

CURRICULUM AND LEARNING GAINS IN MATHEMATICS: A CROSS-COUNTRY ANALYSIS USING TIMSS

Richard T. Houang, Michigan State University, USA
William H. Schmidt, Michigan State University, USA
Leland Cogan, Michigan State University, USA

Abstract

The unique contribution of international studies allows an examination of educational structure, curriculum, and policy frameworks from various countries and how these are related to cross national achievement differences. Measures of both the intended curriculum (derived from country standards and textbooks) and the implemented curriculum (derived from teacher questionnaires) were collected in a way to provide OTL measures in the same topics areas as represented by the 20 sub-test areas. Only three of the 20 tested areas – ‘relations of fractions’, ‘two dimensional geometry’, and ‘probability’ – demonstrated no statistically significant relationship to learning. Gain in the other 17 topic areas were significantly related to measures of the various curricular topics. Curriculum was related to learning in mathematics across countries in 17 of the 20 tested topic areas as measured by the TIMSS Population 2 test. Further, the relationships with gain involved different aspects of the curriculum for different topics but across the topics, all aspects of curriculum other than the content standards measure were represented. All of this has powerful implications for mathematics education and in particular for curriculum policy which is further discussed.

INTRODUCTION

How are differences in learning gains and curriculum coverage in various mathematics topics related across countries? The structure of curriculum as it is shaped in a particular country is a fundamental policy question. The structure of curriculum – at least in school mathematics– is related to achievement gains within a country across topics (Schmidt, et. al., 2001). These are country-level descriptions. However they focus on resource allocations and priorities across topics within a country. This is necessarily a matter of relative gains. Scarce resources of instructional time, teacher attention, and textbook space are allocated among a range of topics in school mathematics.

However, it is also interesting to examine the absolute amount of learning that takes place in eighth grade in specific topic areas (not relative to other topic areas). That is, we are interested not just in the differences among countries but in the difference of their effectiveness in using curricular opportunities to bring about learning in a specific topic area, regardless of how they structured those opportunities relative to others.

For that purpose, this paper focuses on cross-country differences in the allocation of curriculum resources and in their relationships to achievement gains. We are looking for the most effective allocations of resources for different mathematics topics. Most simply, we are interested in why some countries did better than others on the Third International Mathematics and Science (TIMSS) Population 2 test (eighth grade in most countries) and, more generally, why that would be true for other TIMSS-like tests. Why do children in some countries learn more in eighth grade than they do in other countries?

COUNTRY LEVEL RELATIONSHIPS AVERAGED ACROSS TOPICS

Before turning to topic-by-topic analyses, we explore several country effects averaged over topics related to instructional time. By averaging over topics the relationship that emerges reflects a general country level relationship that holds on average across topics.

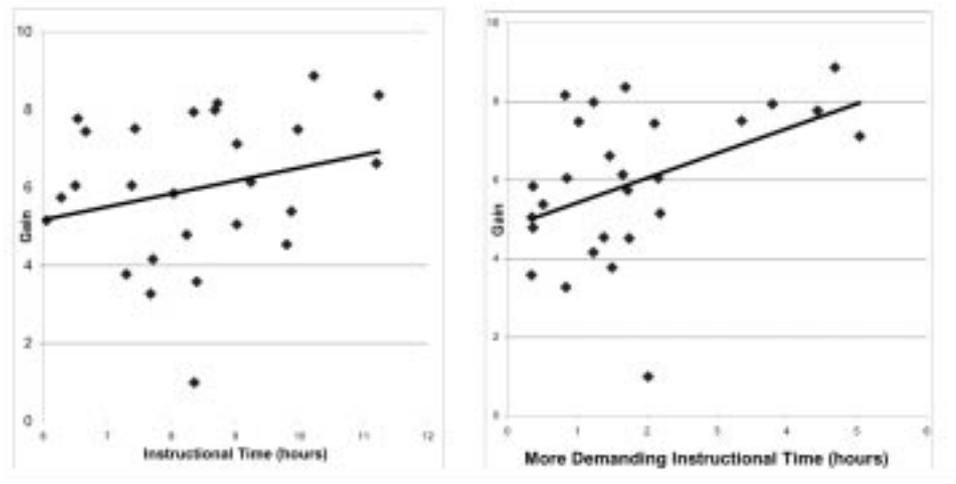
Because of the way in which instructional time was defined – the percent of the instructional year allocated to a topic estimated as the average for a typical teacher at the national level – the country averages over topics were constrained, conceptually, to be equal.¹ To examine the relationship of instructional time to learning on average, we transformed this variable to reflect the total amount of time (in hours) allocated to a topic within a country.² This was to capitalize on what variability there was in actual time amounts across countries, something that could not be done when examining relative time allocations.

Using this definition of instructional time, the two-way analyses indicate the presence of significant country effects in mathematics ($p < .015$). These analyses held constant textbook space and content standards. For mathematics the R^2 index was .24.

Exhibit 1 shows the plot of achievement gain and allocated time in hours for an average mathematics topic. (In this plot the data points represent countries and the other aspects of curriculum were not controlled.) In general there was a positive relationship over the 27 countries whose data are displayed. The straight line represents their linear relationship (the regression line). However, the R^2 associated with the regression analysis was relatively small.

We next examined a more complex model which is represented by the curved line in Exhibit 1.³ This resulted in an R^2 of .12 ($p < .07$). This implies that the pair-wise relationship was such that the rate of increase in gain for more time was not constant (directly proportional to time). Rather, it in fact accelerated rapidly after about nine hours per topic. Before that the amount of time per topic did not seem to appreciably have a differential effect on gain. These simple plots do not hold constant the other aspects of curriculum as was the case with the two-way analyses.

