TIMSS International Curriculum Analysis and Measuring Educational Opportunities

Richard T. Houang and William H. Schmidt
Michigan State University

Abstract

Following the long tradition of attending to the importance of educational opportunity in the understanding of educational progress, 1995 TIMSS International Curriculum Analysis (ICA) was designed to capture country's curriculum standards as *intended curricula* and their textbooks as *potential implemented curricula*. The vision of education opportunity was that of the IEA Tri-Partite Model of Curriculum. Since then, numerous articles and books have been published expanding on the knowledge gained from this study. This paper briefly reviews the ICA dataset and the subsequent advances made toward relating educational opportunity to educational gains.

Keywords: curriculum analysis, secondary analysis, educational opportunity, opportunity to learn, educational policies,

Introduction

In this paper, the TIMSS International Curriculum Analysis (ICA) and its use in research on educational opportunity is reviewed. TIMSS followed the long tradition of International Association for the Evaluation of Educational Achievement (IEA) of connecting educational opportunity and educational progress. Its focus was on the educational opportunities countries afforded their students. Opportunity To Learn (OTL) is framed as a content coverage variable without specific regards to allocated time, specifically, "on the content of what is being taught, the relative importance given to various aspects of mathematics and the student achievement relative to these priorities and content… (Travers & Westbury, 1989)".

"Opportunity" as used in TIMSS ICA is defined as students' encountering an occasion or potential experience in a classroom to study and learn about particular topics. Subject matter specific content is the defining element of an educational opportunity.

Sociologists developed opportunity to learn into a concept not unlike that found in the IEA studies. They, however, also attached other instructional effects – such as, teacher quality, and teaching methods – to their conception of OTL. Eventually, the confluence of both formulations of OTL became appealing to US policy makers (McDonnell, 1995). Although the definition of opportunity used in ICA is more constrained, this broader conception must be taken into account in any discussion of equality of educational opportunity.

The IEA Tri-Partite model of Curriculum provides the conceptual basis for the instrumentation used in TIMSS ICA. The Tri-Partite model defines curriculum at three different levels: the *Intended* – what a system intends students to study and learn; the *Implemented* – what is taught in classrooms; and the *Attained* – what students are able to demonstrate that they know. Using the TIMSS Curriculum Frameworks, which provides an exhaustive list of school topics in mathematics or science, ICA establishes procedures and instruments to encode curriculum documents and textbooks from more that 40 countries. This paper concludes with an example that uses the TIMSS ICA database to characterize educational opportunity and establish its relationship to student achievement.

INVESTIGATING CURRICULAR INTENTIONS

Curriculum is the most fundamental structure for educational experiences. It is a kind of underlying "skeleton" that gives characteristic shape and direction to instruction in educational systems around the world. The daily experiences in the classroom are deliberately shaped based on visions of what education should be, ideas of how to create the formative experiences of education, and intended patterns of opportunities that organize the potential for those experiences. The myriad details of classroom life often hide the fact that instructional activities are planned and orderly implementations of the aims and intentions of educational authorities. Curriculum provides a basic outline of planned and sequenced educational opportunities, a structure imposed by authority for the purpose of bringing order to the conduct of schooling.

Experience and Intentionality

The history of goal-driven group educational attainments makes it clear that much that guides educational experience is alterable. For a category of students (participants in a specific educational system, of a similar age, etc.), similar opportunities (backgrounds, educational settings, tasks, activities, materials, and involvement) typically lead to similar educational attainments. In specific situations, educational attainments are relatively predictable, and the distribution of such attainments is often fairly narrow for similar students. Thus, it is reasonable to infer that the experiences of individual students lying behind specific attainments are shared, common experiences. Once the idea of approximately common experiences is accepted, then examination of those elements that produce common experiences can occur. These elements reflect the intentions of those who created them.

Curriculum Analysis, then, is a study of how curriculum orders student experiences in schooling and of the structures imposed by formal authorities in a variety of national educational systems. It does not examine the direct mechanisms for imposing such order. Rather, it examines some sets of curricular data and materials (curriculum guides and textbooks) for patterns of reflected intentions and for commonalities and diversities of order apparent in materials from various educational systems. It also investigates what may be inferred about common and differing intentions.

In summary, in this paper, curriculum is only considered in its narrower, policy-malleable aspects. It is intentional and related to student experiences in schooling. The aims and intentions of policy makers and curriculum makers articulate visions and seek to guide experiences.

Potential Experience and Opportunity

While ICA describes how national intentions are communicated to help guide students through the course of their studies, acts of curricular intentions are worth investigation in themselves. In fact, as they help determine students' futures, these first stages — the goals and sequences of intended student experiences — *must* be investigated.

The term *opportunity* is used to convey that which is directly manipulated by educational policies and by curricular materials (especially curriculum guides and textbooks), and has certain political connotations, particularly when coupled with the word *education.* It thereby becomes associated with a complex set of values regarding equality and other social norms. Every choice shaping schooling provides some opportunities at the expense of others. The opening tasks of curriculum making involve using vision and insight to define, choose, and sequence opportunities. National choices are clearly the concerns of individual nations and reflect their particular visions.

Curriculum and IEA

TIMSS ICA was a natural extension of the informal analyses of curriculum guides and textbooks found in earlier studies of the IEA. For example, in the Second International Mathematics Study, researchers in each country rated achievement items according to their appropriateness for their nation's mathematics curriculum: many did so by consulting curriculum guides and textbooks. TIMSS ICA employs the framework of IEA Tri-Partite Model of Curriculum and developed more thorough, rigorous, data-gathering methods to document what countries intend to teach.

- Curriculum is considered as *intended*; this is reflected in official documents articulating national policies and societal visions, educational planning, and official or politically sanctioned for educational objectives.

- At the level of teacher and classroom activity, curriculum is considered as *implemented* intentions and objectives.

- At the level of student outcomes, the curriculum is considered as *attained* — the result of what takes place in classrooms. Academic achievement and student belief measures document part of these student attainments.

Defining Curricular Intentions: Curriculum Guides and Textbooks

Curriculum guides articulate official policies as they apply to large groups of students — to all students in a certain grade, in a specific type of school, etc. Each is an official, broad, guiding statement for what a curriculum is intended to be in a specific context. In part, textbook authors write to support implementation of national intentions.

Textbooks have official status in some countries clearly reflecting official curriculum. In other countries, textbooks are developed more independently and may reflect varying interpretations of curricular intentions. Their status varies correspondingly, ranging from official documents supplementing curriculum guides to completely unofficial materials for implementing instruction.

