Review of Educational Research
Winter 2000, Vol. 70, No. 4, pp. 529–545

Is Student Achievement Immutable? Evidence From International Studies on Schooling and Student Achievement

Larry E. Suter
National Science Foundation

International comparative studies of student achievement that have been conducted since 1965 have influenced education policy and research issues in the United States. This article reviews the contribution of recent international studies to an understanding of the role of content in curriculum and thus the way in which schools have an impact on student learning. Studies conducted of U.S. schools during the 1960s by James Coleman and his associates and schools of other countries by the International Association for the Evaluation of Educational Achievement concluded that student performance was determined more by family background than by school characteristics. However, Coleman himself later recanted that finding in a reanalysis of international studies. Recent international studies of student assessment have demonstrated that between-countries differences in how the science and mathematics curriculum is presented may account for differences in student performance. This finding has had an effect on national policy by supporting efforts to reform the content of subject matter introduced in schools.

This article reviews efforts by researchers of cross-national achievement to explain differences in student performance. International studies have introduced techniques for measuring the effects of differences in school emphasis between and within countries on specific content within the subject areas of mathematics and science. In reviewing published analyses of five major international studies conducted between 1964 and 1995, I seek evidence for answers to the following broad questions: Can student achievement, as measured via national assessments, be altered by administrating changes in school system practices? Or is student achievement too strongly established by parental background to be affected by practices in schools?

This paper was originally presented at the Second Annual Conference on Comparative Studies of Educational Achievement, Notre Dame University, November 1999. Most of the paper was completed while the author was on official leave at the School of Education, Stanford University, under the Intergovernmental Personnel Assignment Act. It has benefited from work experiences and written comments from many persons, including Nathaniel Gage, T. Neville Postlethwaite, and Francisco Ramirez and reviewers of this journal. The opinions, findings, and conclusions or recommendations expressed in this article are those of the author and do not necessarily reflect the views of the National Science Foundation.

Do results from international comparative studies provide evidence that changing curriculum content can increase student achievement?

In this article, I examine evidence from past international studies in light of the 1966 Coleman report that attempted to measure the influence of both home background and schooling practices. Coleman et al. (1966) concluded that home background was a much more powerful influence than school on student reading achievement. That study did not include detailed measures of the content of the subject matter provided within classrooms. Many studies of ability grouping or "tracking" have focused on student assignments to different "ability levels" within reading or mathematics classes (see Braddock & Slavin, 1993; Brewer, Rees, & Argys, 1995; Hoffer, 1992; Oakes, 1992). However, none of these studies have broken down the content of instruction into significant areas of emphasis.

Here I report on new research efforts from the Third International Mathematics and Science Study (TIMSS) to designate more clearly the topic areas taught in mathematics and science in different countries and thus establish the amount of influence that curriculum policies have on teacher and student performance. It is recommended that national indicators of educational performance that more adequately measure changes in the content of the U.S. curriculum be constituted and regularly reported.

Curriculum Research

Student achievement is a product of, among many other factors, the amount of exposure students have had to the content of the assessment. National assessments are based on broad frameworks of content areas that are defined by experts in those fields. In the United States, the National Assessment Governing Board sets these content areas (the board's procedures are described on its Web site, http://www.nagb.org). Thus, measures of whether students are given the "opportunity to learn" the content covered by the test are certainly relevant in interpreting test scores. For the purposes of this article, I use the term *curriculum* to refer to the subject content areas that are covered throughout the elementary and secondary school years. In such studies as TIMSS these have been defined, for purposes of data collection, as 44 areas of mathematics and 80 areas of science (Robitaille et al., 1993).

The role of curriculum as an explanatory factor in student achievement has not been completely ignored by researchers (Wittrock, 1986). However, it has been most evident in international comparative studies. For example, Jere Brophy and Thomas Good (1986), in their review of studies on teaching, identified "opportunity to learn" as consistently showing a link between achievement and teaching practices, and they cited the International Association for the Evaluation of Educational Achievement (IEA) study of science as evidence. They pointed out that content coverage was measured directly by Comber and Keeves (1973), who used these measures in 1970 for the first international study of science by asking teachers to state whether or not they covered the specific content of each item that was included on the tests used for assessment. Gustafsson and Undheim (1996), in reviewing studies of individual differences in cognition, also noted that results of international-level studies might be accounted for by differences in curriculum rather than intellectual differences among students. They proposed that examining such international studies might be one way of putting individual differences in perspective.

John Eggleston, in *The Sociology of the School Curriculum*, defined *curriculum* as including the "aims, content, technology (methodology), timing (order) and evaluation that spring, like curriculum itself, from the normative and power systems of society" (p. 4). He further noted that there has been little attention given to the curriculum by sociologists, who have tended to take it as a "given" feature of the school systems they have studied in their investigations of socialization, differential opportunity, and classroom interaction.

International studies of student achievement carried out between 1957 and 1995 contributed to discussions about the performance of U.S. students in elementary and secondary schools by pointing the way toward understanding the role of curriculum in affecting student achievement (Husen, 1967, 1979, 1987). Measurements of differences in content areas taught in each participating country were included in the first international comparisons to help explain observed variations in performance levels of students in different countries. The originators of the studies believed that each country might emphasize different aspects of each content area and that subjects could not be well represented by a single test of achievement. Thus, they initiated development of a measure of curriculum practices and of students' opportunities to learn specific topics.

