0

Overview of Procedures and Results

Introduction

In the constant struggle to improve education in countries everywhere, the issue of school effectiveness has attracted considerable attention in recent years.[1] School effectiveness as an area of study seeks to improve educational practice by studying how schools discharge their role as institutions for learning and instruction, and in particular what makes for a successful school. Since in every country, schools are almost universally the primary institutions for student learning, discoveries about the essentials of effective schooling have the potential to have a great impact, particularly if they can help to raise the performance of the low-achieving schools to match the schools with the highest levels.

Although at first glance it might seem that effective schools are simply those with high average student achievement, since the pertinent literature makes it clear that often high achievement depends mainly upon the composition of the student intake, it is important to take into account the difficulty of the educational task when evaluating the effectiveness of a school. Schools with a high proportion of well-prepared students from homes and communities with strong support for learning are already well on the way to high achievement levels, regardless of the contribution of the school in terms of instruction, facilities and support. Schools in less-advantaged circumstances face a more difficult challenge. Accordingly, studies of school effectiveness typically attempt to disentangle the effects of the organizational and instructional practices of the school from the effects of the abilities and level of preparation of the student body prior to entering the school. Although this can never be completely successful, advances in statistical methodology using hierarchical linear modeling (HLM) have provided powerful techniques for this endeavor.

While a great deal of work has already been accomplished in a range of countries, both with national data and with data from international studies, the data provided by the Third International Mathematics and Science Study (TIMSS)[2] offer unprecedented opportunities for cross-national analyses of school effectiveness. Based on a collaborative venture involving 39 countries, and sponsored by the International Association for the Evaluation of Educational Achievement (IEA), TIMSS is the most ambitious and complex comparative education study undertaken to date. With student achieve-

1 Wyatt, T. (1996), “School Effectiveness Research: Dead End, Damp Squib or Smoldering Fuse?” *Issues in Educational Research*, 6(1), 79-112.

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

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

ment data in mathematics and science collected with the same instruments and following the same procedures in each country, TIMSS provides a cross-national platform of unusual breadth for investigating effective schooling in science and mathematics.

This report presents the findings of an exploratory study of school effectiveness, using data from TIMSS for eighth-grade students. This report has two parts.

- As a starting point to identifying characteristics of effective schools, the first part of the study divided schools in each country into high-performing and low-performing groups on the basis of average student achievement in eighth-grade mathematics and science, and then looked for variables that discriminated between the two groups. Variables that were characteristic of high-performing schools but not of low-performers were retained for further analysis in the second part of the study.

- Building on this work, the second set of analyses sought to identify attributes of effective schools, i.e., those characteristics of schools in each country that were associated with high student achievement even after adjusting statistically for the effect of students' home background on achievement. These analyses made use of hierarchical linear modeling techniques.

Which Countries Were Included in the Report?

Countries participating in TIMSS were required to administer mathematics and science tests to the two adjacent grade levels in their systems containing the most 13-year-old students.[3] This report is based on data from the upper of these two grades, which was the eighth grade in most countries. Exhibit 1 provides information about the grade tested in each country.

Having valid and efficient samples in each country is crucial to the quality and success of any international comparative study.[4] TIMSS developed procedures and guidelines to ensure that the national samples were of the highest quality possible. Standards for coverage of the target population, participation rates, and the age of students were established, as were clearly documented procedures on how to obtain the national samples. For the most part, the national samples

[3] For countries also wanting to participate at the primary and secondary school levels, TIMSS tested students in the two grades with the most 9-year olds (third and fourth grades in most countries), and in the final year of secondary school (twelfth grade in the U.S. and many countries). These data were not included in this report.

[4] The technical aspects of TIMSS are described in a series of technical reports:
Martin, M.O. and Kelly, D.L. (Eds). (1996). *Third International Mathematics and Science Study (TIMSS) Technical Report Volume I: Design and Development*. Chestnut Hill, MA: Boston College.

Martin, M.O. and Kelly, D.L. (Eds.). (1997). *Third International Mathematics and Science Study (TIMSS) Technical Report Volume II: Implementation and Analysis in the Primary and Middle School Years*. Chestnut Hill, MA: Boston College.