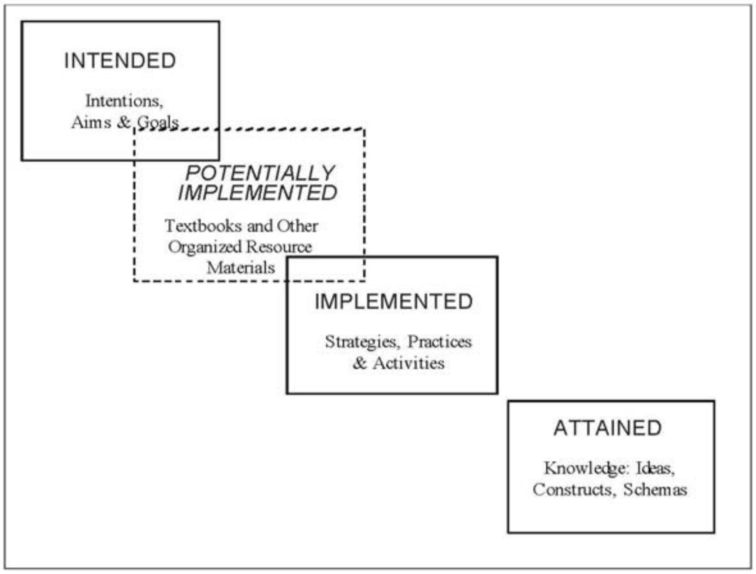
Curriculum guides almost always officially define curriculum for large groups of students, but in many countries, students are not aware of these documents and teachers may not regularly consult them. By contrast, evidence from TIMSS suggests that textbooks were present in almost every classroom in the participating countries and regularly used in instruction. Thus, while textbooks were often not *officially* national in character, they were a common element in most classrooms.

In these same countries, curriculum guides varied in design, function, and detail as much as did the government policies that they articulated. Guides varied so greatly, both between and within countries that a curriculum guide in one country might be unrecognizable as such in another.

While the textbooks' official status, design, and use varied greatly among the TIMSS countries, they were an instantly recognizable resource in classrooms in every country. Typically, textbooks provide a more detailed map of mathematical domains, topics, and performances. They are the "potentially implementable" curriculum (Schmidt, et.al., 1997b). They provide one strategy for navigating among mathematical topics and skills and to support a range of varied implementations within a broad reflection of curricular intention. Figure 1 suggests that textbooks serve as intermediaries in implementing curricular intentions.

***Figure 1.* Textbooks — The Potentially Implementable Curriculum.**

Textbooks served as intermediaries in turning intentions into implementations. They helped make possible one or more potential implementations of curricular intentions.



From Schmidt, et.al, 1997b

Methodologically, TIMSS considered curriculum guides as providing the primary — and purest — official, public statement of curriculum and curricular intentions. For many countries, textbooks were seen as providing a supplemental, supporting reflection of those intentions; they were viewed as a primary bridge between intention and implementation, between the ministry and the classroom. While care must be taken in interpreting textbook data — given different national status for and uses of textbooks — research conducted in many countries indicates the enormous influence textbooks have on student achievement. Nonetheless, neither curriculum guides nor textbooks could stand alone as the sole source of data for the curriculum analyses.

AN APPROACH TO CURRICULUM ANALYSIS

As conceptualized, the scope of TIMSS curriculum analysis effort was broad **-** to delineate the visions and aims helping shape mathematics curricula in close to 50 nations. In smaller scale projects, the “artifacts” of curricular visions — curriculum guides, textbooks, tests, and materials — are commonly used. While these data provide only a partial glimpse of the curricular vision, they do allow certain inferences to be made about curricular intentions. Before TIMSS, curricular documents had never been used as the primary data source for a large-scale project. Due to resource limitations, curriculum studies rely almost solely on expert opinions. While TIMSS also suffered from resource limitations — for example, translation of all documents into a single common language was not possible — methodologies, described in the next section, were developed to investigate curricular intention through careful document analysis supplemented by expert opinion.

Ideally, a full study of curricular documents would include ministerial policy documents, curriculum guidelines, course syllabi, textbooks, syllabi for national examinations, teacher pedagogical plans as they interpret broader requirements, tests, and so on. In TIMSS, the attention, however, was limited to those central documents common to virtually all the participating countries— mathematics curriculum guides and student textbooks. (See Schmidt, et. al., 1997b, c, and the ICA User Guide for a complete list of the curriculum documents that served as the basis for these analyses.)   While the ICA was for both Mathematics and Science, the remainder of this paper will focus on Mathematics.

Analytic Methods

The analytic methods consist of three important components. The *TIMSS Mathematics Framework* codifies the contents and performance expectations of the school Mathematics. The *Topic Trace Mapping* delineates how each *Framework* content topic entered and left the mathematics curriculum across years of schooling. And *Document Analysis* encodes the content of a curriculum document using the *Framework*. The codification facilitates objective comparisons of the encoded documents in their contents and other aspects.

The TIMSS Mathematics Framework

The first methodological requirement was to develop a tool for segmenting and categorizing meaningful chunks of the curricular documents and TIMSS tests. A unified set of categories was developed and the resulting TIMSS mathematics framework focuses on three aspects of mathematics education and testing:

- *Content* refers simply to the mathematics topic area.

- *Performance expectation* refers to expected student performances — that is, common elements of completing mathematical tasks. This aspect specifies what the student is supposed to do with the mathematics content. Expectations were considered less culturally determined than postulating the cognitive processes of performing specific mathematical tasks.

- *Perspective* refers to broad themes that might underlie many tasks, for example, emphasizing the societal importance of mathematics through specific contents or tasks.

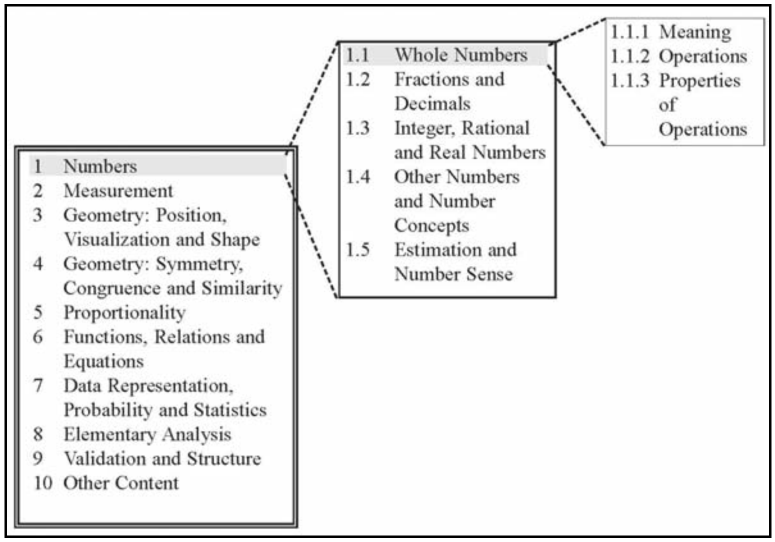
The frameworks were designed to be suitable for all participating countries and educational systems. Representing the interests of many countries, the frameworks were designed cross-nationally and received several iterations of review and revision. Each framework aspect is organized hierarchically using nested subcategories of increasing specificity. Within a given level, the arrangement of topics does not reflect a particular rational ordering of the content. (Figure 2 shows an overview of the content aspect of the mathematics framework.) Each framework aspect was meant to be encyclopedic in terms of covering all possibilities at some level of specificity. No claim is made that the "grain size" — the level of specificity for each aspect's categories — is the same throughout the frameworks. This varying granularity requires special care in designing framework-based methods and interpreting the results of their use.