Exhibit 1: Scatterplot of Achievement Gains and Instructional Time (in Hours) and Instructional Time Related to More Demanding Performance Expectations (Eighth Grade Mathematics)



Instructional Time for More Demanding Expectations

We are pressing analysis close to the limits of our data here but we felt one more effort to further tease out the relationship if it exists was in order. Another measure of instructional time was constructed. Total instructional time in hours was the first measure for time used above. We would like to be able to separate out instructional time devoted to instruction expecting more demanding performances from students. We did this by using textbook space devoted to more demanding student performance expectations as an indicator of how much instructional time was devoted to these more demanding expectations. This involved making an assumption that the proportion of textbook space devoted to more demanding student performance expectations was a reasonably good substitute for a direct measure of the proportion of instructional time devoted to so such more demanding expectations. The data did not contain a direct measure of this more demanding instructional time.

Exhibit 1 also shows a plot of this new time variable with achievement gains. The linear relationship (represented by the line) of more demanding instructional time to achievement gain unadjusted for total textbook space given to a topic and for content standards was stronger than the general relationship of time to achievement gain. The R^2 was .20 ($p < .05$). The two-way analysis adjusted for the other two aspects of curriculum indicated a statistically significant relationship of the new time variable to achievement gain ($p < .016$).

The two plots shown in Exhibit 1 under the umbrella of the two-way analyses suggests the important result that instructional time was related to achievement gain where both measures were averaged over topics. That is, the more time a typical

topic received from teachers on average for a country, the larger the average gain was for a typical topic in that country. That is, those countries that allocated comparatively larger amounts of instructional time to a typical topic also had associated with them correspondingly larger achievement gains.

The positive relationship for instructional time spent on more demanding expectations to achievement gains leads to an important conjecture. It suggests strongly that more instructional time centered on topic-specific problem solving and mathematical reasoning was associated with greater learning at the country level. Since our measure of time was a hybrid involving both instructional time and textbook coverage we can offer this idea only as a conjecture. Certainly this seems to be an area for better curriculum measurement and investigation in future studies. This is consistent with a similar finding from a study done in the United States – which focused at the classroom level (Gamoran, Porter, Smithson, & White, 1997).

We must be very careful here. When we speak of more time spent on problem solving and mathematical reasoning averaged across topics this is not the same as generic, "content free" instruction on problem solving. The measure we used was totally specific to the coverage of specific topics even if it was averaged over several topics. It certainly was not instruction on problem solving in the absence of a specific topic. At least in the United States there are advocates of problem solving instruction that is essentially "content free" focusing on the problem solving process as its own content and with mathematical content merely incidental and subservient. This is not that nor does it come close to offering "hard" evidence about the value of content-free problem solving instruction.

Issues of Focus and Emphasis

The above analyses show, in effect, that the more time allocated to mathematics instruction on a typical topic in a country the greater is the associated overall or average gain. This speaks in a sense to the issue of emphasis or focus in instruction. Since these relationships were defined at the average topic level, the relationship can be simplified to that of the amount of total instructional time over the year. When considering an average over topics, allocating more time nationally is equivalent to providing greater focus or emphasis to mathematics as a whole rather than to some topics and not others. A subtler and more important question is whether this degree of variation in emphasis through instructional time was related to achievement gain on a topic by topic basis. We will turn to this question in a subsequent section.

We have written on the closely related issue of focus elsewhere (Schmidt, McKnight, Raizen, 1997; Schmidt, McKnight, Cogan, Jakwerth, & Houang, 1999). We have referred to the lack of focus or emphasis in US curricula in school mathematics. We referred to such curricula as being "a mile wide and an inch deep." This referred to the US practice of teaching many topics but with little depth or emphasis in the instruction on any of those topics.

The above analyses do not directly address this use of the term "focus" since they deal with the total amount of instructional time for mathematics. This is because in those analyses we always divided by the total number of topics when taking an

average over topics. Thus the only way in that case to obtain greater focus or emphasis for a topic would be to increase the total amount of instructional time available for mathematics over the school year. Of course, this has policy implications and might be used in arguments for increasing emphasis by increasing instructional time overall. However, this can be done only up to a point. Total available school instructional time is a "scarce" or at least limited resource and one that must be shared with instruction for other disciplines.

Another way to achieve greater focus – and a way we have argued for elsewhere – is to teach fewer topics in a given school year (Schmidt, McKnight, Cogan, Jakwerth, & Houang, 1999). Earlier analysis elsewhere suggested that the United States was an outlier in terms of the number of topics it included in eighth grade mathematics compared to other countries. For example, US eighth grade textbooks devote some space to more topics than in any other country except one. This placed them in the 97th percentile at 1.64 standard deviations above the mean. The only country with more topics was Switzerland where the data was a composite across three distinct language-based educational sub-systems. The inclusion of more topics in such a composite does not have the same meaning as that in the United States. This is because no Swiss textbook actually included all the topics represented by the "average" textbook, while in the US what was typical was also actual.⁶

We believe that this effect of a lack of focus is a straightforward inference based on the US position as an outlier in the data and on its relatively poor performance. However, it is an inference that is specific to the United States. The question here is whether a similar inference holds in a more general way. In particular – and in contrast to the foregoing analyses – the question is whether it held without regard to the total amount of instructional time (at least within the limits found in the TIMSS data).

To answer the question of focus and use the findings above we returned to the textbook-based measure of the percent of a topic for which instruction focused on more demanding performance expectations. It was the use of this measure that seemed to convert total teacher time to a more sensitive measure of time – at least more sensitive with respect to learning. Using this variable as we did above works with total instructional time but moves towards reflecting a qualitative difference in coverage (inferred from textbooks but presumed to carry over into classroom instruction).

