

# THE VALIDITY-COMPARABILITY COMPROMISE IN CROSS-CULTURAL STUDIES IN MATHEMATICS EDUCATION

David Clarke

University of Melbourne, Australia

*The pursuit of commensurability in international comparative research by imposing general classificatory frameworks can misrepresent valued performances, school knowledge and classroom practice as these are actually conceived by each community and sacrifice validity in the interest of comparability. The “validity-comparability compromise” is proposed as a theoretical concern with significant implications for international cross-cultural research in mathematics education. The paper uses current international research to illustrate various aspects of the issue and its consequences for the manner in which international research is conducted and its results interpreted. The effects are extensive and constitute essential contingencies on international comparative research in mathematics education.*

Key words: International Research, Mathematics Education, Comparability, Validity

## INTRODUCTION

This paper identifies key considerations affecting the conduct and utility of international comparative research. Central to the design of such research studies are the dual imperatives of validity and comparability. Unfortunately, as will be illustrated, these imperatives are inevitably in tension. This paper identifies, illustrates and discusses these tensions, utilising very specific examples from current international comparative research. It is argued here that any value that might be derived from international comparisons of curricula or classroom practice is critically contingent on how the research design addresses the dual priorities of validity and comparability. It is further argued that since these priorities act against each other, researchers undertaking international comparative research must find a satisfactory balance between these competing obligations.

Perhaps only the drive to categorise is more fundamental than our inclination to compare (cf. Lakoff, 1987). Indeed, the two activities are intrinsically entwined since the act of comparison involves the recognition of the distinctive attributes of the objects being compared, and these attributes represent a form of categorisation. It is a key premise throughout this paper that comparisons are undertaken between constructed representations, whose structure and attributes are reflective as much of the value system of the researcher as of the objects being represented for the purposes of comparison. That these constructions are encrypted in language and effected socially aligns the paper's theoretical orientation with socio-cultural theorists prioritising the role of language as mediating experience and action (Vygotsky (1978) being the obvious example). In this paper, commensurability is interpreted as the right to compare (cf. Stengers, 2011). And it is our central assertion that this right to compare cannot be assumed, but is contingent on our capacity to legitimise both the

act of comparison and the categories through which this act is performed. The need for such legitimisation has been raised for international comparisons of student achievement, but less frequently and less carefully for the cross-cultural comparison of curricula and classrooms. Some examples are cited in this paper of cross-cultural comparisons of limited legitimacy (eg participation or lesson structure) or legitimate comparisons framed at levels of granularity such as to limit or remove explanatory capacity (eg between-desks-instruction or student classroom talk). In each example, I have attempted to suggest how to maximise both the validity and the explanatory power of the categorisation schemes employed for purposes of comparison.

Critical in the legitimisation of these acts of comparison are the validity of the categories we employ and of the act of comparison itself. Much of the focus in this paper is on cultural validity, which is interpreted (with Säljö, 1991) as a key determinant of practice in the international settings we aspire to compare. Research designs, especially data generation and categorisation processes, can misrepresent or conceal cultural idiosyncrasies in the interest of facilitating comparison.

This paper considers this validity-comparability compromise in relation to both curriculum and classroom practice research. Curricular comparisons raise issues related to the structure of school knowledge and the aspirational character of valued performances. Comparisons of classroom practice foreground the performative realisation of school knowledge and introduce the teacher as curricular agent (among other roles), modelling, orchestrating, facilitating and promoting performances aligned with the educational traditions of the enfolding culture. Any cross-cultural comparative analysis faces the challenge of honouring the separate cultural contexts, while employing an analytical frame that affords reasonable comparison.

The paper utilises seven “dilemmas” to reveal some of the contingencies under which international comparative research might be undertaken. The issues raised by each dilemma are not mutually exclusive sets. Specific empirical examples from current international research provide the vehicle by which the entailments of each dilemma can be explored to identify areas of cross-cultural research requiring critical examination.

## **COMPARABILITY AND VALIDITY IN CROSS-CULTURAL STUDIES**

In an international comparative study, any evaluative aspect is reflective of the cultural authorship of the study.

Culture is thus what allows us to perceive the world as meaningful and coherent and at the same time it operates as a constraint on our understandings and activities. (Säljö, 1991, p. 180).

In seeking to make comparison between the practices of classrooms situated in different cultures, the most obvious comparator constructs become problematic.

