Draft Draft Draft

Lack of Focus in the Mathematics Curriculum: A Symptom or a Cause

William H Schmidt
Education Policy Center
Michigan State University

When the original TIMSS study was first released in 1997 preceding it was the report of the curriculum component. That report stressed the importance of curricular differences among countries in interpreting the cross national differences in achievement. One of the original findings achieved significant importance for possibly understanding the relatively poor performance of US students. That finding was characterized by the description of the US mathematics curriculum as one that was "a mile wide and an inch deep." The data indicated a tendency for the US curriculum to cover more topics at each grade level than all of the other countries in TIMSS. This was true of state standards, the NCTM standards, district standards, textbooks and even in terms of the curriculum as implemented by the teachers. The fact was that there was little focus in the US curriculum given the large number of topics that needed to be and actually were covered. In this paper we return to the issue of focus and ask whether such a concept is useful not just as an hypothesis about the US but more generally in understanding cross-national differences in achievement.

SYMPTOM OR CAUSE

Subsequent to those original observations much more insight has been gained into the cross national differences in curriculum and their relationship to achievement (Schmidt, *et al* 2001). One such insight has been labeled coherence and applies to all aspects of the curriculum but in this paper we focus on curriculum or content standards. It is defined as follows:

> Content standards, in the aggregate, 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 the existence of a single coherent sequence, only that such a sequence reflect 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*, 2001).

If content standards reflect the structure of a discipline, then those standards should *increase* in terms of depth and rigor as students move across the grades. Coherence implies replacing repetition of the same standards across grade levels that is found in the US with a set of standards that form a trajectory over those 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 (Schmidt, *et al*. 1997, 1999, 2001).

The emergence of this concept could lead to the conclusion that the concept of focus is merely a symptom or indicator that coherence is lacking in the curriculum. In this interpretation the lack of coherence is the real disease and the lack of focus or the large number of topics in the curriculum is more like a high temperature which only serves as an indicator that something is wrong. To explore this issue we reanalyzed the TIMSS data using indicators of both concepts.

THE DATA

The 1995 TIMSS was the most extensive cross-national comparative study of education ever attempted (Schmidt and McKnight 1995, Beaton *et al*. 1996, Schmidt *et al*. 2001). Discussion of the achievement results have prompted US policy-makers, as well as those from other countries, to consider more carefully the curriculum portraits TIMSS produced, especially those for the highest-achieving countries (Schmidt et al. 1997).

The content standards in mathematics of the TIMSS top-achieving countries were used to develop a model of coherence. A methodology was developed to determine the elements that were common to these countries. We then used these common elements to define an 'international' set of standards. The resulting international standards were then analyzed by research mathematicians in terms of their mathematical coherence. (The description in this section is a paraphrase of Schmidt *et al*, 2001).

Valverde and Schmidt (2000) 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). Six such countries were identified- Singapore, Korea, Japan, Hong Kong, Belgium (Flemish), and the Czech Republic.

The data used to develop the international benchmarks come from the curriculum component of TIMSS and derive from the procedure known as General Topic Trace Mapping (GTTM) (Schmidt et al. 1997). The respondents to the GTTM were 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 the mathematics topic coverage from grades 1-8 is included.

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 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.

The state standards we sampled were subjected to a different content analysis procedure than that used for the development of the international benchmarks. The TIMSS curriculum analysis developed a systematic document analysis methodology. We used this approach to generate the US data using 21 state standards in effect during the 1999-2000 school year.

THE CURRICULUM STRUCTURE OF THE TOP-ACHIEVING COUNTRIES

Figure 1 portrays the set of topics for grades 1-8 that represents only the common topics intended by at least two-thirds 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, 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 and decimals), but with an introduction to the topics of percentages, negative

numbers, integers, proportionality, co-ordinate geometry, and geometric transformations. (Schmidt *et al*, 2001)