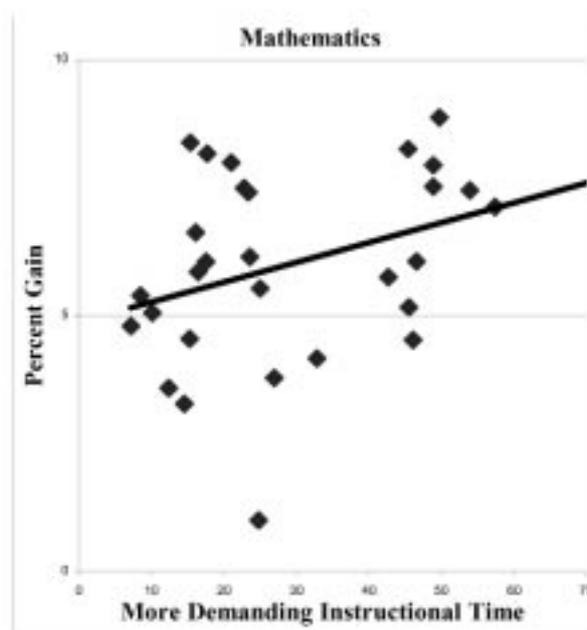
To better make this measure reflect focus, we took the same average but only for those topics for which there was some textbook coverage with more demanding textbook expectations. The average then could be interpreted in terms of the typical topic that received some coverage with demanding performance expectations in a country's textbook. This seems to be a more accurate measure of focus for the textbook. Given our assumption that textbook space is a good rough indicator of emphasis on more demanding performances for classroom instruction, it also becomes a more accurate measure of focus associated with instructional time.

It is as we stated earlier a measure of the qualitative nature of focus and not just its quantitative aspect. A large value for this measure averaged over topics would arise

because only the topics receiving coverage that involved more demanding expectations would be counted in that average. We believe that this is a reasonable definition of focus because we take "focus" to mean not just more coverage but qualitatively different coverage. In this measure increased emphasis would come from including more demanding expectations rather than simply more coverage. This, we feel, is why this is a better indicator of focus (albeit rough) rather than simply emphasis in the sense of increased coverage. Working with more demanding expectations likely takes more instructional time. Since instructional time is a limited resource, increased work with more demanding expectations seems likely to be linked to work with somewhat fewer topics - another sense in which we have used focus elsewhere. This does not mean that the other topics would not be covered but that that they would at best be covered in a less demanding way, perhaps as review or in a cursory introduction.

Now the question becomes, "Is the amount of focus on a topic in the sense of more demanding expectations related to average achievement gain?" Exhibit 2 displays a plot in which the relationship was statistically significant ($p < .044$). The relationship was essentially linear (adding a non-linear component did not appreciably help) and had an R^2 of .14.

Exhibit 2: Scatterplot of Achievement Gains and Instructional Time for more Demanding Performance Expectations Averaged only over Topics Receiving some more Demanding Coverage



This provides some evidence that higher percentages of coverage of a typical topic that involved more demanding performance expectations were associated with larger than average achievement gains. We take this as likely to imply that coverage focused on a limited number of topics that were covered more deeply (in the sense of exploiting more demanding expectations) was related, albeit modestly, to higher average achievement gain. The number of topics covered with more demanding performance expectations was smaller than the number covered in general at eighth grade for all countries. That much of our assumption seems grounded empirically. This is consistent with our earlier discussion of emphasis on topics by increased instructional time but now is extended to the sense of focus as covering fewer topics but with more demanding performance expectations.

TOPIC-SPECIFIC RELATIONSHIPS

We now turn to analyses done for each topic separately. Exhibit 3 displays the pair-wise relationships between each measure of curriculum and between achievement gains in eighth grade mathematics. Eight of the 20 mathematics tested area topics did not show any statistically or marginally significant relationships between any of the aspects of curriculum and achievement gain. These topics included relations of fractions, estimating quantity and size, measurement units, two dimensional geometry, proportionality (two topics), functions, and probability.

For all other topics at least one measure of curriculum was significantly related to learning. Teacher implementation was the aspect of curriculum producing the largest number of statistically significant or marginally significant pair-wise relationships with achievement gain. For the two measures of teacher implementation – instructional time and proportion of teachers covering a topic – there were relationships with achievement gain in six of the tested topic areas.

‘Equations and formulas’ was the topic covered most across all countries and in which the largest achievement gain took place. It had only one significant pair-wise relationship. This was with the measure of the proportion of textbook space covering that topic with more demanding performance associations. The same is true at the other end of the spectrum for three of the arithmetic test area topics – whole numbers, fractions & decimals, and percents. Analysis suggests that more opportunity to learn indicated by any aspect of curriculum was not related to increased learning at the eighth grade for any of these arithmetic topics. This is not surprising since these topics were the focus of instruction for most of the countries at the elementary grades and not in lower secondary (middle) school. However, opportunity that was related to more demanding performances was related to learning in a statistically significant way. Put simply, these data support the conclusion that it was not the quantity of the opportunity that was related to increased learning in arithmetic at eighth grade but rather it was the quality of that opportunity.