Martin, M.O. and Mullis, I.V.S. (Eds.). (1996). *Third International Mathematics and Science Study (TIMSS): Quality Assurance in Data Collection*. Chestnut Hill, MA: Boston College.

were drawn in accordance with the TIMSS standards, and achievement results can be compared with confidence. The 34 countries that satisfied the TIMSS standards with approved sampling procedures at the classroom level are included in the first part of this report. Appendix A (Exhibit A.1) shows the participating countries grouped according to the degree of compliance with the guidelines for sample implementation and participation rates.

What Was the Nature of the Data?

TIMSS was very much a collaboration among countries. Each participant designated a national center to conduct the activities of the study and a National Research Coordinator (NRC) to assume responsibility for the successful completion of these tasks. For the sake of comparability, all testing was conducted at the end of the school year. The four countries on a Southern Hemisphere school schedule (Australia, Korea, New Zealand, and Singapore) tested the mathematics and science achievement of their students in September through November of 1994, which was the end of the school year in the Southern Hemisphere. The remaining countries tested at the end of the 1994-95 school year, most often in May and June of 1995.

The Achievement Tests in Science and Mathematics

Together with the quality of the samples, the quality of the test also receives considerable scrutiny in any comparative study of the magnitude and importance of TIMSS. All participants wish to ensure that the achievement items are appropriate for their students and reflect their current curriculum. Developing the TIMSS tests was a cooperative venture involving all of the NRCs during the entire process. Through a series of efforts, countries submitted items that were reviewed by subject-matter specialists, and additional items were written to ensure that the desired topics were covered adequately. Every effort was made to ensure that the tests represented the curricula of the participating countries and that the items did not exhibit any bias toward or against any country.

Six content areas were covered by the mathematics tests taken by the eighth-grade students. These areas, and the percentage of test items devoted to each, include fractions and number sense (34%); measurement (12%); proportionality (7%); data representation, analysis, and probability (14%); geometry (15%); and algebra (18%). The eighth-grade science test consisted of just five content areas: earth science (16%); life science (30%); physics (30%); chemistry (14%); and environmental issues and the nature of science (10%). About one-fourth of the questions were in free-response format, requiring stu-

Exhibit 1 Grade Tested in Participating Countries – Eighth Grade

Country	Grade Tested
Australia	8 or 9
Austria	4, Klasse
Belgium (Flemish)	2A & 2P
Belgium (French)	2A & 2P
Canada	8
Colombia	8
Cyprus	8
Czech Republic	8
England	9
France	4 eme (90%) or 4 eme Technologique (10%)
Germany	8
Hong Kong	Secondary 1
Hungary	8
Iceland	8
Iran, Islamic Rep.	8
Ireland	2nd Year
Japan	2nd Grade Lower Secondary
Korea, Republic of	2nd Grade Middle School
Latvia (LSS)	8
Lithuania	8
Netherlands	Secondary 2
New Zealand	Form 3
Norway	7
Portugal	Grade 8
Romania	8
Russian Federation	8
Scotland	Secondary 2
Singapore	Secondary 2
Slovak Republic	8
Slovenia	8
Spain	8 EGB
Sweden	7
Switzerland	7 (German); 8 (French and Italian)
United States	8

dents to generate and write their answers. These questions, some of which required extended responses, were allotted approximately one-third of the testing time. Responses to the free-response questions were evaluated to capture diagnostic information, and some were scored using procedures that permitted partial credit.[5]

The TIMSS tests were prepared in English and translated into 30 additional languages using explicit guidelines and procedures. A series of verification checks were conducted to ensure the comparability of the translations.

There were 135 science items and 151 mathematics items developed for the eighth-grade TIMSS tests. The tests were organized so that no one student took all of the items, which would have required more than three hours. Instead, the tests were assembled in eight booklets, each requiring 90 minutes to complete. Each student took only one booklet, and the items were distributed across the booklets so that each item was answered by a representative sample of students.

The Questionnaires

To provide an educational context for interpreting the achievement results, TIMSS used questionnaires to collect descriptive information from students, their teachers, and the principals of their schools. In all, the questionnaires provided data on approximately 1500 variables, and, together with the achievement results, formed the basis for the analyses presented in this report.