As expressed by Travers and Westbury (1990), subject areas were described at three levels: the system level, the classroom level, and the student level. The part of the curriculum that was "intentional," such that it was required by the state or country school administration, was labeled the *intended* curriculum. Since teachers have control over what is actually presented in the classroom, they may choose to implement different aspects; thus, the set of topics implemented in the classroom by teachers was called the *implemented* curriculum. Finally, that part of the curriculum that students actually learn (as measured by proportion of test items answered correctly) was labeled the *achieved* curriculum (see also Garden, 1987a, 1987b).

In a review of the impact of international studies, Kjell Harnqvist (1987) noted that the distinction among the intended, implemented, and achieved curriculums was first made in the Second International Study of Mathematics. Measures of opportunity to learn were originally designed as a check on content validity for individual classrooms; however, these indicators turned out to be an important explanatory variable in regard to between-school differences in achievement in the second IEA science study, and they were found to explain differences between countries in the second IEA mathematics study.

Role of Educational Research

One potentially powerful method by which educational policy may increase student performance in the United States is that of changing the emphasis of teaching priorities of school systems toward specific topic areas more closely aligned with high content standards. For example, the Commission on Excellence (1983) stressed the need to change the "content" of schooling in high schools by raising standards and implementing more rigorous content knowledge requirements for graduation (see http://www.ed.gov/pubs/NatAtRisk/). In more recent years, efforts to more fully define the standards for different subject areas have resulted in formal statements of national standards for mathematics and science as well as other subject areas (see http://www.nas.edu). National programs (e.g., the National Science Foundation efforts in regard to systemic reform) have experimented with implementing these standards in teacher training programs for states and cities and in designing national assessments (see http://www.ehr.nsf.gov/ehr/esr/). Each of these programs and reports has sought policies that would move toward improving student achievement by reforming teachers' expectations of students in specific content areas. Previous international studies had produced evidence of the role of curriculum.

However, educational research has not always provided useful recommendations for policy directions. For example, the results of the IEA six-area study conducted in 1970 seemed to support the Coleman et al. results; that is, schools did not have an effect on student achievement. In analyzing that major international study, C. Arnold Anderson (1972, cited in Walker, 1976), director of the Comparative Education Center at the University of Chicago, suggested that educational research was near its doomsday. He was apparently discouraged by the regression analyses reported in the IEA study showing that family background factors were far more strongly related to student achievement than were school characteristics. According to Anderson:

> The guardians of research and of policy-making on education have experienced a decline in confidence and in prestige during the last two decades. Even those investigations that met high standards of craftsmanship and probity have had the finding distorted into a doomsday mode. Both the utopians (such as the "de-schoolers") and the traditionalists who see no point in any research on education have rivaled each other in proposing sweeping and unwarranted conclusions. (cited in Walker, 1976, p. 259)

Anderson was obviously extremely pessimistic about using the results of educational research in positive ways to identify aspects of schools that might improve their conditions. However, he may have been in error to assume that the IEA study had not produced some useful findings on school effects. As it turned out, there were important aspects of that research that had not yet been widely recognized at the time of Anderson's analysis. I show later that Coleman himself (1975) recognized the importance of specifying the content being measured.

Before turning to a review of the national and international studies and the reanalysis of the IEA work conducted by Coleman after Anderson wrote these discouraging words, a description of the current basis for monitoring the progress of U.S. students is needed. The results of the National Assessment of Educational Progress (NAEP) provide information on changes in student performance in science, mathematics, and reading.

National Assessment of Educational Progress

Regular monitoring of national changes in student achievement has been made possible by NAEP's systematic reporting of results from a national survey of students. NAEP has reported on the results of national tests of students in the content areas of reading, mathematics, and science since it was initiated in 1967 (see http://nces.ed.gov/nationsreportcard/site/whatis.asp). The regular publication of NAEP results in eight subject areas always receives the attention of the national media. Changes in these scores are interpreted as evidence that school systems have or have not improved their delivery of educational subject matter to students. However, the NAEP reports do not directly focus on measures of the specific content areas developed since the *Nation at Risk* report (1983). Instead, NAEP publications usually provide only a single scale score for each of the subject areas.

Some typical NAEP trends in mathematics for three age groups are shown in Figure 1. These line graphs reflect scores on a common test scale that was initiated in 1973 and continued through 1996. The principle followed by the Educational Testing Service (ETS) in terms of NAEP "trends" is that change cannot be measured if the measure is changed. To assess the new forms of curriculum, ETS administers a different combination of mathematics items that do, in fact, reflect changes in the curriculums of schools (for information on all NAEP assessments, see http://nces.ed.gov/nationsreportcard/site/home.asp).

In Figure 1, the scores of all three age groups of students on achievement items have been fitted to a common scale for all of the years shown. Thus, the large differences between age groups and smaller differences across time are represented on the same scale. This reduction in scale size leads viewers to conclude that little change has occurred. To some readers, the graphs might appear "flat."

Yet, some significant changes occurred over the period shown. For example, significant increases in mathematics achievement can be seen for 9-year-olds in the 1980s. Other data (not shown in Figure 1) suggest that African American and Hispanic students at young ages, but not at older ages, have also increased in achievement (National Science Board, 1998). NAEP trends detected some small increases in mathematics achievement for the same age groups between 1973 and 1996 and steady increases since 1982 in science (see National Science Board, 1998). However, such changes are often considered too small to receive the attention of policymakers (National Science Board, 2000).