Exhibit 4 displays the results for the structural relationships. These results are consistent with those of the bivariate analyses. Essentially the same eight topics were found to have statistically significant relationships of curriculum to learning as was

Exhibit 3: Pair-wise Relationships between Aspects of Curriculum and Estimated Achievement Gains in Eighth Grade Mathematics

Topic	Content Standards Coverage & Gain		Textbook Coverage & Gain		Textbook's Complex Expectation Coverage & Gain		Teacher Coverage & Gain		Instructional Time & Gain	
	Coefficient (std. err.)	R ²	Coefficient (std. err.)	R ²	Coefficient (std. err.)	R ²	Coefficient (std. err.)	R ²	Coefficient (std. err.)	R ²
Whole Number	0.00 (0.01)	0.00	0.02 (0.09)	0.00	0.07 (0.02)**	0.28	-0.03 (0.04)	0.02	-0.16 (0.15)	0.04
Common Fractions	-0.01 (0.01)	0.02	0.19 (0.14)	0.06	0.05 (0.02)**	0.16	0.01 (0.02)	0.00	0.08 (0.07)	0.04
Dec Frctns & Percents	0.00 (0.01)	0.00	0.04 (0.07)	0.01	0.05 (0.03)	0.14	0.01 (0.02)	0.01	0.05 (0.05)	0.04
Relations of Fractions	-0.01 (0.01)	0.00	0.26 (0.28)	0.03	0.01 (0.03)	0.00	0.00 (0.02)	0.00	0.07 (0.08)	0.03
Estimating Quantity & Size	0.01 (0.01)	0.02	1.07 (1.59)	0.02	0.01 (0.02)	0.01	0.01 (0.02)	0.01	-0.02 (0.25)	0.00
Rounding	0.00 (0.01)	0.04	0.60 (0.54)	0.04	0.02 (0.04)	0.01	0.03 (0.02)*	0.10	0.59 (0.24)**	0.17
Estimating Computations	0.00 (0.01)	0.00	0.54 (0.30)*	0.10	0.02 (0.02)	0.03	0.02 (0.02)	0.02	0.25 (0.25)	0.03
Measurement Units	0.00 (0.01)	0.01	0.14 (0.09)	0.08	0.02 (0.02)	0.04	0.02 (0.01)	0.07	0.19 (0.16)	0.04
Perimeter, Area & Vol	0.01 (0.02)	0.00	0.10 (0.07)	0.06	0.03 (0.02)	0.09	0.07 (0.03)**	0.18	0.40 (0.15)**	0.19
Measurement Est & Err	0.01 (0.01)	0.01	-2.04 (1.02)*	0.12	0.01 (0.02)	0.03	0.04 (0.02)*	0.11	0.65 (0.54)	0.05
2D Geometry	-0.01 (0.02)	0.04	-0.11 (0.12)	0.03	0.01 (0.03)	0.00	0.06 (0.07)	0.03	0.02 (0.16)	0.00
Polygons & Circles	0.00 (0.03)	0.00	0.14 (0.08)	0.08	0.03 (0.02)*	0.10	0.03 (0.03)	0.03	0.19 (0.13)	0.07
3D Geometry & Trans	0.02 (0.02)	0.00	0.01 (0.04)	0.00	0.03 (0.03)	0.04	0.06 (0.03)**	0.15	-0.01 (0.12)**	0.00
Congruence & Similarity	0.00 (0.02)	0.04	0.23 (0.09)*	0.19	0.03 (0.02)	0.05	0.05 (0.02)**	0.20	0.38 (0.09)	0.38
Proportionality Concepts	-0.01 (0.01)	0.00	-0.01 (0.41)	0.00	0.03 (0.02)	0.06	0.00 (0.01)	0.00	0.10 (0.15)	0.02
Proportionality Problems	0.00 (0.01)	0.06	0.16 (0.12)	0.06	0.01 (0.02)	0.01	0.00 (0.02)	0.00	0.19 (0.16)	0.05
Prrns, Rltns, Functions	0.00 (0.02)	0.00	0.08 (0.06)	0.07	0.01 (0.02)	0.02	0.02 (0.01)	0.08	0.11 (0.08)	0.06
Equations & Form	-0.04 (0.02)	0.09	-0.04 (0.04)	0.04	0.06 (0.02)**	0.15	0.02 (0.05)	0.01	-0.04 (0.08)	0.01
Data Rep & Analysis	-0.01 (0.01)	0.01	0.13 (0.12)	0.04	0.02 (0.02)	0.09	0.03 (0.01)**	0.19	0.36 (0.15)**	0.16
Prob & Uncertainty	0.01 (0.01)	0.02	0.00 (0.19)	0.00	0.00 (0.02)	0.00	0.01 (0.02)	0.01	-0.01 (0.36)	0.00

** significant (p < .5)

* significant (p < .10)

true for the pair-wise analyses. There were four additional topics for which this was true in the pair-wise case because those analyses included the performance expectation and measure teacher coverage that was not included in the multiple regressions.

Mainly either content standards or textbooks have an impact on instructional time and this is in turn related directly to achievement gain. For the other four topic areas with statistically significant relationships either content standards or textbook coverage (or both) were directly related to learning. In all four cases there was a strong relationship between content standards and textbook coverage.

The strength of the relationships both in terms of the pair-wise and the more complex analyses were not very strong, at least as measured by the traditional index of fit – the coefficient of determination (R^2). The average values were all relatively small (.20 or less). For the topics with significant relationships the R^2 index ranged from around .10 to .40. For the structural model analyses the two topics with the largest R^2 values were ‘congruence and similarity’ (.42) and ‘data analysis’ (.35).