In the mathematics framework, *content* involves 10 major categories, each with two to seventeen subcategories. Some subcategories are divided further (see Figure 2 for details). The level of detail and organization reflects a compromise between simplicity (fewer categories) and specificity (more categories). The hierarchical levels of increasing specificity allow some flexibility in detail level and generalization.

Complex signatures reveal important differences in how curricula are meant to achieve their goals. They show differences in how subject matter elements are combined — and differences in what students are expected to do. Each framework can reveal subject matter presented in an integrated, thematic way, with a rich set of performance expectations for students — as recommended by curriculum reformers in many countries. However, it also allows simpler signatures, for example, those often associated with more traditional curricula and many traditional achievement test items.

Figure 2 **Content Categories of the Mathematics Framework.**

Each aspect of the framework contains a set of main categories. Each main category contains one or more levels of more specific sub-categories. This figure shows the main content categories, with some sub- categories expanded to provide a better insight into the framework's structure.



From Schmidt, et.al, 1997b

Topic Trace Mapping

TIMSS achievement testing took place in each country the two consecutive grades with the majority of 9-yearolds (hereafter called Population 1), the two grades with the majority of 13-year-olds (Population 2), and the mathematics specialist courses in the final year of secondary school (Population 3). Some curricular information is essential for all grade levels, however. It was decided to obtain these data for all content topics in the mathematics framework. Because this action was feasible only by using expert opinions rather than directly by document-based methods (except for a few mathematics topics covered in depth), each country was asked to designate a panel of experts to determine each grade in which each topic was covered. The experts were asked to base their responses on curriculum guides and other curricular materials, but their methods were not directly monitored.

For each topic, the experts were asked to provide two kinds of information. First, they were to indicate all grades in which aspects of that topic were covered. Second, they were asked to indicate at which grade the topic was the focus of more extensive or intensive curricular attention. Experts provided similar data for all of the different major geographical regions in their countries and for differing streams within their curricula. These coverage and focus data were collected for mathematics content topics only; results of pilot work indicated that this method did not provide reliable data on performance expectations or perspectives. This collection effort cumulated in a map of the way in which each content topic entered and left the mathematics curriculum across years of schooling. Any special focus or curricular attention was also mapped. To ensure reliability, countries were asked to provide their data more than once, to review the data they had provided, and to answer direct questions about any apparent discrepancies.

Figure 3 shows "tracings" or "maps" of the intended curricular coverage for two representative mathematics topics in a sample of countries selected to illustrate the diversity typical of these data. A "-" indicates aspects of the topic that were typically part of mathematics instruction for students at that age, according to the experts. A "+" indicates aspects of that topic received a "focus" — special curricular attention and emphasis of some sort — for students at that age.

Figure 3. **Two Representative Topic Trace Maps.**

These data are typical topic trace maps for a sample of countries selected to show representative diversity. The results are typical of those for other topics and countries.

Congruence and Similarity

	Age 6	7	8	9	10	11	12	13	14	15	16	17	18
Argentina									+				
Canada	-	-	-	-	-	-	-	-	-	-	-	-	
Cyprus							+		-	+			
Denmark							+	-	-	+	+	-	
Hungary			-	-	-	-	-	+	-	-	+	-	
Iceland								+	-	-	+	-	
Iran						+	-	-	+	-	-		
Ireland								+	+	+		-	+
Japan					+	+		+	-				
New Zealand							-	-	-	-	-		
Spain						+	+						
Tunisia								+	-	-	-	-	+
USA						-	-	+	-	+			

Equations and Formulas

	Age 5	6	7	8	9	10	11	12	13	14	15	16	17	18
Argentina											+	+	-	
Canada					-	-	-	-	+	+	+	+	+	-
Cyprus		-	-	-	-	-	-	-	+	+	+	+	+	
Denmark			-	-	-	+	-	-	+	+	+	+	+	-
Hungary				-	-	+	-	-	+	-	-	+		
Iceland									-	-	+	-	+	-
Iran								+	-	-	-	+		
Ireland	-	-	-	-	-	-	-	-	+	+	+		+	+
Japan				+	+	+	+	+	+	+	+	+	+	
New Zealand	-	-	-	-	-	-	-	-	-	+	+	+	+	
Spain								-	-	-	-	-	-	
Tunisia										-	+	+	+	+
USA			-	-	-	-	-	+	+	+	-	+	+	

Legend: - indicates intended coverage at that age.
+ indicates intended focal coverage at that age.

From Schmidt, et.al, 1997b

'Congruence and similarity' refers to congruent and similar geometric figures — especially triangles, quadrilaterals, and polygons — and their properties. Figure 3 shows, for example, that content from 'congruence and similarity' received only two focused years of coverage for students ages 11 and 12 in Spain. By contrast, Hungarian experts indicated that 10 years of study were intended for this content beginning with students at age 8 with special emphasis at ages 13 and 16. Other patterns emerged, ranging from countries planning only one year on this topic in upper secondary (Argentina) to those that intended coverage of the topic almost from the beginning of primary to the end of upper secondary school (Canada and Hungary). Other countries showed patterns of "interrupted" or multi-sited coverage, e.g., Cyprus and Japan. The patterns in Figure 3 are typical of other countries and many other topics.

'Equations and formulas' provides a clear contrast. This topic includes the elementary representation of numerical situations; formulas; algebraic expressions including linear, quadratic, and polynomial equations; systems of equations; radical equations; inequalities; and systems of inequalities — in short, all equation-related algebraic content. Experts indicated that this topic was taught from very early in primary to the end of advanced secondary school in most countries, a logical finding for a topic so inclusive. Ireland and New Zealand, for example, began this introduction very early. A few countries, for example, Iceland, Tunisia, and Iran, waited until comparatively late.

Document Analysis

Given practical and economic constraints, document analysis was conducted using a sample of documents. The sampling design had to balance several conflicting concerns and address the need to

- obtain some information on intentions at all grade levels;
- obtain detailed information at some grade levels — particularly at the focal grades at which achievement testing would occur; and
- be sensitive to important national differences in educational governance and textbook role, and to curricula and textbook differences within countries.