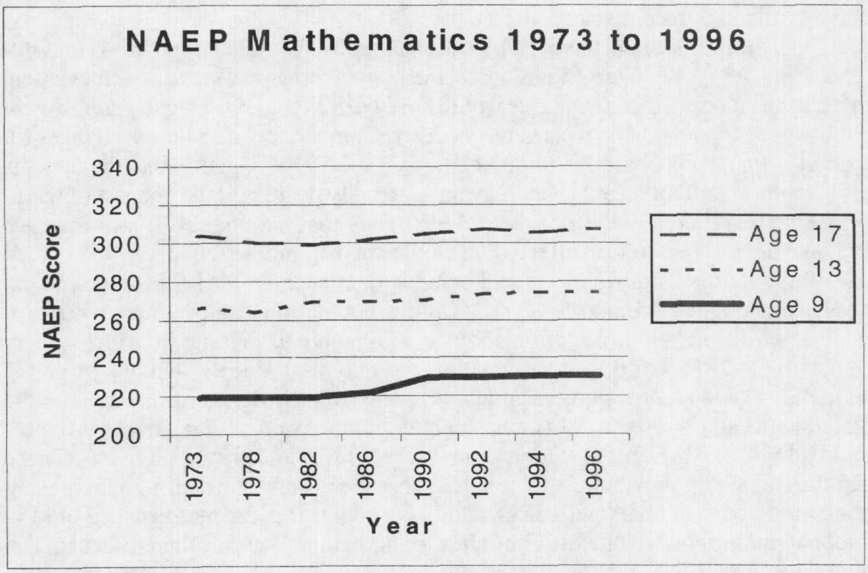

FIGURE 1. *National Assessment of Educational Progress trends in mathematics: 1973–1996.*

Note: Data were derived from Campbell, Reese, O'Sullivan, and Dossey (1996).

The trends illustrated in Figure 1 reflect changes over time for the same age group of students but do not represent changes for the same students. Indicators of change could be created, for example, by following the achievement of 9-year-olds until 4 years later, at the age of 13 years. In an analysis of this cohort, Barton and Coley (1998) showed that there was no increase in growth rates between the 1970s and the 1990s. They compared the cohort growth of 9- to 13-year-olds between 1973 and 1977 with that between 1992 and 1996 and found either a declining growth rate or a steady growth rate in mathematics and science. This steady cohort phenomenon occurs because recent age groups began their schooling at higher levels than did earlier cohorts and then evidenced about the same or slightly less pronounced NAEP score increases in the next 4 years of school. Thus, Barton and Coley's analysis is another example in which significant changes in student achievement levels were not found. Their analysis of cohort trends dismissed any evidence of recent increases in student achievement.

One plausible reason for an increase in the achievement of students is that today's parents have a higher level of education on average than parents of children attending school in the 1970s (Suter, 1995). The higher educational level of the parents of younger cohorts of students has an effect on the opportunities for each new cohort of students and explains about half of the growth in student achievement observed in the overall NAEP trends. Barton and Coley (1998) found that there are small differences in cohort growth rates of achievement in mathematics between students who report that their parents are college graduates and students who report lower education levels for their parents. Barton and Coley reported that cohort growth rates in science were slightly lower for every educational level of parents in the 1990s than they were in the 1970s, suggesting that all influences of achievement have declined in recent years.

Thus, the impression conveyed by national trends in NAEP test scores is that student achievement is nearly immutable. Although some increases in achievement have been detected, these increases would have been expected owing to parents' generally higher educational levels in recent years. Comparison of test score changes for cohorts of students in the 1970s and cohorts in the 1990s suggests that the benefits of being in school in recent years may have actually decreased the amount of learning. One could conclude from the NAEP trends that national efforts to improve achievement have not been effective in elementary or middle schools.

While national studies of student achievement such as NAEP provide indicators of student achievement levels for various population groups, they do not provide equally detailed information about school practices that might affect student achievement. Most of the survey results continue to produce statistically reliable evidence that student achievement levels change very little over time and that the greater a family's resources (as measured by income, educational level, or household goods), the higher the student achievement level in nearly all subjects. These results can lead policymakers to conclude that the primary difficulty in increasing the productivity of the American education system is that the performance of low-income and minority students continues to lag behind that of other students. The message conveyed is that the education system successfully serves students from families with sufficient resources but not others and that efforts to change the system itself are fruitless since no evidence exists to suggest that changes in schools would result in student improvement.

The Coleman Report

The "Coleman report," *Equality of Educational Opportunity*, represents one of the most detailed studies of the determinants of student achievement. Coleman and his colleagues examined the effects of school characteristics and family background characteristics on achievement among students in Grades 3, 6, 9, and 12. The study used sophisticated models to examine the effects of various factors on a test of verbal achievement. Coleman and his colleagues (1966) concluded that "schools bring little influence to bear on a child's achievement that is independent of his background and general social context" (p. 325). Thus, family background is of great importance, and schools represented only a small independent effect due, they reported, to variations in facilities, curriculum, and staff achievement. They also reported that the social composition of the student body (number of minority students) was more highly related to student achievement (independently of the student's own social background) than any school factor. It was this last observation that led to new policies regarding efforts to reduce segregation of students by race.