**Dilemma 1: Cultural-specificity of cross-cultural codes**

Use of culturally-specific categories for cross-cultural coding (eg participation, mathematics).

In the Chinese adaptation of the research design for the Middle School Mathematics and Institutional Setting of Teaching (MIST) project, the decision was made not to use the Instructional Quality Assessment (IQA) (Silver & Stein, 1996), but instead to develop a local instrument for the evaluation of mathematics classroom instruction. The reason for the rejection of the IQA instrument for use in Chinese school settings reflected the embeddedness, within the instrument, of particular values characteristic of the cultural setting and educational philosophy of the authoring culture (USA). For example, for the measurement of students' participation in classroom instruction, new criteria are needed that accommodate the larger class size and norms of social interaction of the Chinese mathematics classroom. Figure 1 shows the criteria for evaluating the level of student participation in teacher-facilitated discussion in mathematics classes.

#### A. Participation

Was there widespread participation in teacher-facilitated discussion?

4	Over 50% of the students participated consistently throughout the discussion.
3	25 to 50% of the students participated consistently in the discussion OR over 50% of the students participated minimally.
2	25 to 50% of the students participated minimally in the discussion (that is, they contributed only once.)
1	Less than 25% of the students participated in the discussion.
N/A	Reason:

**Figure 1. Participation criteria from the *Instructional Quality Assessment (IQA)* instrument (Silver & Stein, 2003).**

In countries such as China and Korea, teachers in both primary and secondary schools make extensive use of elicited student choral response as a key instructional strategy (Clarke, 2010). In the lessons analysed from one Shanghai classroom, a large number of choral responses (~ 80) were used in each lesson. In the analysis of a classroom in Tokyo, there were a similar number of individual student public statements, but no evidence of choral response. Applying the IQA participation criteria (Figure 1), the regularity and frequency of the use of choral responses would characterise this classroom as participatory at a level comparable with the classroom in Tokyo. Yet the students in the Tokyo classroom participate primarily through individual contributions rather than choral response and the type of teacher-facilitated discussion and the nature of student participation in that discussion in the two

classrooms are sufficiently different to make their comparability with respect to participation highly questionable. The cultural authorship of research instruments and their cross-site legitimacy has implications for both data generation and interpretation and must be accommodated carefully through revision or replacement, or through reconception of the nature of the comparison being undertaken. While Dilemma 1 arises through the misapplication of a culturally-authored categorisation scheme, Dilemma 2 arises from the need to categorise at a level of granularity sufficiently large as to accommodate (mask or gloss over) more fine-grained differences between the objects of comparison. The difficulty here arises not because the category scheme is culturally-specific, but because the objects to be categorised are sufficiently disparate or varied at one level of granularity as to defy comparison, but able to be included in a more inclusive (less discriminating) categorisation scheme that then supports the act of comparison, but disengages the act of comparison from the more fine-grained descriptions that could explain the origins of identified differences.

### **Dilemma 2: Inclusive vs Distinctive**

Use of inclusive categories to maximise applicability across cultures, thereby sacrificing distinctive (and potentially explanatory) detail (eg. mathematical thinking, lesson structure).

In a recent study undertaken by the author and his colleagues, we compared the ways in which mathematics curricula are framed in Australia, China, Finland and Israel. We sought to identify the similarities and differences in the organisation of mathematics curricula in the four countries in terms of their aims, content areas and performance expectations. In particular, we investigated the ways in which “mathematical thinking” was framed through curricular statements.

The key documents analysed in this study were: the Victorian Essential Learning Standards (VELS), the Chinese Mathematics Curriculum Standards (CMCS), the Finnish National Core Curriculum (FNCC) and the Mathematics Curriculum (Israel) (MCI). The four curricula are structurally quite different and prioritise different performance types. The excerpts below capture some of these qualitative differences.

See mathematical connections and be able to apply mathematical concepts, skills and processes in posing and solving mathematical problems (VELS).

[Translation] Obtain important mathematics knowledge that is essential for functioning in society and further development (including mathematical facts and experience in participating in mathematics activities) and basic mathematical thinking skills as well as essential skills of application (CMCS).

The task of instruction in mathematics is to offer opportunities for the development of mathematical thinking, and for the learning of mathematical concepts and the most widely used problem-solving methods (FNCC).

[Translation] Mathematics is not only a collection of calculated algorithmic operations that serve an applied purpose but also a subject with its own structure that includes unique thinking and investigation methods. The goal of the curriculum is to generate a change in the way that students view the subject (MCI).

