice. In commenting on the proliferation of OECD-initiated international comparative research projects, Cohen characterised the OECD as "a club of 29 of the world's richest countries" (Cohen, 1998, p. 4). Even when less affluent countries participate in international studies, it is frequently as the objects of investigation rather than as partners in the research. Research is conducted from a Western perspective and evaluates the practices it studies by Western criteria.

> This research remains largely bounded by the Western conception of (teacher-centred) pedagogical practice and by implicit social rules pertaining to authority and social participation (Fuller & Clarke, 1994, pp. 143-144).

Among the volumes of text prompted by recent international comparative studies such as TIMSS, no country it seems has been as prolific in generating papers as the USA. These papers warrant a separate analysis of their own, for in them one finds the narratives of nationhood (see Thorsten, 2000). We find such narratives in the media reporting of other large-scale international comparative studies such as PISA (the Programme for International Student Assessment; website: www.pisa.oecd.org). The situatedness of these narratives is inevitable and even illuminating. But their situated character should be acknowledged and allowed for in the reading of any report of international comparative research. It might even be the focus of such research.

Globalisation seeks to minimise international differences (whether by consensus or imposition) whereas internationalisation seeks to celebrate both the similarities and the differences and to learn from them. This difference can be illustrated by comparing the goal of aspiring to standardize instructional practice in mathematics classrooms internationally and the goal of aspiring to optimize local practice through critical reflection stimulated by

consideration of best practice elsewhere. Of all the approaches to international comparative research, it is the video studies that we might expect to have the greatest potential to inform classroom practice. The methodological challenges of international "video survey" studies have been usefully discussed by Stigler, Gallimore, and Hiebert (2000).

Watanabe (2001) quotes White (1987) as writing "we should hold Japan up as a mirror, not as a blueprint". This powerful and appealing metaphor can serve as a general characterisation of one of the major uses of international comparative studies of classroom practice. Use of this metaphor places the agency for the interpretation and adaptation of any documented practice with the person looking in the mirror. There is no invocation of absolute best practice – the judgement is a relativist one, and an instructional activity with a high degree of efficacy in Hong Kong may retain little effectiveness when employed in a Swedish classroom, where different cultural values inform and frame the actions of all classroom participants. Most importantly, we are encouraged to study Japanese (or South African or German) classrooms not solely for the purposes of mimicking their practices but for their capacity to support us in our reflection on our own practice. The mutuality of the potential benefit provides further motivation for such research.

The Way Forward: Learning From Similarity and Difference

International comparative research is open to misuse in at least three ways: (i) Through the *imposition* on participating countries of a global curriculum against which their performance will be judged; (ii) Through the *appropriation* of the research agenda by those countries most responsible for the conduct of the study, the design of the instruments, and the dissemination of the findings; and, (iii) Through the *exploitation* of the results of such studies to disenfranchise communities, school systems, or the teaching profession through the implicit denigration of curricula or teaching practices that were never designed to achieve the goals of the global curriculum on which such studies appear predicated. Each item would be cause for significant concern if it were shown that research agency resided exclusively with a particular cultural perspective (for example, Western or East-Asian).

"Adaptive potential" is a key consideration by which researchers and educators from one school system consider the relevance and utility of the policies and practices of another. Hatano and Inagaki (1998) remind us that the adaptation of pedagogical practice requires consideration of both the practicality of technical implementation and the extent to which the beliefs underlying the adapted practice are in harmony with local cultural values.

If, as I am arguing, the adaptability of a policy or a practice is dependent on the degree of consonance between the settings and the beliefs of the originating and adapting cultures, then it is incumbent upon the researcher to provide details of those settings and beliefs. This detail is also essential if the research is to support our interrogation of our own practice. If a key criterion for the consideration of a practice as valuable is the learning it promotes, then our research designs must afford plausible connection between a particular classroom practice and consequent learning outcomes. The methodological approach of the Learner's Perspective Study (http://www.edfac.unimelb.edu.au/DSME/lps/) provides one example of how this might be done.

International comparative research must be undertaken on a basis of mutual benefit to all participants. The fundamental reflexivity that is embodied in the metaphor of the mirror

rather than the blueprint should underlie the function of all such studies. We must guard against the cultural imperialism of an implicit global curriculum (whether Western or Asian in character) and, instead, stress the centrality of local interpretation of all findings. Adoption of such a relativist approach avoids the competitive dichotomisation of Asian and Western cultural traditions.

The utility of international comparative testing rests with the reader's endorsement of the test instrument as the legitimate operationalisation of valued (and locally relevant) educational goals. Even with this endorsement, research into international differences and similarities in student mathematical performance (such as TIMSS and PISA) has limited utility, except as a form of national report card, unless it is accompanied by data that suggest cultural, societal, or instructional variation that might be used to explain such differences and similarities and then to promote improved mathematical learning and associated performance. Case studies offer an alternative to the dangers inherent in survey research of superficiality and indiscriminant aggregation. It is not obvious to everyone, however, that comparative case studies can inform policy. LeTendre et al (2001) argue that they can.

> Studies such as Spindler (1987), Tobin, Wu and Davidson (1989), or Shimahara and Sakai (1995) provide powerful insights into the way teacher practice and belief is shaped within different national contexts. Comparative studies that provide this level of ethnographic detail hold much potential for educational policy (p. 23).

From the studies that have been done, we have every reason to believe that it is in these interrelationships that the character and function of culture will most clearly emerge: In the teacher practice that mediates between curriculum content and the student, through the actions and the lesson structure that constitute the enactment of that curriculum in the classroom, together with the beliefs and expectations on which the student's participation is predicated, culminating in the learning of which student achievement is simply the most evident socially-constructed and culturally-mediated correlate. Culture is not outside these things. It is in the combination of these and other elements that culture itself is constituted. Nor is culture a synonym for nationality. As several studies have shown, the culture of the classroom can be constructed differently within a particular country or school system. There are, however, cultural values and beliefs that frame each country's educational endeavours. International comparative research must do more than document cultural differences, it must accommodate them.

Acknowledgements

The literature review on which this paper is based was informed by the work of Harsh Suri (2002, and in preparation).

References