Exhibit 4: Estimated Structural Coefficients for the Curriculum and Achievement Model for Each Topic (Eighth Grade Mathematics)

Topics	R^2	Textbook Coverage to Achievement Gain δ	Content Standards Coverage to Achievement Gain ξ	Instructional Time Coverage to Achievement Gain Y_t
Congruence & Similarity	0.42	-0.02 (0.12)	-0.02 (0.01)	0.43 (0.14)**
Data Representation & Analysis	0.35	0.25 (0.12)**	0.00 (0.01)**	0.51 (0.16)**
Measurement Estimations & Error	0.24	-2.54 (1.04)**	0.01 (0.01)*	0.47 (0.50)
Rounding	0.20	0.36 (0.58)	-0.01 (0.01)	0.66 (0.28)**
Perimeter, Area & Volume	0.19	0.01 (0.09)	0.00 (0.01)	0.39 (0.19)**
Equations & Formulas	0.19	-0.07 (0.04)*	-0.05 (0.02)**	-0.01 (0.09)
Estimating Computations	0.17	0.59 (0.32)	-0.01 (0.01)	0.29 (0.28)
Common Fractions	0.16	0.21 (0.19)	-0.02 (0.01)*	0.08 (0.10)
Measurement Units	0.12	0.11 (0.11)	-0.01 (0.01)	0.19 (0.22)
Polygons & Circles	0.12	0.11 (0.09)	0.01 (0.03)	0.14 (0.14)
Proportionality Concepts	0.10	-0.11 (0.45)	-0.02 (0.01)	0.18 (0.17)
Relations of Fractions	0.09	0.21 (0.31)	-0.02 (0.01)	0.09 (0.09)
Proportionality Problems	0.09	0.15 (0.13)	-0.01 (0.01)	0.14 (0.17)
Patterns, Relations, & Functions	0.08	0.05 (0.07)	-0.01 (0.02)	0.07 (0.11)
Whole Numbers	0.07	0.09 (0.10)	0.00 (0.01)	-0.27 (0.19)
Estimating Quantity & Size	0.07	0.75 (1.70)	0.01 (0.01)	-0.23 (0.29)
Decimal Fractions & Percents	0.05	-0.03 (0.10)	-0.01 (0.02)	0.10 (0.09)
3-D Geometry & Transformations	0.05	0.03 (0.06)	0.02 (0.02)	-0.08 (0.17)
2-D Geometry	0.04	-0.13 (0.13)	-0.01 (0.02)	0.07 (0.17)
Uncertainty & Probability	0.04	0.11 (0.34)	0.01 (0.01)	-0.41 (0.69)

** significant ($p < .5$)

*significant ($p < .10$)

It must be kept in mind that these analyses are not related to how curriculum policy played out within a country but rather are related to exploring cross-national differences. The general conclusion is similar to those of the within country, across-topic priorities and their relationship to learning (Schmidt, et.al., 2001). That is, more curricular opportunity to learn a topic was related to larger achievement gains. This occurred primarily through textbook coverage or instructional time although content standards were related to both of these two other aspects of curriculum.

‘Congruence and similarity’ was a tested area of geometry among those with the largest achievement gains at eighth grade. It was also a topic that was first introduced in most countries at eighth grade. Substantively this topic is related to other geometry topics but does not build as incrementally from them or depend on them as much as many other topics do. It is essentially more "insular" than other geometry topics. As a relatively insular topic introduced at eighth grade for most countries it is an area "ripe for learning" and thus it was not surprising to see such large achievement gains associated with it at eighth grade.

As a relatively insular topic it could be mostly influenced by curricular opportunity provided directly to it and not as much as provided indirectly through other areas. Both textbook space and instructional time had statistically significant relationships to achievement gain in this tested area. However, of the two only instructional time was statistically significant when the other was controlled.

Using a 10 percent increase in textbook coverage argument, this criterion would predict a two percentage points (2.3) increase in achievement gain considering only textbook space. How much increase does this represent? The average gain for this topic was 7.5 percentage points during eighth grade. The increase of two percentage points would be an increase of two-thirds of a standard deviation.

Instructional time was significantly related to learning for this topic area both in the pair-wise sense and in terms of the structural relationships. The estimated effect was a predicted increase in gain of 6.5 percentage points. Even when controlling for textbook coverage, a 15 percent increase in instructional time would have been associated with an increase in achievement gains of about two standard deviations.

For data analysis the predicted effect of instructional time (controlling for textbook coverage and content standards) was even larger – a 7.5 percentage points increase in achievement gain. This would represent an increase of over two standard deviations. Unlike ‘congruence and similarity’, ‘data analysis’ was a topic area covered by most countries from the primary grades onward. It was neither new nor insular. However, there were several countries that reported not covering the topic at all – literally there were no teachers that reported teaching this topic. In many other countries the percent of instructional time was trivially small. So across countries a strong relationship existed between opportunity and learning.

The other curricular area worth mentioning is equations. This was the topic both with the greatest opportunity to learn provided at eighth grade and the topic with the largest gains in achievement. From the pair-wise analyses, the main relationship

for this topic involved textbook coverage that involved more demanding performance expectations. This topic was one that was covered to some extent across the entire curriculum from the early grades onward and in many countries it received a large focus at seventh grade. It was also a topic that was related explicitly to other parts of mathematics (especially number theory). All of this may make it more difficult to establish a relationship between curricular indicators at eighth grade and learning.

It seems reasonable to assume that at eighth grade countries whose textbooks (and presumably to some extent their typical classroom instruction) focused on the more demanding aspects of problem solving and mathematical reasoning in dealing with this topic were beginning to deal with it in a way qualitatively different from their treatment of it in previous grades. Thus it also seems reasonable that this was associated with greater achievement gains for this topic.

OTHER MEASURES AND THEIR RELATIONSHIPS TO ACHIEVEMENT GAIN

There is a long-standing tradition in cross-national studies that achievement is explained to a large extent by more general background variables such as those associated with a country's economic wealth rather than by variables related to curriculum and schooling. Taken to its extreme this has often been considered as a reason why it has been difficult to find empirical support for the effect of schooling and of curriculum specifics when more general background considerations have already been taken into account.