Any attempt to characterise the relative emphasis given to particular types of valued performance at different grade levels can only be undertaken if a common classificatory framework can be imposed on all curricula. But such a general framework must not be allowed to mask the significant emphasis given to Geometry in grades 7 to 9 in China, or to “Communicating” in grades 3 to 5 in Finland, or the idiosyncratic prioritizing in grades 7 to 9 in Israel of “the evolution of phenomena from the perspective of mathematics.” The danger is that the commensurability demands of such comparisons conceal major conceptual differences in the curricular expression of categories of school knowledge. For example, it can be argued that the curricula in Australia, China, Finland and Israel are similar in that they advocate the development of “mathematical thinking” but this conceals important differences in the nature of the mathematical thinking that each curriculum seeks to promote. The act of developing more inclusive categories by combining more fine-grained culturally-specific categories in order to enable cross cultural comparisons runs the risk of distorting the knowledge categories we seek to compare. In cross-cultural research the imposition of an “external” classification scheme for the purposes of achieving comparability can sacrifice validity by concealing diversities reflecting cultural characteristics and by creating artificial distinctions. Comparability is achieved through processes of typification and omission, and each has the potential to misrepresent the setting.

### **Dilemma 3: Evaluative Criteria**

Use of culturally-specific *criteria* for cross-cultural *evaluation* of instructional quality (eg. Student spoken mathematics).

Where research is specifically constructed to be evaluative, the question arises as to the legitimate application of criteria developed in one culture to the practices of another culture. The use of evaluative criteria posits an ideal of effective practice that should be substantiated by reference to research. Problems arise when the research on which a criterion is based is itself culturally-specific.

For example, despite the emphatic advocacy in Western educational literature, classrooms in China and Korea have historically not made use of student-student spoken mathematics as a pedagogical tool. In research undertaken by Clarke, Xu and Wan (2010), classrooms were identified in which student spoken mathematics was purposefully promoted in public but not in private interactions (eg Shanghai classroom 1), in both public and private interactions (eg Melbourne 1) and in neither

public nor private interactions (eg Seoul 1). Each of these classrooms models a distinctive pedagogy with respect to student spoken mathematics.

If the occurrence of student-spoken mathematics is identified with quality instruction, then the instructional practice of the classroom in Seoul would be judged to be deficient. The classrooms in Shanghai and Melbourne differed significantly in the extent to which private student-student interactions were encouraged, but the teachers in both classrooms prioritized student facility with spoken mathematics. In the Shanghai classroom, promotion of this capability was developed solely through public discourse, whereas in the Melbourne classroom, private student-student mathematical speech was an essential pedagogical tool. Interestingly, in post-lesson interviews, the students from Melbourne and Shanghai showed comparable fluency in their use of the language of mathematics, while students from the classrooms in Seoul showed little evidence of such a capacity. The comparability issue here is whether it is legitimate to undertake evaluative comparison of frequency and sophistication of student spoken mathematics in Melbourne, Shanghai and Seoul, if the operative pedagogy in the Seoul classrooms does not actually value student facility in spoken mathematics as a learning outcome and therefore cannot be presented legitimately as having “failed” to develop this capacity in the students. Evaluative judgments of instructional quality made in the context of international comparative research must justify the model of accomplished practice implicit in the criteria employed and provide evidence of the cross-cultural legitimacy of these criteria.

#### **Dilemma 4: Form vs Function**

Confusion between form and function, where an activity coded on the basis of common form is employed in differently situated classrooms to serve quite different functions (eg kikan-shido or between-desks-instruction).

Kikan-shido (a Japanese term meaning “between-desks-instruction”) has a *form* that is immediately recognisable in most countries around the world. In kikan-shido the teacher walks around the classroom, while the students work independently, in pairs or in small groups. Although kikan-shido is immediately recognisable to most educators by its form, it is employed in classrooms around the world to realise very different functions. A teacher undertaking kikan-shido in Australia, will do so with very different purposes in mind from those pursued by a teacher in Hong Kong, or, for example, a teacher in Japan. In reporting the frequency of occurrence of an activity such as kikan-shido for the purposes of comparative analysis, the researcher conflates activities that are similar in form but which may be employed in differently-situated classrooms for quite distinct functions. Such conflation can create an impression of similarity although differences in practice are actually quite profound (for more detail, see Clarke, Emanuelsson, Jablonka & Mok, 2006). The distinction between this Dilemma and those preceding it, derives from the combination of the

legitimate equivalence of form (between-desks-instruction) across cultural sites, and the disparity of function (rather than of yet more fine-grained form) that exists at the same level of granularity between sites (rather than between different levels of granularity across all sites).