FIGURE 1

Mathematics Topics Intended at Each Grade by Top-Achieving Countries

	Grade							
Topic	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Whole Number: Meaning	●	●	●	●	●			
Whole Number: Operations	●	●	●	●	●			
Measurement Units	●	●	●	●	●	●	●	
Common Fractions			●	●	●	●		
Equations & Formulas			●	●	●	●	●	●
Data Representation & Analysis			●	●	●	●		●
2-D Geometry: Basics			●	●	●	●	●	●
2-D Geometry: Polygons & Circles				●	●	●	●	●
Measurement: Perimeter, Area & Volume				●	●	●	●	●
Rounding & Significant Figures				●	●			
Estimating Computations				●	●	●		
Whole Numbers: Properties of Operations				●	●			
Estimating Quantity & Size				●	●			
Decimal Fractions				●	●	●		
Relation of Common & Decimal Fractions				●	●	●		
Properties of Common & Decimal Fractions					●	●		
Percentages					●	●		
Proportionality Concepts					●	●	●	●
Proportionality Problems					●	●	●	●
2-D Geometry: Coordinate Geometry					●	●	●	●
Geometry: Transformations						●	●	●
Negative Numbers, Integers, & Their Properties						●	●	
Number Theory							●	●
Exponents, Roots & Radicals							●	●
Exponents & Orders of Magnitude							●	●
Measurement: Estimation & Errors							●	
Constructions Using Straightedge & Compass							●	●
3-D Geometry							●	●
Geometry: Congruence & Similarity								●
Rational Numbers & Their Properties								●
Patterns, Relations & Functions								●
Proportionality: Slope & Trigonometry								●

Intended by more than half of the top-achieving countries ●

The ‘upper triangle’ appearance of the display in figure 1 implies a hierarchical sequencing of topics in the top-achieving countries over the first eight grades. As we have indicated, this sequencing moves from elementary to more advanced topics. Not only is the progression of the topics over grades logically consistent with the nature of the mathematics but that the progression culminates at 7th and 8th grade in more rigorous topics than those usually intended in the US, at least for the majority of students. (Schmidt *et al*., 2001).

METHOD

To explore the issue of whether focus is itself an important and separate indicator of the curriculum or if it is essentially redundant with coherence (merely a reflection of the lack of coherence) we developed statistical indicators of both concepts. For each country their GTTM data were mapped into the matrix defined by Figure 1. The rows and columns of Figure 1 were considered as fixed and remained constant. The resulting map for each country placed an indicator (represented graphically by a dot as found in Figure 1) into various cells defined by the 32x8 matrix (256 possible cells or opportunities to cover math topics). Those “dots” indicated that the topic defined by the row was covered in the content standards of that country in the grade defined by the column. In effect this map indicated what topics were covered at which grades for each country.

The region of the matrix with “dots” in Figure 1 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 coherence. One way to think of this process is the model region was highlighted creating a “silhouette” as shown in Figure 1 which was then superimposed on the country maps and 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 discussed previously we are not arguing for only one model of coherence, so 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 1. It should be remembered that Figure 1 is a composite of the A+ countries and as such does not completely represent the curriculum of any country even the A+ countries. In this way the "silhouette" serves the role that a mean plays in other statistical analyses.

To define coherence at specific grade levels the 'upper triangle' of Figure 1 must be correspondingly partioned. The eighth grade 'upper triangle' includes 99 cells in the matrix. 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 second set of statistics which we used as indicators of focus measured the extent to which a country map includes "dots" outside of the silhouetted area. Large values here indicated a lack of focus typically the result of including a large number of topics outside those needed to reproduce the coherent pattern.

For any country the 256 cells (32x8) 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 and, as described above, are used in the definition 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 covered or introduced after coverage in the A+ scenario is finished

The three groups can be used to define focus. One straightforward indicator is the sum of the three values which defines the total number of topics 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 portioned accordingly for each separate grade.

The second statistic is based on number 2 above. The number of cells before the ‘upper triangle’ with a positive indicator for intended coverage reflects a lack of focus by adding more topics to earlier grades. It is a particular type of lack of focus, one which implies an early treatment of topics where the necessary prerequisite mathematics would likely be being simultaneously covered or worse yet not covered until a later grade. Examples of both kinds are found in the data.

The third estimator we developed is one which combines the concepts of coherence and focus. We took the ratio of numbers 1 and 2 as defined previously. The number of cells with a match to the ‘triangle’ area was divided by the total number of topics intended up to the particular grade in question. For a country this indicates the proportion of intended opportunities to cover topics that focuses on the ideal scenario. Put another way it is a measure of the relative coverage associated with the ‘upper triangle’. Since most of the available cells outside the silhouetted area are below the upper triangle’s “quasi diagonal” we limit this statistic to only those cells.

This indicator of focus is such that high values imply a lack of focus by calling for the coverage of content at lower grades than called for by the ideal scenario. This definition of focus is tied conceptually to that of coherence. The inclusion of topics outside the silhouetted area adds to the number of topics to be covered at the grade assuming the A+ definition of coherence. In essence we define the lack of focus as the coverage of topics outside of those needed to progress over the grades in a coherent fashion.

The outcome measure was the TIMSS eighth grade scaled total test score for the country. It was available at third, fourth, seventh and eighth grade.

RESULTS

To examine the curriculum effects as defined by these indicators of coherence and focus, country level regression analyses were performed. Complete data were available on 19-32 countries depending on the grade level.

Consider first 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 1.