A large literature exists showing that achievement status is related to measures of a country's wealth or level of economic development (see Sweetland, 1996 for a review of human capital theory and inquiry). This has in part encouraged agencies such as the World Bank to support educational development projects in economically developing countries. One question that must be answered is whether such measures of wealth or development are related to learning as reflected in measures of achievement gain at the national level rather than merely being related to measures of achievement status. A "yes" answer would not only suggest that wealthier, more economically developed countries would stand higher in comparative achievement but also that "the wealthy would get wealthier." It would certainly say a great deal about the "playing field" in deliberate efforts to help educational development in less wealthy or developed countries.

A second kind of question is whether such measures of wealth or development are so powerfully related to achievement gain that, as they appear to do for achievement status, they account for gains so completely that little room is left for differences made by curricular factors. Since this is a book about the effects of schooling and not about economic development, the main question of interest here is whether the inclusion of such national economic measures in the analyses relating curriculum to learning would change any of the results discussed thus far. When measures of a country's wealth or level of development are statistically controlled for (that is, held constant), would this eliminate the relationships of curriculum to learning that we have seen so far?

In this section we examine this issue by doing the same type of regression analyses to examine the structural relationships of curriculum and their relationship to achievement gain while adjusting for measures such as the wealth of a nation. Several measures of wealth and development were explored. The choice did not seem to matter for our purposes of including such a measure only as a control rather than a substantive focus. Similar results were obtained for various measures explored. The variable we finally used most thoroughly and that will be discussed here is a country's gross national product (GNP) in US dollars.

Somewhat more in line with the kinds of issues explored in this book we also include other variables related to providing opportunity to learn or delivering curricular intentions as actual instruction. One of these is a measure of classroom instruction. The analyses presented thus far have included a measure of the proportion of textbook coverage of a topic related to more demanding performance expectations. No directly parallel questions asked teachers about their classroom instruction.

However, there were some relevant questions that teachers were asked and that can help provide additional measures of the nature of classroom instruction as a purveyor of opportunity to learn. Teachers were asked how often their classroom instruction had students practice basic computational skills. They were also asked how often students were asked to write equations to express a relationship. We believe that we can take teachers' spending a great deal of time on the former as indicative of instruction focused on less demanding performance expectations. We also believe that we can take teachers' spending a great deal of time on the latter as indicative of a focus on more demanding performance expectations. Unfortunately these two questions were not asked in relation to specific topics.

An entirely different sort of variable that might have been included in a model exploring factors other than the curriculum to explain achievement gains would be a measure of teachers' subject matter knowledge. The content of the curriculum is delivered through teachers. Not only is the amount of time relevant to this delivery but also the quality of that delivery. We believe that there are at least two aspects related to the quality of classroom instruction that might affect achievement gains. The first deals with the nature of how demanding teachers are in what they expect of students. The second deals with how knowledgeable teachers were about the subject matter on which they were providing instruction. The first we attempted to get at through the questions discussed above that were asked of teachers about computation, ordering events, and so on. The second was not directly measured in TIMSS. At the time of the design of TIMSS some argued for the inclusion of such a measure but as one can easily imagine this issue was politically sensitive and unlikely to result in inclusion of such a measure in a design subject to the constraints of cross-national consensus.

Since no direct measure of teacher subject matter knowledge was included in the TIMSS design, we attempted a rough surrogate through a measure that reflects teachers' beliefs about the nature of subject matter. This measure does not directly reflect how knowledgeable teachers were about their subject matter. It does,

however, indicate how they conceived of the subject matter and likely reflected qualitative differences in the instruction they enacted.

This approach was described in detail elsewhere (Schmidt, McKnight, Cogan, Jakwerth, & Houang, 1999). Three country-level measures based on teacher beliefs about subject matter are included here. The first is the percent of teachers in a country who focused on the procedures, that is, who believed a more algorithmic approach to their subject matter was appropriate for classroom instruction. The second was the percent of teachers in a country who were focused on process. This is different from focusing on procedures. It is not the procedures of the discipline that are the focus but rather the processes by which students come to understand the discipline that are regarded as important. This is in a sense a valuing of students over subject matter in a teacher's belief about appropriate instruction. The third measure was the percent of teachers who focused on subject matter as a discipline with all of its formal aspects.

The previous analyses relating curriculum to learning examined a series of pair-wise relationships involving five measures – textbook coverage, textbook coverage related to more demanding performance expectations, content standards, instructional time, and the percentage of a nation's teachers covering a topic. One thing that can be generalized from all these analyses is that different measures of curriculum were most strongly related to achievement gain for different topics. When we model the amount of learning (achievement gain) for a topic in the present section we include only one aspect of curriculum in the analysis for each topic – the one curricular variable that the previous analyses showed had the strongest relationship to achievement gain for that topic. This will be different for different topics.

We did analyses for each topic area separately. Each analysis included a measure of opportunity to learn –OTL – (the one previously shown most strongly related to achievement gain for that topic). It also included a measure of GNP, one of the measures of classroom instruction taken as a surrogate for more demanding instruction, and one of the measures characterizing the disciplinary focus of the country's teachers.

The first important generalization from these data is that GNP as a measure of a country's wealth was not strongly related to achievement gain. It was only statistically significantly or marginally significantly related to achievement gain for about half of the topic areas. Those areas included 'whole number', 'fractions', 'relations of fractions', 'rounding', 'two dimensional geometry', 'polygons and circles', 'equations and formulas', 'data analysis', and 'probability'.

More importantly from the point of view of this paper, in no case did controlling for GNP eliminate any significant relationship of curriculum to learning found in Exhibit 3 and Exhibit 4. This was also true when several other measures of a country's wealth or development were used in the analyses. The conclusion is that even when controlling for a country's wealth, curriculum is still significantly related to learning.