The student questionnaire elicited information from the students about resources for learning in their homes, their attitudes towards mathematics and science, and their learning experiences in school. With regard to schooling, the questionnaire asked about the frequency of occurrence of a range of classroom instructional activities, and the students' perception of their school's social climate.

TIMSS administered questionnaires to mathematics and science teachers to gather information about their backgrounds, education and training, and attitudes towards mathematics and science. The questionnaires asked how the teachers divide their time among their teaching tasks, about their level of preparation to teach specific subject matter, and about the instructional approaches that they use in their classrooms. Information also was collected about the materials used in instruction, the activities students do in class, the use of calculators and computers, the role of homework, and the reliance on different types of assessment approaches.

5 TIMSS scoring reliability studies within and across countries indicate that the percent of exact agreement for correctness scores averaged over 85%. For more details, see Mullis, I.V.S. and Smith, T.A. (1996). "Quality Control Steps for Free-Response Scoring," in M.O. Martin and I.V.S. Mullis (eds.), *Third International Mathematics and Science Study: Quality Assurance in Data Collection*. Chestnut Hill, MA: Boston College.

The school questionnaire asked principals to provide information about the school's location, organization and structure, and resources for learning.

How Was the Analysis Conducted?

As a first step in preparing this report, TIMSS researchers reviewed the contents of the TIMSS database in the light of the effective-schools literature to identify variables that were likely to characterize effective schools. These variables were correlated with student achievement in science and mathematics in an extensive exploratory analysis. Variables that were significantly related to achievement were retained for further study. This exercise reduced the number of variables under consideration to fewer than 100. Where possible, individual variables were combined to form an index that was more global and more stable than the original variables. For example, the school questionnaire contained several questions that pertained to student misbehavior. On the basis of a principal component analysis, these variables were combined to provide two indices of student misbehavior: "student administrative violations" and "serious student misconduct."

For the analyses reported in the first chapter of this report, schools within each country were first ranked by their average achievement, separately for mathematics and science. Schools in the top third of the average achievement distribution were assigned to the high-achieving group, and those in the bottom third were assigned to the low-achieving group. Again, this was done separately for mathematics and science. The idea was to work through the variables and indices identified in the exploratory stage to see which of them could discriminate effectively between the high-achieving schools and the low-achieving schools. Each variable and index was dichotomized at a point that seemed to maximize the discrimination between the two groups of schools, and a t-test was applied to the data from each country to determine whether the frequency of occurrence of the variable differed significantly between the two groups. Variables and indices that showed significant differences in most of the participating countries, or showed particularly big differences in a few countries, were included in this report.

Contrasting the characteristics on which high- and low-achieving schools differ most is a useful device for highlighting areas that might prove fruitful for further study of school effectiveness. However, the fact that a school is in the low-achieving group in the study does not necessarily mean that it is ineffective in providing mathematics or science instruction. The educational burden varies from school to school, and so it is possible that a school that is not well-resourced, and that has a socially and economically disadvantaged student body, might be very effective in overcoming its handicaps without managing to raise average student achievement sufficiently to make the high-achieving group. Such a school still would be regarded as an effective one. In contrast, a well-resourced school in an affluent area might be "resting on its laurels" and producing average student achievement below what could be expected given its student intake. Such a school could make it into the high-achieving group largely on the strength of the student body, but might be regarded as less than effective as an organization for teaching and learning.

Seen in this light, a school is effective to the extent that it "adds value" by realizing the potential of the student body through efficient organization and effective instruction. From this perspective, the student body may be considered the raw material that the school has to work with as an organization for promoting learning. If all schools had students with the same initial level of advantage and preparation, then school effectiveness would simply be a matter of comparing average student achievement at the end of the school year. However, since schools in many countries vary considerably in the composition of their student bodies, any study of school effectiveness must take this fact into account.