This study, more than any other, led to a belief among policymakers in Washington that educational research was unlikely to uncover tools that might direct national policy and that student achievement might be immutable. Subsequent policies designed to improve student achievement tended to concentrate on efforts to revise the composition of student bodies.

The IEA Six-Area Study

Other survey studies of the effects of family background and school characteristics on student achievement were being conducted at the same time as the Coleman study, including an international study of mathematics conducted in 1965 in 14 countries and a study of six areas (science, literature, reading comprehension, English as a foreign language, French as a foreign language, and civic education) in 21 countries between 1966 and 1973. British researcher Gilbert Peaker analyzed the data collected from the participating countries to answer the same questions about the effects of schools as those addressed in the Coleman study. The results provided another opportunity to examine the effects of school background on achievement. Peaker (1975) reported the following observations:

1. What is taught is very much the same from country to country. There are no striking differences in curriculums, despite differences in language and history.
2. There is not much variation among the countries with long-established universal education systems in regard to either average achievement or range of achievement. In countries where the educational system is relatively new, the standard of achievement is much lower.
3. The more teaching, the more learning.
4. Efforts made to subdivide teaching into different aspects or elements were restricted by the amount of intercorrelation among factors. A great many variables were brought into the analysis, but most of them appeared to have little separable effect, and only the quantitative measure stood out in all countries.

Peaker, however, addressed the problems of reform by noting that the idea of establishing a common curriculum for all students had detractors. He stated that "the continued existence of the backward child has been one of the causes of the present climate of educational opinion. . . . Every reform has disappointed the hopes of some of its advocates" (Peaker, 1975, pp. 173–174). He further noted that some reform advocates in England had suggested that their current curriculum was not well suited for all students and that a curriculum for the "working class" should be instituted. According to an even more extreme view he noted, the attempt to make all children literate and numerate is misguided and should be abandoned, with attention concentrated on making children happy at school.

Thus, the IEA study researchers seemed to have been extremely gloomy about the prospect of finding school- or policy-relevant indicators that might be used to direct education policy. Few researchers or administrators were able to find evidence that schools could be reformed in ways that would lead to improvements in student achievement. Some of the data produced by the IEA six-area study were later reexamined by Coleman, and the conclusions he reached were more instructive. I turn to that analysis after a review of Comber and Keeves's (1973) report on the international study of science achievement.

The First IEA Science Study

In the introduction to their study of science achievement in 14 countries, Comber and Keeves (1973) reviewed the other studies undertaken during the 1960s. They noted that the Plowden National Survey in England (Peaker, 1967, 1971), Project Talent (Flannigan, 1964), the Coleman report (1966), and the Mosteller and Moynihan reanalysis of Coleman's data (1972) had considerable influence on educational policy. In all of these inquiries, the influence of between-school variance had been found to be comparatively small. Comber and Reeves pointed out that science knowledge might be different from knowledge in other areas (e.g., reading and literature) in regard to degree of impact of the school, since achievement in science is more likely to be dependent on school instruction.

The researchers examined the role of schools and found that the effect of family background on science achievement differed greatly among the participating countries. (Note that this was not the same conclusion reached by Peaker.) They also found evidence that school practices do have an effect on learning. In particular, their measure of opportunity to learn (i.e., whether students were instructed in the topics of the test) was directly related to science score levels in middle school and high school when the separate science topics were delineated. The study confirmed that student background was the most important influence, accounting for approximately 37% of the variance between schools in science achievement.

Comber and Keeves (1973) pointed out that schools modify habits present early in life and that learning is a continuous and cumulative process over generations. They suggested that whatever influence schools have on achievement cannot be observed in the regression coefficients created from individual differences in students observed during a single year. Differences created by the influence of the home are stronger and thus more observable. Comber and Keeves (1973) concluded:

> As for the school factors, measurement of learning conditions within the school accounts for enough variation in achievement to support the argument, no longer taken as self-evident, that schools do have an impact on the learning of science. Home background is a good predictor of science achievement, though this varies among the countries. USA, England, and Scotland had the highest levels of family background explaining achievement in population 2 (14-year-olds). (p. 261)

The IEA science study of 1970 acknowledged that the influence of home background persisted as a strong indicator of total achievement through the age of 14 years. However, it was the first study to suggest that the effect of family background may be different for different countries and that measurement of school effects might depend on the subject being measured. The researchers attempted to locate the specific activities of schools that contributed to attainment independently of the home environment.

The IEA science study obtained teacher estimates of the degree to which test items had been covered in the classroom. This variable was examined for a number of countries and was found to be predictive of aggregate student science scores for each country (Comber & Keeves, 1973). This relationship suggests that what is taught in the schools directly influences subsequent test scores and that increasing the fit of the test to the curriculum, and vice versa, can increase achievement scores.