The grades selected for document collection were those for which TIMSS achievement testing would take place. Each national document sample required selecting curriculum guides and textbooks to represent those used for at least half of the students in the targeted grades. Major regional, school type, and other strata were to be represented appropriately in the sample. National curriculum guides were sampled if they existed; otherwise, the sample consisted of appropriate guides for each grade that were regional, provincial, or the like. Similarly, official national textbooks were sampled if they existed; otherwise, the most widely used commercial textbooks were selected. The documents included for each country were chosen by national representatives in consultation with TIMSS. Documents were marked for analysis and coded (without translations) for the three aspects of the TIMSS mathematics framework in the individual countries.

Next, informational questionnaires in English were provided, and marked archival copies of the documents were sent to the curriculum analysis center for filing and referencing. Forty-eight countries reported at least some document-based curriculum data. The database consist of 241 mathematics curriculum guides and 318 textbooks and — 251 science curriculum documents 312 textbooks in use in 1990-91, a total of 1,122 documents.

The documents were written in over 30 different languages. In lieu of translation, each country used its own team to mark and code important features of their documents. Pilot work and field trials identified features that were important for capturing document intent, structure, and essential content.

Lengthy documents were divided into smaller segments in two stages. First, each document was separated into a series of *units* to capture major structures. The unit type was determined and recorded. To capture more finely grained structures in the documents and to attach framework codes to small, narrowly defined segments, each unit was subdivided into one or more segments called *blocks.* For a listing and detailed description of the unit and block types coded, see Schmidt, et. al., (1997b, 1997c) and the ICA User Guide.

Next, category codes for mathematical content, performance expectations, and perspectives were assigned to blocks. These data were recorded on appropriate forms and cross-referenced to marked and numbered segments of the original documents. All marked documents, coding forms, and supplemental questionnaires were sent to the curriculum analysis center for data entry and archiving.

These methods involved considerable decision making by each country's document analysis team. To ensure uniformity and reliability, detailed manuals were prepared for every procedure. Face-to-face training was provided to representatives of every country through a series of regional training sessions lasting almost 3 days each. An initial quality assurance phase required national teams to code a selected random sample of units, which were then translated and coded by international referees. Failures to meet satisfactory criteria in coding were noted and analyzed. Only when the criteria had been met (by further practice, training, coding new samples, etc.) were countries authorized to begin the main analysis. All received materials were carefully examined and discrepancies between documents and forms pursued individually to obtain clarifications. Finally, after the collection was complete, a further sample of submitted materials was selected, translated, and coded for each country. An appropriate reliability estimating procedure was developed and applied. In virtually every case, concordance of over 80 percent (and often over 90 percent) resulted. (Schmidt, et. al., 1997b).

CHARACTERIZING EDUCATIONAL OPPORTUNITIES

Three reports were published based on the TIMSS ICA Database. *Many Visions, Many Aims, Volume 1 and 2* (Schmidt, et. al., 1997b, 1997c) contrasted school mathematics and science curricular intentions cross-nationally. *According to the Book* (Valverde, et. al., 2002) focused on mathematics and science textbooks.

In addition, the analytic procedures to extract content coverage and performance expectations from documents have been used in numerous studies: comparing standard assessment and district curricula (Schmidt, 1999a, 1999b), Longitudinal Study of American Youth (Miller, 2000), Adolescent Health (Schiller, et. al., 2007), Achieve Inc.(2004) examining high school exit exams, and most recently NAEP. The document analysis procedure was used to code textbooks and assessment instruments. Instead of coding every page of a textbook, a modified procedure that focused on introduction, summary and exercises was used. This has the advantage of requiring only 1/10 of the time normally needed to code a textbook. And comparison with the original coding procedure showed up to 80% agreement in aggregated coverage. The results were used to characterize the educational opportunities for the course that the textbook was used. The exhaustiveness of the Framework allowed detailed categorization of these courses. Coded textbooks for Algebra I courses, for example, revealed a great disparity in the content covered (Schiller, et. al., 2007). Similar results were also used as curricular indicators in structural equation modeling (e.g., Miller, 2004). For assessment instrument, items were coded as blocks and the content aggregated. The results were used for example in the California Validity Study, an alignment study contrasting district curricular guides and standardized assessments.

For the remainder of this paper, we focus on the use of Top Trace Maps in creating an international benchmark. And using the benchmark as reference, indicators of “coherence” and “focus” are defined and used in analyses relating to student achievement.

“World Class” Mathematics Curriculum

In TIMSS, US 4th-graders performed relatively better in mathematics than 8th-graders, and 8th-graders performed relatively better than 12th-graders (Beaton, et.al., 1996, Schmidt, et.al., 1999). Discussion of such achievement results have prompted US policy-makers, as well as those from other countries, to consider more carefully the curriculum portraits TIMSS has produced, especially those for the highest-achieving countries (Schmidt *et al*. 1997b, 1997c). This has led to an effort to discern just what it might mean to have a ‘world-class’ mathematics or science curriculum (Valverde and Schmidt 2000).

Valverde and Schmidt (2000) examined the content standards in mathematics and science of the TIMSS top-achieving countries. The common elements defined an ‘international’ set of standards which reflected those countries demonstrating excellence on the TIMSS Population 2 test, i.e. grades 7 and 8 in the US. The resulting international standards (benchmarks) were analyzed in terms of their coherence. Valverde and Schmidt termed the top-achieving countries identified for this study as the ‘A-plus’ (A+) countries. These countries had the highest mean middle-school student achievement (total score) without identifying more than five additional TIMSS countries that could be statistically equivalent to them; that is, to include the next lower-achieving country would make it necessary to include more than five additional countries that have statistically equivalent achievement. In mathematics, six such countries were identified—Singapore, Korea, Japan, Hong Kong, Belgium (Flemish), and the Czech

Republic; in science there were four such countries—Singapore, the Czech Republic, Japan, and Korea.

The data used to develop the international benchmarks were derived from the Topic Trace Maps. As mentioned above, Topic Trace Maps were completed by education officials (typically curriculum officers in the national ministry) of each nation who, using their national content standards or an aggregate of regional standards, indicated for each grade level whether or not a content topic was included in their country's intended curriculum. The result was a map reflecting the grade-level coverage of each topic for each country. Only topic coverage from grades 1–8 are presented here.

Topic trace maps were available for each of the A+ countries. While none were identical they all bore strong similarities. The following procedures were followed to develop an international benchmark. First, the mean number of intended topics at each grade level was determined across the countries. Next, the topics were ordered at each grade level based on the percentage of the A+ countries that included it in their curriculum. Those topics with the greatest percentage were chosen first, and only as many were chosen as were indicated by the mean number of intended topics at that grade level.