The opportunity variables were statistically significant or marginally significant for

11 of the 20 mathematics tested topic areas. This was true when controlling for GNP and the other two measures. This is essentially consistent with the results of Exhibit 3 in which there was a significant bivariate relationship for 12 topics. The only difference was that for the area of 'decimals and fractions' the relationship between opportunity and achievement gain was no longer significant when the analysis controlled for the other variables. For the other topics statistical significance remained even when holding constant those other variables.

Another interesting finding associated with these results centers on the classroom instruction and teacher subject matter orientation measures. The classroom instruction measure for mathematics was significantly related to gain for only five of 20 topic areas when controlling for curriculum through the opportunity measure. These five areas were 'whole numbers', 'estimating quantity and size', 'measurement errors and estimation', 'proportionality concepts', and 'proportionality problems'.

For 'whole numbers' and the two proportionality topics countries in which teachers spent a larger proportion of their classroom instructional time on writing equations were also those with larger achievement gains in the three topic areas. For 'estimating quantity and size' and for the measurement topic the relationship was negative and involved the time spent on computational skills. That is, when a larger proportion of time was spent on emphasizing computational skills for those two topics, there were correspondingly smaller achievement gains. Since it seems likely that there is a trade-off between emphasizing simpler performances such as computation and more demanding performances such as expressing relationships in equations, the results for the five topics were essentially similar. More demanding instruction was associated with greater gains for these few topic areas.

The measure of teacher's subject matter orientation (taken here as a rough surrogate for teacher subject matter knowledge) was significantly or marginally significantly related to achievement gain for about half of the topic areas (11 of 20). This was true for the following seven arithmetic topics – 'whole number', 'fractions', 'relations of fractions', 'estimating quantity and size', 'rounding', 'estimating computations', and 'measurement units'. It was also true for the more advanced topics of 'functions', 'equations and formulas', as well as the data analysis and probability topics.

Unfortunately, however, there was no consistent pattern to the relationship over topics that would allow a nice generalization. This may well be because this variable did not reflect teacher subject matter knowledge as we had hoped. However, there was a relationship to achievement gain of teacher beliefs about subject matter averaged over a country's mathematics teachers. That relationship appears to have been very specific to individual topics.

The R² index for these twenty analyses ranged from .25 to .70, with most of the values around .40 to .50. This implies that for most topic areas around half of the variation in achievement gain across countries was related to measures of a country's wealth, OTL, type of classroom instruction, and teachers' subject matter orientation. All but four of these multiple regressions were statistically significant or marginally significant. These four were 'two dimensional geometry', 'three dimensional geometry and transformations', 'functions', and 'probability'.

Relationships for Mathematics at the Most Detailed Level of Topic Specificity

Results for a series of analyses with topics defined at the most specific level possible are presented here. Achievement gain for each tested topic area was related to opportunity in a specific topic helping to define the tested area but at the most specific level possible. These analyses included topics both directly and not directly included in defining the tested topic area. The determination of which topics to include and which curricular aspects to include was based on the insights gained from the empirical analyses thus far in this paper as well as a conceptual analysis related to the mathematics involved.

Many of the results are consistent with those reported in Exhibit 3 and Exhibit 4. From those analyses there were eight topics for which no significant relationship to opportunity could be found. Of these only three still show no relationship to learning. The topics are 'relations of fractions', 'two dimensional geometry', and 'probability'. Gain in the other five topic areas can now be shown to be related to different aspects of curriculum for certain topics.

Since many of the relationships of curriculum to learning are consistent with the results reported previously we will not discuss each of the topic areas separately. The R^2 index used as an indication of the strength of the relationship varied from around .1 to .4. This indicates that as much as 40 percent of the cross-national variation in learning could be accounted for by taking into account only the measures of curriculum opportunities. The estimated regression coefficients were all positive. This indicates that greater opportunity was associated with greater learning.

The aspects of curriculum most involved in significant relationships were textbook coverage, more demanding performance expectations, teacher coverage, and instructional time. Only the content standards measure was not involved in significant cross-national relationships to achievement gain. It is possible that the direct or indirect relationship of content standards to teacher implementation and their direct relationship to textbook coverage within countries were the primary means by which they had a relationship to learning across countries.

Several interesting relationships emerged. Gain in 'perimeter, area and volume' was related to more than just teacher coverage in that topic area. It was also related to textbook coverage, the performance expectation measure, and teacher coverage of 'two dimensional geometry' and 'three dimensional geometry'. This makes great conceptual sense since perimeter and area are typically defined in terms of two-dimensional geometric figures and volume in terms of three-dimensional geometric solids.

The set of relationships for the two proportionality topic areas were perhaps the most interesting to emerge from these analyses. In neither of these two areas was gain found to be significantly related to curriculum in any of the previous analyses that took into account curriculum topics directly for the tested areas. Here significance was noted primarily for curricular coverage in two other topic areas. One, 'congruence and similarity', was from geometry. The other, 'slope and

trigonometry', was from algebra. Although the category is 'slope and trigonometry', for most countries coverage in this area was probably related mainly to slope given the strong emphasis on algebra at the Population 2 level.

The common thread between 'congruence and similarity' and 'slope and trigonometry' is proportionality. There is a strong conceptual relationship between similarity and proportionality. This is also true for slope and proportionality and even for simple trigonometric ratios and proportionality. The curricular aspect through which a relationship between these two areas and achievement gain in the proportionality topics was significant was that of textbook coverage related to more demanding performance expectations such as problem solving and reasoning.