Coleman's Analysis of the IEA Studies

In 1975, James Coleman conducted a detailed reanalysis of the correlation coefficients among family background, student characteristics, and school factors reported for three subjects in the IEA six-area study and found that the IEA analysts had made a number of errors in the presentation of their results. He carefully re-created the analyses and summarized his results in three tables. His analysis compared the student achievement regression estimates for home background and school characteristics in the three subjects: reading, civics, and science. He also examined the results for two populations: 9-year-olds and 14-year-olds. Coleman (1975) concluded:

> These data show fairly conclusively, then, that reading achievement is more fully an outgrowth of home influences than either of the other two subjects (literature and science), less a function of what takes place at school. This is a rather important result, because it indicates that the general finding in this study and in others that home background is a much more powerful influence than school influences in determining achievement is a result that is subject-specific. The tests ordinarily used in examining school effects on achievement are reading or vocabulary. The results of Tables 2 and 3 suggest that, if other tests, less related to reading, were used, school effects would be larger than those estimated in previous studies. These studies, by using as a criterion that subject least affected by school variables and most affected by home variables, have probably underestimated the general effects of the school relative to the home. (p. 382)

Three important findings emerged from Coleman's secondary analysis. First, the strength of relationships between school factors and achievement depends on the topic of the test. Smaller school effects are found in reading than in science, a subject that is more likely to be directly taught at school than at home. Thus, Coleman's analysis shows how important it is to pay attention to the subject matter of the study.

Freudenthal (1978) had raised related concerns following the publication of results from the first international study of education that focused on mathematics achievement. He suggested then that a study of education should not regard mathematics as an outcome measure for all aspects of education and that mathematics itself has important variations that must be extricated and understood if we are to understand how schools and teachers influence students.

Second, the effects of schools on achievement are greater in some countries than in others. However, developing countries are even more likely than highly industrialized countries to have large differences in the effect of the school curriculum on achievement (see also Heyneman & Loxley, 1982).

Third, an especially large influence of family background on student achievement was noted in England. This finding supports the notion that national behavior affects how students learn and which institutions are likely to have the greatest effect on altering learning behavior.

The Second IEA Mathematics Study

Until 1982, studies had relied on indirect measures of curriculum and thus were unable to measure different effects of school curriculum choices on actual student performance. The Second International Study of Mathematics measured the opportunity to learn of eighth-grade students in mathematics by asking teachers to indicate whether test items had been covered in the classroom in that year or previously. This activity was strenuous for the teachers to perform; they had to examine 156 items and judge the probability that each item's content had been covered (Garden, 1987a). The results of the survey showed significant variation between countries in the number of items on the international test covered by the teachers. U.S. teachers tended to cover less of the test content than did those in many other countries. It was tempting to conclude that a causal relationship existed between the amount of a topic covered and the amount learned by students, each as measured by a set of test items that had been approved by a knowledgeable group of mathematics educators (McKnight et al., 1987). These empirical results certainly encouraged others to examine the effects of curriculum on achievement.

However, no attempt was made to directly compare statistical results with comparable models between the earlier international studies and the second mathematics study. The authors of the second study paid more attention to attributes of teacher instruction and curricular emphasis that might affect changes in student achievement within a single school year. For example, seven countries conducted pretests and posttests of student achievement for eighth-grade students over a year so that effects of curriculum differences could be related to student achievement growth or decline. This effort showed that countries stressing particular topics (e.g., algebra) exhibited large increases in student achievement on those topics (Burstein, 1989). This was one of the few studies that attempted to provide a baseline measure of growth in student achievement.

The TIMSS Curriculum Study

The TIMSS was specifically designed to identify components of mathematics and science stressed in each participating country. The analysis of that study is still under way; I report on studies in progress that help provide an understanding of whether and how choices made in what is taught in school might affect final achievement levels.

William Schmidt and colleagues (1996) conducted a study in six countries of how the school curriculum in mathematics and science was organized. The 3-year study, called the Survey of Mathematics and Science Opportunities (SMSO), exchanged data collection ideas with curriculum specialists in each of the six participating countries (France, Japan, Norway, Spain, Switzerland, and the United States) and resulted in plans to develop instruments to measure basic concepts of curriculum and of opportunities to learn mathematics and science. These instruments were implemented in the 1994 TIMSS.

The originating concepts were based on general ideas contained in previous IEA studies that any country's national curriculum can be defined by topics that are intended by the school system, implemented in the classroom, and attained by the students. The project began by defining a set of 40 topics in mathematics and 60 in science taught throughout the elementary and secondary school systems of the study countries. Only a few prior attempts had been made to experiment with new definitions of the content of curriculum; thus, there was little previous experience in regard to a measurement method for these concepts. The SMSO project involved many in-depth discussions and experiments on collection procedures with specialists in the six participating countries to improve measurement validity. The general hypothesis was that curriculum content decisions can be described as flowing from the national level to the classroom level, and thus national decisions about what content to teach in a particular grade affect the ultimate level of student achievement.

The SMSO produced a rich description of the decisions made by teachers in the six participating countries. The material collected demonstrated that the model used to create the data collection instruments was much too simplistic to adequately describe all of the processes of curriculum selection for each country. In a concluding set of remarks about descriptions of classroom content, Schmidt et al. (1996) wrote:

> Countries have developed their own ways of engaging students in the substance of mathematics and science. There appear to be strong cultural components, even national ideologies, in the teaching of these subjects. The French emphasis on formal knowledge occurs because secondary school teachers identify strongly with their disciplinary counterparts at universities. In Japan, primary children are exposed to disciplinary concepts in an informal but orchestrated way to prepare the child for more rigorous and formal study at lower secondary school. U.S. educators appear to prolong childhood by extending exposure to more basic science and mathematics topics into lower secondary school. The Swiss school system emphasizes local autonomy to the extent that both parents and teachers are involved with the education decisions of students. (p. 132)

The TIMSS continues with extensive data collection from the 49 countries that participated in the middle school aspects of the study in 1995. Preliminary analyses of the curriculum data provide some evidence of the differences between countries. Indicators chosen to represent the cumulative effect of decisions made by educators in each country include the number of mathematics or science topics taught at each grade level and the proportion of textbooks that emphasize understanding simple or complex information.