Figure 4 portrays the set of topics for grades 1–8 that represents only the common topics intended by a majority of the A+ countries. The data suggest a three-tier pattern of increasing mathematical complexity. The first tier, covered in grades 1–5, includes an emphasis primarily on arithmetic, including whole-number concepts and computation, common and decimal fractions, and estimation and rounding. The third tier, covered in grades 7 and 8, consists primarily of advanced number topics, including exponents, roots, radicals, orders of magnitude, and the properties of rational numbers, algebra, including functions and slope, and geometry, including congruence and similarity and 3-dimensional geometry. Grades 5 and 6 appear to serve as an overlapping transition or middle tier marked by continuing attention to arithmetic topics (especially fractions, decimals, estimation, and rounding), but with an introduction to the topics of percentages, negative numbers, integers and their properties, proportional concepts and problems, 2-dimensional co-ordinate geometry, and geometric transformations, all of which, except for percentages, were also topics found in the third stage. Thus, grades 5 and 6 serve as a point of transition where attention to topics such as proportionality and co-ordinate geometry led to the formal treatment of algebra and geometry that is characteristic of the third stage.

The implied curriculum structure also included six topics that provided a form of continuity across all three stages. These topics—measurement units; perimeter, area, and volume; algebraic equations, including the representation of numerical situations and the informal solution of simple equations; data representation and analysis; and basic two-dimensional geometry including points, lines, angles, polygons and circles—appear to insure stability across the three tiers, serving as 'buttresses' supporting the overall curriculum structure. Those 'buttresses' include the fundamentals of algebra, geometry, measurement and data analysis, and, by way of the implied breadth of these topics, could move from their most elementary aspects to the beginnings of complex mathematics.

The 'upper triangular' appearance of the display in Figure 4 implies a hierarchical sequencing of the topics in the top-achieving countries over the first eight grades. As indicated, this sequencing moves from elementary to more advanced topics in a way that appears to be based mostly on the inherent logic of the involved mathematics itself. Not only is the progression of the topics over grades logically consistent with the nature of the mathematics but that same progression culminates at 7th and 8th grade in more rigorous topics. These two characteristics combine to provide an example of the coherence and rigor.

***Figure 4*. Mathematics topics intended at each grade by a composite A+ curriculum.**

Topic	*Grade 1*	*Grade 2*	*Grade 3*	*Grade 4*	*Grade 5*	*Grade 6*	*Grade 7*	*Grade 8*
Whole number meaning	●	●	●	⊙	⊙			
Whole number operations	●	●	●	●	⊙	•		
Measurement units	○	●	●	●	●	●	⊙	•
Common fractions		•	○	●	●	⊙		
Equations and formulas	•	•	○	⊙	⊙	⊙	●	●
Data representation and analysis	•	•	○	○	⊙	⊙	•	○
2-D geometry: basics	•	•	○	⊙	⊙	⊙	●	●
Polygons and circles	•	•	•	○	⊙	⊙	●	●
Perimeter, area and volume			•	⊙	⊙	⊙	⊙	○
Rounding and significant figures			•	⊙	⊙	•	•	•
Estimating computations	•	•	•	⊙	⊙	⊙	•	
Properties of whole number operations			•	⊙	⊙	•	•	
Estimating quantity and size	•	•	•	○	○		•	
Decimal fractions			•	⊙	●	⊙		
Relationship of common and decimal fractions			•	•	●	⊙	•	
Properties of common and decimal fractions				•	⊙	⊙		•
Percentages					⊙	⊙	•	
Proportionality concepts					⊙	⊙	⊙	○
Proportionality problems					⊙	⊙	●	●
2-D coordinate geometry					⊙	○	⊙	⊙
Geometry: transformations						⊙	⊙	⊙
Negative numbers, integers and their properties						○	⊙	•
Number theory						•	⊙	○
Exponents, roots and radicals							⊙	⊙
Exponents and orders of magnitude					1		○	○
Measurement estimation and errors							○	•
Constructions w/ straightedge/ruler and compass							●	○
3-D geometry					•	•	⊙	●
Congruence and similarity							•	●
Rational numbers and their properties							•	○
Patterns, relations, and functions							•	○
Slope and trigonometry								○[2]

Intended by less than the majority. •
Intended by 4 out of the 6 top Achieving Countries ○
Intended by *all but one* of the A+ countries (5 out of 6). ⊙
Intended by *all* of the A+ countries. ●

Notes:
1. The data in table 1 indicate the inclusion of this topic at grade 5, but since it is not intended in grade 6, which results in a break in its coverage (the only additional topic for which this is the case) we did not include it here.
2. Two additional topics not included in the original 32 of figure 1—real number and validation and justification—are intended at grade 8.

From Valerde and Schmidt, 2000

Coherence and Student Achievement

"Coherence" is defined as follows:

> Content standards, taken together, are coherent if they are articulated over time as a sequence of topics and performances consistent with the logical and, if appropriate, hierarchical nature of the disciplinary content from which the subject matter derives. This is not to suggest that there is only one coherent sequence, but rather that any such coherent sequence reflects the inherent structure of the discipline. This implies that for a set of content standards 'to be coherent' they must evolve from particulars (e.g., simple mathematics facts and routine computational procedures associated with whole numbers and fractions) to deeper structures. It is these deeper structures by which the particulars are connected (such as an understanding of the rational number system and its properties). This evolution should occur both over time within a particular grade level and as the student progresses across grades. (Schmidt, et.al., 2005)

If content standards reflect the structure of a discipline, then the "depth" of those standards should *increase* as students move across the grades. Failure to increase in depth, sophistication and complexity across the grades would indicate a lack of coherence. Extensive repetition of virtually the same standards across grade levels is found in the US. This goes against the idea of "coherent" development, is unwarranted, and contributes to a lack of focus. Such repetition can be replaced with standards that form a trajectory by linking coverage of the topics over grades and by reducing the repetition over the same grade levels. Such an approach would represent a "continuing penetration of the discipline moving to a deeper structure that makes things 'simpler' in Bruner's (1995) terms". That is, coherent development is, in the long run, fundamentally simpler than virtual repetition without development in depth and sophistication and without attaching to fundamental, unifying ideas of the discipline. Schmidt and Houang (2007) found that these curricular characteristics related significantly to student achievement at the country level.

Indicators of Focus. The Topic Trace Map data for each country were mapped into the matrix defined by Figure 4. The rows and columns of Figure 4 were considered as fixed and remained constant. The resulting map for each country placed an indicator into various cells defined by the 32 by 8 matrix (256 possible cells or "opportunities to cover mathematics topics"). In effect, this map indicated what topics were covered at which grades for each country.

The region of the matrix with dots in Figure 4 was taken as a model or ideal scenario defining coherence. For each country, the number of "hits" within the model region was considered an indicator of the degree of coherence. One way to think of this process is to view the region with majority of top achieving countries (4 out of 6) creating a "silhouette". This silhouette was then superimposed on the country maps. Then, the degree of overlap was used to estimate the coherence of that country's curriculum. A high value on this statistic indicates that the part of the curriculum map of the country dealing with the silhouetted region is very similar to the ideal scenario.