### **Dilemma 5: Linguistic Preclusion**

Misrepresentation resulting from cultural or linguistic preclusion (eg Japanese classrooms as underplaying intellectual ownership).

The analysis of social interaction in one culture using expectations encrypted in classificatory schemes that reflect the linguistic norms of another culture can misrepresent the practices being studied. This can occur because characteristics of social interaction privileged in the researcher's analytical frame may not be expressible within the linguistic conventions of the observed setting. For example, the Japanese value implicit communication that requires speaker and listener to supply the context without explicit utterances and cues. This tendency is typically found in leaving sentences unfinished. As a consequence, in Japanese discourse, agency or action are often hidden and left ambiguous. In English, when introducing a definition, the teacher might employ a do-verb: "We define". In a Japanese mathematics classroom, the teacher often introduces a definition in the intransitive sense (*Sou Natte Iru* = "as it is" or "something manifests itself") as if it is beyond one's concern. Such differences in the location of agency, embedded in language use, pose challenges for interpretive analysis and categorisation of classroom dialogue.

### **Dilemma 6: Omission**

Misrepresentation by omission, where the authoring culture of the researcher lacks an appropriate term or construct for the activity being observed (eg. Pudian).

The Sapir-Whorf hypothesis suggests that our lived experience is mediated significantly by our capacity to name and categorise our world.

We see and hear . . . very largely as we do because the language habits of our community predispose certain choices of interpretation (Sapir, 1949).

Marton and Tsui (2004) suggest that "the categories . . . not only express the social structure but also create the need for people to conform to the behavior associated with these categories" (p. 28). Our interactions with classroom settings, whether as learner, teacher or researcher, are mediated by our capacity to name what we see and experience. Speakers of one language have access to terms, and therefore perceptive possibilities, that may not be available to speakers of another language. For example, in the Chinese pedagogy "Qifa Shi" (Cao, Clarke, & Xu, 2010), the activity "Pudian" is a key element. Pudian can take various forms: Connection, Transition,

Contextualising, but its function is to help students develop a conceptual, associative bridge between their existing knowledge and the new content. There is no simple equivalent to Pudian in English, although teacher education programs delivered in most English-speaking countries would certainly encourage the sort of connections that Pudian is intended to facilitate. Many such pedagogical terms have been collected in a variety of languages (Clarke, 2010), describing classroom activities central to the pedagogy of one community but unnamed and frequently absent from the pedagogies of other communities. It follows that an unnamed activity will be absent from any catalogue of desirable teacher actions and consequently denied specific promotion in any program of mathematics teacher education. It is also likely that such activities will go unrecognised in reports of cross-cultural international research, where the authoring culture of the research report lacks the particular term.

### **Dilemma 7: Disconnection**

Misrepresentation through disconnection, where activities that derive their local meaning from their connectedness are separated for independent study (eg. teaching and learning (cf obuchenie), public and private speech).

Whether we look to the Japanese “gakushu-shido”, the Dutch “leren” or the Russian “obuchenie”, we find that some communities have acknowledged the interdependence of instruction and learning by encompassing both activities within the one process and, most significantly, within the one word. In English, we dichotomise classroom practice into Teaching or Learning. One demonstration of the consequences of the inappropriate disconnection of actions that should be seen as fundamentally connected is evident in the comparison of two published translations involving Vygotsky’s use of the term “obuchenie” (discussed in Clarke, 2001).

From this point of view, *instruction* cannot be identified as development, but properly organized *instruction* will result in the child's intellectual development, will bring into being an entire series of such developmental processes, which were not at all possible without *instruction* (Vygotsky, as quoted in Hedegaard, 1990, p. 350).

From this point of view, *learning* is not development; however, properly organized *learning* results in mental development and sets in motion a variety of developmental processes that would be impossible apart from *learning* (Vygotsky, 1978, p. 90).

The analogous disconnection of public and private speech in classrooms, and of speaking and listening (Clarke, 2006) has the same effect of misrepresenting activities that may be fundamentally interrelated (not just conceptually but also functionally connected) in their enactment in particular classroom settings.