In an analysis of the relationship between curriculum and student achievement, Schmidt, Jakwerth, and McKnight (1998) presented further evidence that the organization of mathematics and science content and how it is presented in classrooms determine the level of student achievement on those topics. He showed that changes in student test scores for specific topic areas in mathematics and science between Grades 3 and 4 and between Grades 7 and 8 are higher in countries that stress more specific aspects of the content and lower in countries, such as the United States, that do not. Schmidt et al. also showed that all of the topics of mathematics and science are eventually covered in elementary school and secondary school; however, in the United States, many more of the topics are covered frequently throughout the school years. Furthermore, every country other than the United States exhibits a higher achievement level than all of the other countries in at least one area of emphasis. Thus, Schmidt et al. argued that the United States does not have a sufficiently concentrated focus on a given content area to ensure significant growth in achievement.

Coleman's observation of the importance of background in student achievement led some policymakers to assume that the diversity in U.S. ethnic populations leads to the country's lower achievement in comparison with other countries. According to Schmidt et al. (1998), the TIMSS suggests that this is not the correct interpretation. The variability in student achievement levels in the United States is about the same as that in other countries. At Grade 8, for example, Japan exhibits greater variations in TIMSS scores than the United States. In fact, the correlations of mean achievement levels with variance in achievement are about .5 for the fourth grade and .7 for the eighth grade. Across countries, greater diversity in student achievement is related to higher mean achievement levels. This means that average scores of whole countries can be high, regardless of the diversity of their populations.

Furthermore, a special tabulation of TIMSS results by Houang for the 1995 high school population provides evidence that the ranking of the United States in comparison with other countries at the end of secondary school is only slightly affected by the performance of minority groups. The scores for the White population only would rank U.S. performance at the end of secondary school in the lowest one third of all countries. Moreover, U.S. high school physics students were among the lowest performing of the 17 participating countries. Thus, student performance in the most selective science courses at the end of secondary school, representing the highest performing students in the United States, did not match the average achievement levels of students in other countries (Mullis et al., 1998).

Results from the TIMSS suggest that policies of assigning students to classes and effectiveness of instruction within classrooms contribute to U.S. achievement levels. An analysis of sources of variance within the United States for eighth-grade students shows that 45% of the variance in achievement in mathematics can be attributed to either the school or the class rather than to differences in student characteristics. Along with only 6 of 22 other countries (Australia, Belgium, Germany, Hong Kong, Netherlands, and Switzerland), the United States exhibits low levels of student variance in total test scores (Schmidt et al., 1998). Thus, student achievement policies are best focused on activities occurring within classrooms.

Schmidt et al. argued that policymakers are misled by the general scores of national tests that "scale" a large number of items into a single index of achievement. A single score suggests that all of mathematics or science learning can be expressed through the results of a single test. However, various components of these subjects are learned with different degrees of efficiency in different countries. By disaggregating TIMSS test scores into specific detailed content areas, it is possible to detect differences between countries in emphasis of the intended and implemented curriculums and the effect on student performance.

Schmidt et al.'s analysis of international differences in coverage of material in textbooks and classrooms showed that decisions made in U.S. elementary schools about mathematics and science topics result in an extremely variable pattern of school offerings. His analysis related specific school policies regarding textbook coverage, teacher presentation methods, and national standards to direct changes in student performance. Other countries have managed to maintain a focus on topics such as geometry and algebra at early ages and a focus on physical science concepts as early as middle school.

This review of the recent evidence from TIMSS that activities carried out within schools affect student performance is not yet complete. Final results of the analysis of the role of curriculum have not been published, and some unanswered questions remain. For example, some countries (e.g., Germany) appear to have a more focused curriculum but do not have accompanying higher achievement levels. Conversely, other countries (e.g., Switzerland) have a very dispersed and unfocused curriculum, yet they evidence high achievement levels for all students. The relationships between achievement and measures of topic coverage are not so simple to unravel as the first attempts to measure opportunity to learn suggest. The information on curriculum gathered in the TIMSS and the many analyses that remain to be done are likely to increase our understanding of the role of curriculum and training for teachers.

Discussion and Conclusion

The following questions were raised at the beginning of this review: Can student achievement, as measured by national assessments, be altered by administrating changes in school system practices? Or is student achievement too strongly established by parental background to be affected by practices in schools? Do results from international comparative studies provide evidence that changing curriculum content can increase student achievement? The purpose of this article has been to review four decades of research on student assessment conducted across nations to find how international studies have addressed these questions.

The present review was conducted with the knowledge that indicators of student performance in the United States have been changing slowly (see Figure 1) in spite of efforts at large-scale school reform. Moreover, many observers of trends in the past 25 years have not been convinced by evidence of the influence of school practices. For example, in a recent review of research on teaching-centered instruction, Branson (1998) cited observations that student performance had reached an upper limit in the 1950s. According to Bracey (1987), "The lines on a graph of average student performance are as flat as the surface of a frozen lake" (p. 109). Researchers who were contemplating the effect of schooling on student achievement following the first round of national and international studies in the 1960s also seemed to conclude that student achievement was immutable (Peaker, 1975).