As previously discussed, we are not arguing for only one model of coherence. Coherence could take on another form. Low values on the proposed indicator should not necessarily be interpreted in the absolute as a lack of coherence but rather as a deviation from the empirically derived ideal scenario of coherence presented in Figure 4. As Figure 4 is a composite of the top achieving countries, it does not represent the curriculum of any individual country, even those countries that contributed to its definition. Therefore, the "silhouette" serves a role similar to the role that a mean plays in other statistical analyses.

Since coherence is cumulative over the grades, its definition at specific grade levels demands that the 'upper triangle' of Figure 4 must be correspondingly partitioned. The eighth grade 'upper triangle' includes 99 cells in the matrix as represented by the silhouetted area. The portion of the 'upper triangle' that goes up to seventh grade includes 81 cells, excluding the 18 cells found in the triangle at eighth grade. In the same way, the fourth grade contains 28 of the cells and third grade only 13 cells.

The statistic used as an indicator of focus was a cumulative count of the total number of dots in the country map up to and including the grade in question. Large values indicated a lack of focus — typically the result of including a large number of topics in grades outside those needed to reproduce the empirical pattern of "coherence."

For any country, the 256 cells (32 by 8) in the matrix can be divided into three groups:

1. Cells that overlap with the 'upper triangle' area — those that reflect an exact match with the ideal scenario of coherence — and, as described above, a count of these is used as the indicator of coherence.

2. Cells 'before the upper triangle'**-** those topics that are introduced too soon compared to the ideal scenario.

3. Cells 'after the upper triangle'**-** those topics that are introduced or continued after

coverage in the ideal scenario is finished.

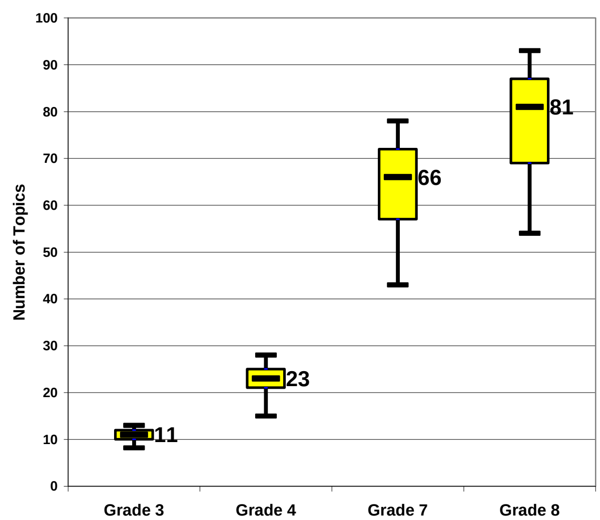
The sum of the three groups is used to define ***focus***. It indicates the total number of cells or topic/grade combinations intended to be covered by the country up to and including that grade. It is a cumulative index and as a result the matrix must be partitioned appropriately for each separate grade.

A second statistic for estimating focus is based on a count of the number of intended topic/grade combinations 'before the upper triangle.' A positive value of this indicator for intended coverage (number 2 above) reflects a ***lack of focus*** created by adding more topics to earlier grades. It is a particular kind of lack of focus —it implies an early treatment of topics where the necessary prerequisite mathematics would likely be concurrently covered or, what is even more problematic, not covered until a later grade. Examples of both kinds are found in the data. The latter is potentially more detrimental to student learning.

The outcome measure was the TIMSS scaled total mathematics score for the country. It was available at grades three, four, seven and eight. Country-level regression analyses were performed to examine the curriculum effects as defined by these indicators of coherence and focus. Complete data were available for 20 to 33 countries, depending on the grade level.

Figure 5 illustrates the distribution (by box and whisker plots where the extreme values (hinges) are the fifth and ninety-fifth percentiles) for the measure of coherence at each of grades three,

Figure 5. **Number of Topic/Grade Combinations Aligned with Ideal Scenario of Coherence**

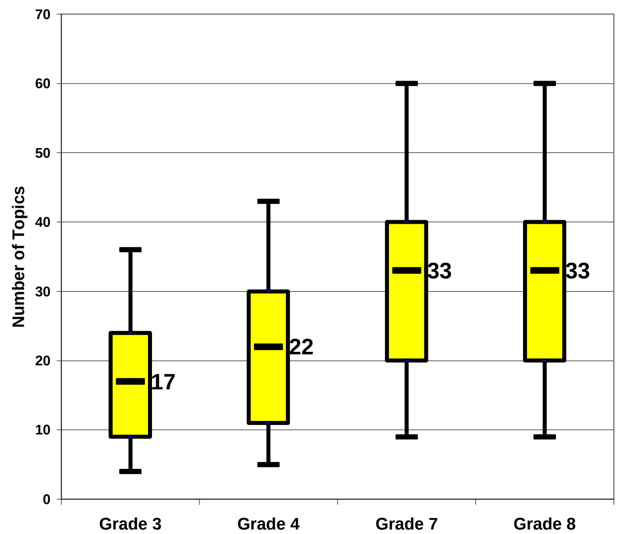


From Schmidt and Houang, 2007

four, seven and eight. The variability across countries in grades three and four, but especially grade three, is quite small. Remember that the total size of the coherence model at these two grades is 13 and 28, and, as seen in Figure 5, the median values are 11 and 23 respectively. However, at grades seven and eight the variability is much larger with an inter-quartile range at grade eight of almost 20. At these grades, no country achieves the maximum values.

Figure 6 represents the distribution of the specific measure of focus also across all four grades. No country has the value of zero indicating that all countries have some topic/grade combinations that occur before those specified in the coherence model. Some countries do have as few as ten or less such "premature" topic/grade combinations. The median values at grades seven and eight are 33 such combinations rising to as many as more than 60.

Figure 6. **Number of Topic/Grade Combinations Intended to be Covered Prematurely**



From Schmidt and Houang, 2007

From the point of view of mathematics, the variation portrayed in Figure 6 is extremely disturbing because it is in this area of the matrix where the hierarchical nature of mathematics can be most violated by the sequencing of school topics in mathematics.