## CONCLUSIONS

The pursuit of commensurability in international comparative research by imposing general classificatory frameworks can misrepresent valued performances, school knowledge and classroom practice as these are actually conceived by each community and sacrifice validity in the interest of comparability. In this paper, the “validity-comparability compromise” has been proposed as a theoretical concern that has significant implications for international comparative research. The identified dilemmas offer different perspectives and illustrate some of the consequences of ignoring this central concern. Some dilemmas are not specific to cross-cultural comparative contexts (eg inclusive vs distinctive - Dilemma 2), but should be considered any time difficulties caused by diversity at the intended level of comparison can be avoided by undertaking the comparison at a larger (more inclusive) level of granularity, sacrificing distinctive (and potentially explanatory) detail in the interests of legitimate comparison. Other dilemmas are a direct consequence of the idiosyncrasies of culture and language (eg Linguistic Preclusion or Omission – Dilemmas 5 and 6). The examples also illustrate the steps that can be taken to improve the legitimacy of the intended comparison. Partnerships with those being compared can minimise misrepresentation, but the necessity of the compromise is inescapable. The interpretation and application of international comparative research is critically contingent on researchers’ capacity to address those “dilemmas” pertinent to their particular design. As they have been framed in this paper, the dilemmas relate to methodological concerns. However, each dilemma can also serve as an interrogatory instrument: a tool directing the researcher’s attention to salient characteristics that, while presenting impediments to comparison, simultaneously provide insight into nuances of meaning and practice. This paper is intended to fuel a wider engagement in the critical interrogation of international comparison as a socio-material knowledge practice.

## REFERENCES

- Cao, Y., Clarke, D. J., & Xu, L. H. (2010). Qifa Shi Teaching: Confucian Heuristics. In M. M. F. Pinto & T. F. Kawasaki (Eds.), *Proceedings of the 34<sup>th</sup> Conference of the International Group for the Psychology in Mathematics Education, Vol. 1*, pp. 232-234. Belo Horizonte, Brazil: PME.
- Clarke, D.J. (2001). Teaching/Learning. Chapter 12 in D. J. Clarke (Ed.). *Perspectives on practice and meaning in mathematics and science classrooms*. Kluwer Academic Press: Dordrecht, Netherlands, 291-320.
- Clarke, D.J. (2003). International Comparative Studies in Mathematics Education. Chapter 5 in A.J. Bishop, M.A. Clements, C. Keitel, J. Kilpatrick, and F.K.S. Leung (eds.) *Second International Handbook of Mathematics Education*, Dordrecht: Kluwer Academic Publishers, 145-186.

- Clarke, D.J. (2006). Using International Comparative Research to Contest Prevalent Oppositional Dichotomies. *Zentralblatt für Didaktik der Mathematik*, 38(5), 376-387.
- Clarke, D.J. (2010). The Cultural Specificity of Accomplished Practice: Contingent Conceptions of Excellence. In Y. Shimizu, Y. Sekiguchi, & K. Hino (Eds.). *In Search of Excellence in Mathematics Education. Proceedings of the 5<sup>th</sup> East Asia Regional Conference on Mathematics Education (EARCOME5)*, Tokyo: Japan Society of Mathematical Education, pp. 14-38.
- Clarke, D.J., Emanuelsson, J., Jablonka, E., & Mok, I.A.C. (Eds.). (2006). *Making Connections: Comparing Mathematics Classrooms Around the World*. Rotterdam: Sense.
- Clarke, D. J., Xu, L., & Wan, V. (2010). Spoken mathematics as a distinguishing characteristic of mathematics classrooms in different countries. *Mathematics Bulletin – a journal for educators (China)*, Volume 49, Special Issue, 1 -12.
- Hedegaard, M. (1990). The zone of proximal development as basis for instruction. Ch. 15 in L.C. Moll (Ed.) *Vygotsky and Education*. Cambridge: CUP.
- Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
- Marton, F. & Tui, A.B.M. (2004). *Classroom Discourse and the Space of Learning*. Mahway NJ: Erlbaum.
- Säljö, R. (1991). Introduction: Culture and Learning. *Learning and Instruction Vol. 1*, 179-185.
- Sapir, E. (1949). *Selected writings on language, culture and personality*. Berkeley: University of California Press.
- Silver, E. A. & Stein, M. K. (1996). The QUASAR Project: The Revolution of the Possible in Mathematics Instructional Reform in Urban Middle Schools. *Urban Education*, 30(4), 476-521.
- Stengers, I. (2011). Comparison as a matter of concern. *Common Knowledge* 17(1), 48-63.
- Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.