This review has shown that international comparative studies of student achievement have led to the development of new methods for studying the impact of schooling. These studies have provided a basis for questioning previous conclusions that school practices do not influence student achievement. While no experiment in education has been devised that randomly assigned students, schools, and teachers to different practices, the international studies have taken advantage of many "natural experiments" and have challenged conclusions from studies carried out in a single cultural setting.

International studies have not always produced the most insightful analyses of their own data, and thus they have not always had a strong impact on research questions or policy considerations. Sometimes secondary analyses of their data have led to improved interpretations of policy implications. That researchers did not always pay close attention to the discoveries apparent in their data is illustrated by Coleman's reanalysis of the IEA data.

The international comparative studies have contributed to educational research by improving the questions posed by researchers about how schools operate. The observations of cultural differences provided by the international studies lead to a richer examination of the nature of schooling than is possible within a single country. International comparisons provide a better focus than most other studies on the content of the curriculum. Thus, the international studies were the first to show cross-national differences in the effects of school influences on those subject areas that are most likely learned in school and less influenced by home background (Coleman, 1975; Walker, 1976).

Empirical measurement of curriculum aspects was first introduced by the IEA studies as a way of explaining cross-national differences in performance on a set of test items (Harnqvist, 1987). Measurement of curriculum content areas led to the conceptual model of intended, implemented, and attained curriculums and became an important framework for studying national differences in achievement (Travers & Westbury, 1990). Studies of curriculum coverage in mathematics and science carried out during TIMSS have identified large differences between countries in regard to the coverage of science or mathematics topics (Schmidt et al., 1996; Schmidt, McKnight, Cogan, Jakwerth, & Houang, 1999). While these cross-sectional studies suggest that student achievement might be a result of the organization of the curriculum, the causal link has yet to be completely demonstrated. In fact, specific indicators of curriculum such as number of topics covered and length of time spent on a topic have now been shown to be insufficient in explaining between-countries differences in achievement. Further studies of country differences in curriculum and achievement are under way to provide a better understanding of the conditions under which a linkage between aspects of intended coverage of a subject and student achievement can be made (Schmidt et al., 1999).

The fact that the average level of student achievement in some countries is higher than in other countries itself suggests the possibility that events other than background (e.g., organization of school systems) may have some impact on student performance. The 1995 TIMSS results for U.S. science achievement are especially striking, since student performance in 4th grade was about as high as the highest performing country, while the performance of 8th-grade students in science was lower (at about the international average). The performance of 12th-grade students was even lower than that of 8th-grade students and was, in fact, below the international average (National Science Board, 2000). Such differences in relative performance levels for these different grades cannot be accounted for by economic conditions or background characteristics of students. More likely, some aspect of the United States' method of implementing the science curriculum leads to relatively lower science achievement levels as students spend more time in the school system.

The TIMSS addressed curriculum differences with more statistical rigor than it applied to teacher characteristics and teaching practices and, thus, may not have taken into account all of the critical aspects of influences on student performance. The video studies of Stigler, Gonzales, Kawanaka, Knoll, and Serrano (1999) provided new insights into how teachers behave when engaging students directly, but these findings were not integrated into a single model that could be related to aggregate student achievement levels. However, examination of those tapes easily leads to the hypothesis that teacher subject matter knowledge may be more important than curriculum content. The 1999 repeat of TIMSS included new information on how well teachers perceive their own performance. Experimental studies linked to known policy changes in schools will be necessary to ultimately determine whether student achievement is actually immutable or can be influenced by policy. Thus, the questions posed here have a partial answer. The research associated with international studies has helped provide a basis on which to conduct further research into the causes and consequences of student achievement differences.

International studies have provided both a stimulus and a road map for understanding individual differences in achievement. New observations, as with the two-subject, three-population study of TIMSS, have added insights as well as complexity to the investigation. The role of curriculum has been more clearly specified, but the relative importance of other aspects of teaching practices has not yet been as successfully measured. New studies of teaching practices have been designed in connection with international studies that may extend our understanding of the role of teaching methods as well as curriculum and standards in student learning (Stigler et al., 1999).

References

Barton, P. E., & Coley, R. J. (1998). *Growth in school: Achievement gains from the fourth to the eighth grade*. Princeton, NJ: Educational Testing Service.

Bracey, G. W. (1987). Why the schools can't improve: The upper limit hypothesis. *Journal of Instructional Development, 10,* 7–12.

Braddock, J. H., & Slavin, R. E. (1993). Why ability grouping must end: Achieving excellence and equity in American education. *Journal of Intergroup Relations, 20,* 51–64.

Branson, R. K. (1998). Teaching-centered schooling has reached its upper limit: It doesn't get any better than this. *Current Directions in Psychological Science, 7*(4), 126–135.

Brewer, D. J., Rees, D. I., & Argys, A. M. (1995). Detracking America's schools: The reform without cost? *Phi Delta Kappan, 77,* 210–215.

Brophy, J., & Good, T. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching*. New York: Macmillan.

Burstein, L. (1989). *Second International Study of Mathematics* (Vol. 3). London: Pergamon Press.

Campbell, J., Reese, C., O'Sullivan, C., & Dossey, J. (1996). *NAEP 1994: Trends in academic progress*. Washington, DC: National Center for Education Statistics.