Table 1 presents the number of topic/grade combinations in the three categories related to the 'upper triangle' as well as the sum of these — which is one of the focus measures- for a sample

TABLE 1
Alignment with Ideal Scenario of Coherence at 8th Grade for a Sample of Countries

Country	*Total Number of Topic/Grade Combinations in Curriculum*	*Number of Topic/Grade Combinations Not Aligned (Introduced Before the Ideal Scenario)*	*Number of Topic/Grade Combinations Aligned with Ideal Scenario*	*Number of Topic/Grade Combinations Not Aligned (Introduced Following the Ideal Scenario)*
Australia	136	32	75	29
Belgium (Fl)	155	37	87	31
Czech Republic	119	32	83	4
France	156	40	85	31
Germany	117	33	74	10
Hong Kong	79	9	62	8
Hungary	158	49	93	16
Japan	128	39	86	3
Korea	125	27	90	8
Netherlands	97	14	54	29
Norway	159	33	96	30
Portugal	144	48	84	12
Singapore	115	11	84	21
Spain	109	23	69	17
USA	186	63	98	25

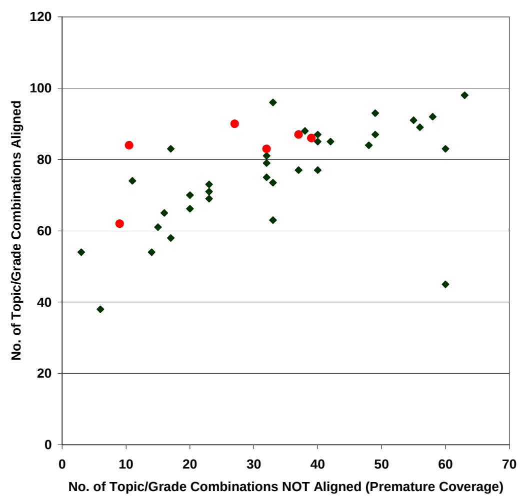
From Schmidt and Houang, 2007

of countries at grade eight. Using the data here one can see that the countries used to develop the model scenario are not necessarily the ones with the greatest degree of consistency with it. Hong Kong, for example, has only about 60 percent agreement with the model of coherence and the Czech Republic about 80 percent. Norway and the US have a higher value for the coherence measure than any of the six top achieving countries. The US also has the most topics introduced too early (63).

One concern with this methodological approach was whether using the top-achieving countries to help define the model scenario was "stacking the deck," in terms of the regression analyses to be presented. Table 1 helps to allay those concerns. The other factor that also helps alleviate the concerns is that the top achieving countries chosen were chosen based on their eighth grade performance. These were not all among the top achieving countries at grades three, four and seven. Thus, if there were bias, it would mostly show up at eighth grade.

The regression analyses relating the focus measure (the total number of topic/grade combinations intended) and the coherence measure, each separately with country level achievement, were not statistically significant. The significance levels were as follows: grade 3 ($p < .11$); grade 4 ($p < .13$); grade 7 ($p < .16$) and grade 8 ($p < .09$). Neither focus nor coherence considered alone was systematically related to country level achievement. This was true at all four grades. Coherence and focus were themselves correlated and Figure 7 shows the scatter plot of the measure of coherence with the focus measure at eighth grade that deals with early

***Figure 7*. Number of Topic/Grade Combinations NOT Aligned (Premature Coverage) vs. Number of Topic/Grade Combinations Aligned with Ideal Scenario of Coherence**



From Schmidt and Houang, 2007

coverage of the topics ($r = .77$). The estimated correlation coefficients for the other grades are: grade 3 ($r = .67$); grade 4 ($r = .76$) and grade 7 ($r =.75$). The six, top-achieving countries are noted by circles in Figure 7. We present this measure of focus graphically since, as stated previously, it represents the biggest threat to coherence and the largest contribution to the total focus measure (other than the count of the match with coherence).

Consider the results of the regression analyses relating achievement to coherence and focus with the latter being defined as the total number of topics intended up to the grade in question.

The coherence measure also is adjusted, as described previously, to each of grades three, four, seven and eight to correspond to the achievement test results at the same grades. These results are presented in Table 2.

The significance level of the model fitting varied from $p < .12$ to $p < .01$ across the four grades. At grades three and four, the significance level is greatly influenced by the small sample sizes of 20 and 21 countries (fewer countries in TIMSS participated at the lower grades). The estimated R^2 values indicated that around 20 to 25 percent of the variance was accounted for by both indicators. In all four instances, the focus measure was statistically significant, indicating that, when controlling for coherence, the larger the number of topic/grade combinations intended for coverage, the lower the mean country achievement.

TABLE 2
Regression Analyses relating Coherence and Focus to Achievement

Model	***Grade 3***			***Grade 4***			***Grade 7***			***Grade 8***		
Predictor	**Estimate**	**Std Erorr**	**p**	**Estimate**	**Std Erorr**	**p**	**Estimate**	**Std Erorr**	**p**	**Estimate**	**Std Erorr**	**p**
Number of Topic/Grade Combinations Aligned with Ideal Scenario	**9.94**	**7.38**	***0.196***	**7.48**	**4.04**	***0.081***	**3.57**	**1.48**	***0.023***	**2.83**	**1.12**	***0.017***
Total Number of Topic/Grade Combinations in Curriculum	**-2.42**	**1.13**	***0.047***	**-2.58**	**1.05**	***0.025***	**-1.64**	**0.57**	***0.008***	**-1.39**	**0.45**	***0.004***
Model Fit												
R-Square	**0.2191**			**0.2574**			**0.2205**			**0.249**		
Residual Mean Squares	**1652.9**			**1489.0**			**1476.0**			**1331.4**		
p<	***0.1222***			***0.0687***			***0.0270***			***0.0136***		
Number of Countries	**20**			**21**			**32**			**33**		

From Schmidt and Houang, 2007

The coherence measure was significant ($p < .02$) at grades seven and eight and only marginally so at the fourth grade ($p < .08$). Here the coefficients are all positive indicating the greater the degree of match of a country's intended curriculum to the ideal scenario, the higher the country's mean performance on the TIMSS test. The lack of significance at grade three is probably a combination of two factors. One is the small sample size. The other, and likely more important, reason is the small size of the ideal scenario (only 13 cells) and the corresponding lack of much variability at grade three for that scenario.

The results of these analyses at eighth grade imply that if a country were to increase its coverage of the ideal scenario by ten topic/grade combinations or opportunities intended to learn and hold constant at its current level of focus, this would predict an increase of about one-fourth of a standard deviation in mathematics achievement (or over a third of a standard deviation at seventh grade**-** the predicted increase at grades three and four is even greater but the precision of these estimates is marginal at best).

For the focus measure (where the variation is much larger as displayed in Figure 6) a decrease in the number of intended topic/grade combinations by 50 would predict an increase of almost three-fourths of a standard deviation. The later is especially relevant to the US where intending coverage of 50 fewer topic/grade combinations would not be difficult given the fact that the US has the highest total and a reduction of 50 would bring it more into line with Japan and Korea (see Table 1). The predicted effect for focus would be even greater at grades three, four and seven. For both coherence and focus, ignoring the marginal nature of the significance at grade four, the implication seems to be that the predicted effects are greater at the earlier grades. Perhaps premature coverage of topics has more of a deleterious effect at the earlier grades. This, of course, is only a hypothesis in need of further study.