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 19 and 20 countries. The R^2 values indicated around 20 to 25 percent of the variance being accounted for by both indicators. In all four instances the focus measure was statistically significant, indicating that the larger the number of topics intended for coverage the lower the mean country achievement.

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 TIMMS test. The lack of significance at grade three is probably a combination of two factors one being 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 around that scenario.

The next set of analyses used the same measure of coherence but included the special focus measure (See Table 2). Here the significance level of the model fitting varied similarly to the other model fitting ($p<.12$ to $p<.03$). The same was true of the similarities in the estimated R^2. The lack of focus resulting from the premature (at least from the point of view of the model scenario) coverage of topics at earlier grades was statistically significant at each grade level. The estimated coefficients were all negative as well.

Conditioning on this measure of focus, coherence was not significant at either third or fourth grade. But for seventh and eighth grade the coherence measure was significant with estimated positive coefficients.

The final set of analyses included the measure that combined total coverage (focus) with degree of overlap with the ideal scenario. The model varied from marginally significant at grade three ($p<.10$) to statistically significant at the other three grades. In all cases the estimated regression coefficients were positive indicating that if a country spent more of its total effort (defined as the combination of all cells for which coverage was intended) on the topics related to the model scenario the greater the mean performance of the students at each of grades four, seven and eight.

DISCUSSION

We have contended elsewhere that coherence is one of the most critical, if not the single most important, defining element of high-quality content standards. Content standards were defined in terms of the disciplines that underlie the school subject matter articulated in those standards. We suggest that standards are coherent if they specify topics, including the depth at which the topic is to be studied as well as the sequencing of the topics, both within each grade and across the grades, in a way that is consistent with the structure of the underlying discipline (Schmidt, *et al*, 2001).

The results of the above analyses are consistent that with contention 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 inconsistency of the statistical significance ($p<.05$) of the relative measure (coherence over total effort)with the marginal significance of the other analyses ($p<.07$) warrants further investigation especially with a larger number of countries.

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 coherence. The amount of 'clutter' to that vision created by covering topics too early or before their time from a mathematics 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 as reported above when coherence is included in the model. This implies that focus only has a statistically significant negative impact conditional on coherence.

The policy implications for the US are profound. When we examined in other work (Schmidt, *et al*, 2001 and PROM/SE 2005) the mathematics national professional standards developed by the NCTM, the contrast with the ideal scenario was striking. This contrast is equally striking when we examined the standards of more than 100 local districts or the individual standards in over 40 states. Not only is the organizing principle underlying these curricula unlike that of the ideal scenario, but the organizing principle for these US standards seems to be qualitatively different rather than simply differing in degree.

Given the results of these analyses this bodes particularly poorly for the US predicting poor levels of performance especially in the middle grades coming from the lack of both coherence and focus. This is what one can predict and in fact what we have seen again and again as recently as 2003.

Table 1

Model	*Grade 3*	*Grade 4*	*Grade 7*	*Grade 8*
Predictor	**Estimate (StdError, p)**	**Estimate (StdError, p)**	**Estimate (StdError, p)**	**Estimate (StdError, p)**
Number of Topics 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 Topics 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**	**0.2490**
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
Standardized Coefficient				
Number of Topics Aligned with Ideal Scenario	**0.4374**	**0.7331**	**0.8481**	**0.8240**
Total Number of Topics in Curriculum	**-0.6969**	**-0.9703**	**-1.0044**	**-1.0204**

Table 2

Model	*Grade 3*	*Grade 4*	*Grade 7*	*Grade 8*
Predictor	**Estimate (StdError, p)**	**Estimate (StdError, p)**	**Estimate (StdError, p)**	**Estimate (StdError, p)**
Number of Topics Aligned with Ideal Scenario	**7.52** (6.57 , *0.268*)	**4.91** (3.19 , *0.141*)	**1.84** (1.06 , *0.094*)	**1.66** (0.87 , *0.068*)
Number of Topics Not Aligned (Introduced Prior to Ideal Scenario)	**-2.42** (1.13 , *0.047*)	**-2.58** (1.05 , *0.025*)	**-1.79** (0.71 , *0.017*)	**-1.97** (0.70 , *0.009*)
Model Fit				
R-Square	**0.2191**	**0.2574**	**0.1826**	**0.2122**
Residual Mean Squares	**1652.9**	**1489.0**	**1547.9**	**1396.6**
p<	*0.1222*	*0.0687*	*0.0538*	*0.0279*
Number of Countries	20	21	32	33
Standardized Coefficient				
Number of Topics Aligned with Ideal Scenario	**0.3309**	**0.4807**	**0.4382**	**0.4825**
Number of Topics Not Aligned (Introduced Prior to Ideal Scenario)	**-0.6209**	**-0.7646**	**-0.6405**	**-0.7151**