Coleman, J. S. (1975). Methods and results in the IEA studies of effects of school on learning. *Review of Educational Research, 45,* 335–386.

Coleman, J. S., & Associates. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Department of Health, Education, and Welfare.

Comber, L. C., & Keeves, J. P. (1973). *Science education in nineteen countries: An empirical study*. Stockholm: Wiley.

Eggleston, J. (1977). *The sociology of the school curriculum*. London: Routledge & Kegan Paul.

Flannigan, J. C. (1964). *The American high school student: The identification of development and utilization of human talents*. Pittsburgh, PA: Project Talent Office, University of Pittsburgh.

Freudenthal, H. (1978). *Weeding and sowing: Preface to a science of mathematical education.* Amsterdam: Reidel.

Garden, R. (1987a). *The opportunity to learn.* Washington, DC: National Center for Education Statistics.

Garden, R. (1987b). Second IEA mathematics study. *Comparative Education Review, 31,* 47–68.

Gustafsson, J.-E., & Undheim, J. O. (1996). Individual differences in cognitive functions. In D. Berliner & R. Calfee (Eds.), *Handbook of research in educational psychology* (pp. 243–310). New York: Macmillan.

Harnqvist, K. (1987). The IEA revisited. *Comparative Education Review, 31,* 48–55.

Heyneman, S. P., & Loxley, W. A. (1982). Influences on academic achievement across high and low income countries: A re-analysis of IEA data. *Sociology of Education, 55,* 13–21.

Hoffer, T. B. (1992). Middle school ability grouping and student achievement in science and mathematics. *Educational Evaluation and Policy Analysis, 14,* 205–227.

Husen, T. (Ed.). (1967). *International study of achievement in mathematics* (Vol. 1). New York: Wiley.

Husen, T. (1979). An international research venture in retrospect: The IEA surveys. *Comparative Education Review, 23,* 371–385.

Husen, T. (1987). Policy impact of IEA research. *Comparative Education Review, 31,* 29–47.

McKnight, C., Crosswhite, F., Dossey, J., et al. (1987). *The underachieving curriculum: Assessing U.S. school mathematics from an international perspective.* Champaign, IL: Stipes.

Mosteller, F., & Moynihan, D. P. (1972). *On equality of educational opportunity.* New York: Random House.

Mullis, I., Martin, M., Beaton, A., Gonzalez, E., Kelly, D., & Smith, T. (1998). *Mathematics and science achievement in the final year of secondary school: IEA's Third International Mathematics and Science Study (TIMSS).* Boston: Center for the Study of Testing, Evaluation, and Education Policy, Boston College.

National Center for Education Statistics. (1998). *Linking The National Assessment of Educational Progress (NAEP)and the Third International Math and Science Study (TIMSS): Eighth-grade results.* Washington, DC: Author.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* Washington, DC: U.S. Government Printing Office.

National Science Board. (1998). *Science and engineering indicators: 1998.* Arlington, VA: National Science Foundation.

National Science Board. (2000). *Science and engineering indicators: 2000.* Arlington, VA: National Science Foundation.

Oakes, J. (1992). Can tracking research inform practice? Technical, normative, and political considerations. *Educational Researcher, 21*(6), 17.

Peaker, G. (1975). *An empirical study of education in twenty-one countries: A technical report.* New York: Wiley.

Robitaille, D. F., Schmidt, W. H., Raizen, S., McKnight, C., Britton, E., & Nocol, C. (1993). *Curriculum frameworks for mathematics and science.* Vancouver: Pacific Educational Press.

Schmidt, W. H., Jakwerth, P. M., & McKnight, C. C. (1998). Curriculum-sensitive assessment: Content does make a difference. *International Journal of Educational Research, 29*(6).

Schmidt, W. H., Jorde, D., Cogan, L., Barrier, E., Gonzalo, I., Moser, U., Shimizu, K., Valverde, G., McKnight, C., Prawat, R., Wiley, R., Raizen, S., Britton, T., & Wolfe, R. (1996). *Characterizing pedagogical flow: An investigation of mathematics and science teaching in six countries.* Dordrecht, the Netherlands: Kluwer.

Schmidt, W., McKnight, C., Cogan, L., Jakwerth, P., & Houang, R. (1999). *Facing the consequences: Using TIMSS for a closer look at U.S. mathematics and science education*. Dordrecht, the Netherlands: Kluwer.

Stigler, J., Gonzales, P. A., Kawanaka, T., Knoll, S., & Serrano, A. (1999). *The TIMSS Videotape Classroom Study: Methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan, and the United States*. Washington, DC: National Center for Education Statistics.

Suter, L. (1995). *Indicators of science and mathematics education 1995*. Arlington, VA: National Science Foundation.

Travers, K. J., & Westbury, I. (1990). *The IEA study of mathematics I: Analysis of mathematics curricula*. Oxford, England: Pergamon Press.

Walker, D. A. (1976). *The IEA six subject survey: An empirical study of education in twenty-one countries*. New York: Wiley.

Wittrock, M. C. (Ed.). (1986). *Handbook of research on teaching* (3rd ed.). New York: Macmillan.

Author

LARRY E. SUTER is Program Director (Statistics), Division of Research, Evaluation and Communication, National Science Foundation, 4201 Wilson Blvd., Arlington, VA 22230; Lsuter@nsf.gov. He specializes in educational research.