Perhaps in many countries the problems or applications used in such problem solving or mathematical reasoning involved similarity, slope, or even simple trigonometric ratios. Anecdotally the authors have been told that proportionality is not taught as a separate topic in many countries. Persons from various countries informed us that proportionality appeared in the curriculum only through such relationships as similarity in geometric figures and direct proportionality in functions.

There were, in fact, two other topic areas related to the tested area of 'proportionality problems' in these analyses. They were 'functions' and 'equations and formulas'. The reciprocal relationship was at least marginally significant. That is, the relationship of gain in 'function' was related to the performance expectation aspect of textbook coverage related to 'proportionality problems'.

'Measurement units' and 'estimating quantity and size' were two other topic areas where new significant relationships not seen in previous analyses emerged. The curriculum opportunity measures related to these two topics came from the same higher-order category of the TIMSS mathematics framework as the topics defining the tested areas for these topics. Perhaps the more interesting of the two involves the relationship of textbook coverage of 'exponents and orders of magnitude' to achievement gain in 'estimating quantity and size'. Both 'exponents and orders of magnitude' and 'estimating quantity and size' came from the same higher-order category of the (hierarchical) mathematics framework indicating a presumed conceptual relationship.

Putting these results together with those described earlier in this paper seems to lead to a powerful, straightforward conclusion. Curriculum was related to learning in mathematics across countries in 17 of the 20 tested topic areas as measured by the TIMSS Population 2 test. Further, the relationships with gain involved different aspects of the curriculum for different topics but across the topics, all aspects of curriculum other than the content standards measure were represented.

In areas such as 'congruence and similarity' and 'proportionality problems' almost half of the variance in achievement gain across countries was accounted for by curriculum measures. From a previous section we also know that instructional practices and teacher beliefs added to the strength of some relationships. We also know that the inclusion of wealth or development measures did not dampen the relationship of curriculum coverage to national achievement gains in mathematics.

References

- Gamoran, A., Porter, A. C., Smithson, J., & White, P. A. (1997). Upgrading high school mathematics instruction: improving learning opportunities for low-achieving, low-income youth. *Educational Evaluation and Policy Analysis, 19*(4), 325-338.
- Schmidt, W. H., McKnight, C., Cogan, L. S., Jakwerth, P. M., & Houang, R. T. (1999). *Facing the Consequences: Using TIMSS for a Closer Look at US Mathematics and Science Education*. Dordrecht/Boston/London: Kluwer.
- Schmidt, W. H., McKnight, C., & Raizen, S. (1997). *A Splintered Vision: An Investigation of U.S. Science and Mathematics Education*. Dordrecht/Boston/London: Kluwer.
- Schmidt, W. H., McKnight, C., Valverde, G. A., Houang, R. T., & Wiley, D. E. (1997). *Many Visions, Many Aims, Volume I: A Cross-National Investigation of Curricular Intentions in School Mathematics*. Dordrecht/Boston/London: Kluwer.
- Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H. A., Wiley, D. E., Cogan, L. S., & Wolfe, R. G. (2001). *Why Schools Matter: A Cross-National Comparison of Curriculum and Learning*. San Francisco: Jossey-Bass.
- Sweetland, S. R. (1996). Human capital theory: Foundations of a field of inquiry. *Review of Educational Research, 66*(3), 341-359.

NOTES

1. The actual sum of these percents across topics within a country, however, does not equal 100 percent and, as a result, the averages across can vary. This has to do with the fact that the tested topic areas do not coincide with the teacher questionnaire topic categories. Since the focus of the two-way analyses and those of this paper are on the relationship of curriculum to learning we use the tested topic area designations. For several tested topic areas, the same teacher questionnaire category is appropriate, i.e., the independent variable (curriculum) is the same for more than one dependent variable (gain). In effect, the teacher content categories are at a higher level of aggregation in the curriculum frameworks for some tested topic areas. The result is that in the two-way matrices, since some teacher categories are repeated, the sum will be more than 100 percent. Country differences on this variable averaged over topics would, as a result, be artifactual.
2. Using a set of questions asked in the school questionnaire relative to the allocation of time to an instructional period and the number of such periods in a year, we formed a variable – the total number of hours of mathematics in a school year. This was then averaged over schools (properly weighted) to provide a country-level indication of the total instructional time allocated. This was then multiplied by the average percent of teacher time for each topic to obtain an estimate of the total amount of time in hours allocated to a tested topic.
3. We fitted the simplest polynomial model – the quadratic – so that we could see whether the relationship was merely linear or whether there was evidence of a non-linear

component as well. Other models such as cubic, exponentials, etc., might have been fitted. However, since the degrees of freedom were so few, the latter was enough to determine if a non-linear component to the relationship existed.

4. The total time for a topic was multiplied by the country average for the textbook performance expectation variable. We recognize that this is a textbook-defined variable rather than directly and uncomplicatedly about instructional time. However, we used it at the country level as a surrogate measure for what was likely to happen in the classroom. This we view as a not unreasonable assumption given the relationship of textbook coverage to instructional variables. This assumption can certainly be questioned, however.
5. The measure used to convert the total time was a conditional percent of the two categories of performance expectations given the coverage of a specific topic. This measure was specific for each topic.
6. To obtain representative Swiss "textbooks", we took a composite of the textbooks from each system. These textbooks were quite different. In the US, we also took a composite of widely used textbooks but this was just an average of very similar textbooks.
7. We do this for two reasons. First, we have data for a limited number of countries and hence limited degrees of freedom for the error term in the analyses. Secondly, and perhaps more importantly, we wish to avoid the problem of collinearity. Classic evidence of collinearity was present for all measures that were included in the model here. The largest latent root of the covariance matrix was always very large and when using measures related to the latent vectors the condition disappeared.