The results of the above analyses are consistent with the hypothesis — especially at the upper grades where we found the measure of coherence to be positively related to achievement. There is also some indication that coherent standards are at least marginally related to student performance at fourth grade as well.

The other striking result is that even when controlling for coherence the focus measure was statistically significant at all grades. Even at third grade, where the overall model was not statistically significant, the focus measure was ($p < .05$). This implies that for a country to have high mean level performance it is not enough to have a high degree of coherence. The amount of 'clutter' to that vision created by covering topics too early or before their time from a mathematical point of view must be kept small. Covering too many topics does have a negative impact on student learning even when controlling for coherence.

As further evidence of the interplay between coherence and focus in terms of their joint relationship to achievement, the bivariate relationship between focus and achievement is not statistically significant at any of the four grade levels. In other words, focus by itself is not related to country level achievement but it is related to achievement (as reported above) only when coherence is included in the model. This implies that focus only has a statistically significant negative impact conditioned on coherence. Put simply, no matter what the degree of coherence evident for a country, covering topics at grades outside those defining coherence has a negative impact.

SUMMARY AND CONCLUSION

TIMSS ICA captures the curriculum from more than 40 nations and contains valuable information about cross-national curriculum. It provides a source of data that can be used to create international benchmarks and be used to assist in curricular reform. ICA also provides the methodology to capture the intended and implemented curriculum. The data have been successfully used to examine educational opportunity structure within and across educational systems to facilitate educational reform. When the TIMSS Curriculum Frameworks were employed to measure the curriculum at these different levels, comparisons could be made not only with respect to international benchmarks for each level at which curriculum was measured. This approach was adopted in a NSF funded 60 school district project in the United States (Houang, et al 2005).

There have been successes. Utilizing curriculum analysis to generate indicators of learning is still in its infancy. As the Top Trace Map example shows, measures of curriculum that are not embedded in the substance of the subject matter — here mathematics — and that measure more generalized (and thus, in some ways, superficial) aspects of subject matter are not adequate descriptors of curriculum. When combined with a measure that does derive directly from the subject matter such as coherence, however, it does predict achievement. The measure of coherence can only be defined in terms of the substance of the particular subject matter. Its definition would be different for other subjects. Measures of focus, breadth or depth do not have this characteristic — they are essentially the same regardless of subject matter. Types of measures derived *from the content itself* do characterize important cross-national differences that are related to cross-national differences in achievement. Content, at least in mathematics, has yet to become so globalized and homogeneous in substance and sequence that it minimizes the need for identification of cross-national differences and related to achievement differences.

References

Achieve, Inc. (2004) "Do Graduation Tests Measure Up?: A Closer Look at State High School Exit Exams." Washington, DC: Achieve Inc.

Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Smith, T. A. and Kelly, D. L. (1996) *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS)* (Chestnut Hill, MA: Boston College).

Houang, R.T, L.S. Cogan, J. Ferrini-Mundy and W.H. Schmidt. (2005) Triangulating on Curriculum to Motivate and Inform Systemic Curricular Reform. Paper presented at the MSP Evaluation Summit, Minneapolis, MN, September 15-16, 2005.

Miller, Jon. (2004) "Student achievement in context: the influence of parent and home factors on student achievement in mathematics." Paper presented at the AERA Annual Meeting.

McDonnell, L. M. (1995). "Opportunity to Learn as a Research Concept and a Policy Instrument." Educational Evaluation and Policy Analysis **17**(3): 305-22.

Robitaille, D. F., Schmidt, W. H., Raizen, S., McKnight, C., Britton, E., and Nicol, C. (1993). *Curriculum Frameworks for Mathematics and Science*. Vancouver: Pacific Educational Press.

Schiller, K.S., W.H. Schmidt , R.T. Houang , and C. Muller. (2007, April) Hidden Disparities: Underestimating Differences in Learning Opportunities in High School. Paper presented at AERA Annual Meeting.

Schmidt, W. H. and McKnight, C. C. (1995) Surveying educational opportunity in mathematics and science: an international perspective. *Educational Evaluation and Policy Analysis*, 17(3), 337–353.

Schmidt, W. H., McKnight, C. C. and Raizen, S. A. (1997a) *A Splintered Vision: An Investigation of U.S. Science and Mathematics Education* (Dordrecht, The Netherlands: Kluwer).

Schmidt, W. H., McKnight, C. C., Valverde, G. A., Houang, R. T. and Wiley, D. E. (1997b) *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Mathematics* (Dordrecht, The Netherlands: Kluwer).

Schmidt, W. H., Raizen, S. A., Britton, E. D., Bianchi, L. J. and Wolfe, R. G. (1997c) *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Science* (Dordrecht, The Netherlands: Kluwer).

Schmidt, W. H., McKnight, C. C., Cogan, L. S., Jakwerth, P. M. and Houang, R. T. (1999) *Facing the Consequences: Using TIMSS for a Closer Look at US Mathematics and Science Education* (Dordrecht, The Netherlands: Kluwer).

Schmidt, W. H. (June, 1999) *Using Off-the-Shelf Tests in Large-Scale Assessment Systems*, National Conference on Large Scale Assessment, Snowbird, UT.

Schmidt, W. H. (June, 1999) *Benchmarking Standards and Comparing Results: An Update on the Work of Achieve*, National Conference on Large Scale Assessment, Snowbird, UT.

Schmidt, W. H. (June, 1999) *The Alignment of Standards and Assessments*, National Conference on Large Scale Assessment, Snowbird, UT.

Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H. C., Wiley, D. E., Cogan, L. S. and Wolfe, R. G. (2001) *Why Schools Matter: A Cross-National Comparison of Curriculum and Learning* (San Francisco, CA: Jossey-Bass).

Schmidt, W. H., H.A. Wang, and C.C. McKnight, (2005) "Curriculum Coherence: An Examination of US Mathematics and Science Content Standards From An International Perspective," *Journal of Curriculum Studies*, 37(5).

Schmidt, W. H. and R.T. Houang. (2007) Lack of Focus in the Mathematics Curriculum: Symptom or Cause. In Tom Loveless (Ed.), *Lessons Learned: What International Assessments Tell Us about Math Achievement*, pp. 65-84, Brookings.

TIMSS International Curriculum Analysis Database. Retrieved March, 2008, from http://ustimss.msu.edu/

Travers, K. J., & Westbury, I. (1989). *The IEA study of mathematics I: Analysis of mathematics curricula* (Vol. 1). Oxford, England: Pergamon Press.

Valverde, G. A. and Schmidt, W. H. (2000) Greater expectations: learning from other nations in the quest for 'world-class standards' in US school mathematics and science. *Journal of Curriculum Studies*, 32(5), 651–687.

Valverde, G. A., Bianchi, L. J., Wolfe, R. G., Schmidt, W. H. and 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).