In the second part of this report, school effectiveness was studied through a multilevel analysis that examined the relationship between a range of school and student variables and student achievement, while simultaneously adjusting for differences in the home background of the students. More specifically, the analysis compared the efficacy of a series of models of home, home/school, and school factors in accounting for the adjusted (for student home background) difference between schools in achievement. The multi-level analysis was conducted using the HLM program.[6]

For the modeling of school effectiveness, the analysis concentrated on countries where there was substantial variation between schools in average achievement. In countries such as Japan, Korea, Norway, and Cyprus, the difference between schools was very small, and so there was little to be gained from an analysis of school differences. Accordingly, these countries were not included in the multilevel analysis. Also excluded were countries with relatively high levels of missing data, or with low average student achievement. In all, 18 countries were included in the multilevel analysis for mathematics, and 14 for science.

Since this report is entitled "Effective Schools in Science and Mathematics," it is important to be aware of how schools were characterized in the data. The basic TIMSS sampling plan called for a random sample of approximately 150 schools in each country, and for at least one intact mathematics class in each sampled school. In most

[6] See Appendix A for details.

countries, therefore, a school was represented by a single, randomly selected, mathematics class. In Australia, Cyprus, Sweden, and the United States, however, two mathematics classes were sampled in each school. Each was treated as a separate unit for the purposes of the analyses in this report. In England, students were sampled at random from across the entire grade, without reference to classes.

Whereas in most countries the school average in mathematics is based on the students in a single mathematics class, in science the situation is more complicated. School averages in science also were based on the students in the sampled mathematics class, but while sometimes this group also formed an intact science class, frequently it was made up of students from a range of eighth-grade science classes in the school. If this study had been about mathematics only it might have made sense to think of classes rather than schools as the unit of analysis, but this was not possible for science. Consequently, it was decided to talk about schools, while keeping in mind that school and class effects cannot be analytically disentangled.

The Structure of This Report

The following two chapters present the findings for the two sets of analyses conducted for this report. A procedural appendix, Appendix A, describes the methods used to perform the analyses presented in Chapters 1 and 2. Appendices B and C present detailed multilevel analysis results for science and mathematics, respectively, for participating countries.

Summary of Results

The contrast between the highest- and lowest-achieving schools in science and mathematics in each country showed that home background indicators of socioeconomic status and of parental support for academic achievement most consistently distinguished between the two groups of schools. In almost all countries, students in the high-achieving schools had higher levels of book ownership, study aids, possessions in the home, and parental education, and spent less time working in the home. Another distinguishing factor, related to the home, was student aspirations for higher education. In most countries, plans to attend university after secondary school were much more frequently reported by students in the high-achieving schools.

Factors more directly related to the school were less uniformly effective in distinguishing between the high- and low-achieving schools. Although factors such as school size and location, school social climate, student attitude to science and mathematics, and instructional activities in science and mathematics class did discriminate between the high- and low-achieving schools in some countries, few school variables worked consistently across all countries. This indicates that analyses of characteristics of effective schools are likely to be most fruitful using different variables in different countries, or groups of countries, rather than common variables that operate in the same way across all countries.

The results presented in the second chapter show that the extent to which achievement in science and mathematics can be related to school factors varies considerably from country to country, and that the extent to which schools differed in the home background of their students also is not the same in all countries. It is clear that the way student home background relates to student achievement, and the way the school system moderates or magnifies this relationship, are closely linked to societal and school organizational factors unique to each country, and any cross-national analytic efforts should take this into account.

Although only a small set of classroom-related variables survived the variable-selection process, they accounted for quite a large proportion of the differences between schools in most countries. The most prominent indicator was doing daily homework in a range of subjects (language, mathematics, and science). Schools where eighth-grade students were expected to spend time on homework in a range of subjects had higher average achievement in science and mathematics, even after adjusting for the home background of the students in the school. Teacher characteristics, school social climate, and demographic characteristics such as school location and class size were less consistent predictors of achievement across countries. Among variables that arguably may be influenced by both the home and the school (the home-school interface), the average level of students' aspirations for further education was a significant predictor of school achievement in science in most countries and in mathematics in almost all countries.

While the results show that classroom-related variables are related to average school achievement even after adjusting for the home background of the students in the school, the strong relationship that persists between the average level of home background and adjusted student achievement also serves as a reminder that, in many countries, home background, schooling, and student achievement are closely intertwined, and that teasing out the influences of the various contributing factors remains a major challenge.