DOCUMENT RESUME

ED 244 822 SE 044 547

TITLE Second Study of Mathematics. Bulletin No. 4.
INSTITUTION International Association for the Evaluation of Educational Achievement, Wellington (New Zealand).
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Dec 79
GRANT NIE-78-0068
NOTE 160p.; For Bulletin No. 3, see ED 176 948.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC07 Plus Postage.
DESCRIPTORS *Curriculum Development; *Educational Assessment; Elementary Secondary Education; *International Programs; *Mathematics Achievement; Mathematics Curriculum; *Mathematics Education; Mathematics Instruction; *Research Methodology; Research Projects; Student Attitudes
IDENTIFIERS *International Assn Evaluation Educ Achievement; International Mathematics Committee

ABSTRACT

The Mathematics Bulletin Series of the International Association for the Evaluation of Educational Achievement (IEA) is intended to serve a number of purposes: (1) to consolidate the decisions of the International Mathematics Committee; (2) to provide a historical record of the development of the Second IEA Mathematics Study; and (3) to provide information and guidelines for National Centers and National Mathematics Committee members. The Second IEA Mathematics Study, which is an investigation of the teaching and learning of mathematics in schools, consists of three main components: a curriculum analysis, an examination of classroom processes, and an analysis of cognitive and affective student outcomes in light of the nature of the curriculum and instructional practice. Previous bulletins have described the background of the Second IEA Mathematics Study, reported on various developments concerning the study, and elaborated on the study's design. This bulletin describes the updated model for the study, gives results of pilot testing of the cognitive instruments used, reports on the sampling and methodological issues associated with the two proposed versions (cross-sectional and longitudinal) of the study, and gives further timetable amendments to allow for a full pilot study of the longitudinal version. (JN)

Second Study of Mathematics
Bulletin Number 4
December 1979

This document was produced under Grant 78-0068 of the National Institute of Education, Washington, D. C., U.S.A. The opinions expressed in this publication do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

1. Introduction

The Mathematics Study Bulletin series of the International Association for the Evaluation of Education Achievement (IEA) is intended to serve a number of purposes:

1. To consolidate the decisions of the International Mathematics Committee.
2. To provide a historical record of the development of the project.
3. To provide information and guidelines for National Centers and National Mathematics Committee members.

Although the bulletins are written primarily for use within the framework of IEA countries and committees, they may at times be found useful for general informational purposes.

Three previous Mathematics Study bulletins have already been published. Bulletin No. 1 (October 1976), described the background and evolution of the Second IEA Mathematics Study, the issues to be addressed, and the procedures proposed to address those issues, and outlined a tentative timetable for the Study.

Bulletin No. 2 (September 1977), reported on developments concerning the Study, including funding and organizational details, discussed activities of the International Mathematics Committee, provided an international grid summarizing initial responses from countries about their mathematics curricula, and presented an updated timetable for the Project.

Bulletin No. 3 (December 1978), elaborated upon the design for the Study, provided a detailed timetable, and summarized sampling and instrument specifications. The document also contained sufficient information for National Centers to prepare proposals and proceed with other arrangements preparatory to participation in the Study.

This bulletin describes the updated model for the Study, gives results of pilot testing of the cognitive instruments, reports on the sampling and methodological issues associated with the two proposed versions (cross-sectional and longitudinal) of the study and gives further timetable amendments to allow for a full pilot study of the longitudinal version. The results of the attitude trials together with the recommended items for the attitude scales for the cross-sectional study are included in a separate publication, "Results of the Attitude Trials" (December 1979).

2. Activities during 1979

2.1 Methodology Seminars

The Second International Mathematics Study raises many methodological issues, predominant among them being the matter of student growth in achievement during the school year. How should growth be measured? What analyses shall be employed to account for growth? What are the implications of this interest in growth for sampling and for the structure of the instruments, particularly the cognitive test?

As a response to this important problem, distinguished methodologists from various countries including Japan, Sweden, Australia, New Zealand, Canada and the United States were invited to prepare papers to address the growth issue.

Early in the planning, it became evident that it was impracticable to attempt to gather together the principle participants in one location. Hence, it was decided to convene two somewhat parallel sessions: one at Michigan State University, USA, and another at the Australian Council for Educational Research, under the direction of John R. Schwille and John P. Keeves, respectively.

Michigan State Seminar: February 1-2, 1979

Overview of the Second IEA Mathematics Study.
K. J. Travers, University of Illinois

Overview of Descriptive Analyses for the Study.
E. Kifer, University of Kentucky

Relating Classroom Process Variables to Measures of Student Achievement and Learning in the Second IEA Mathematics Study.
Graham Nuthall, University of Canterbury

Nuthall discussed a number of apparent contradictions between the logic of relationships that exist between significant classroom process variables and the analytical procedures which are available and advocated a procedure based on student performance on individual items. He argued that the appropriate form of summary aggregation of data is across students, and not across items, using the item scores as records of what is known, or not known, rather than as measures of a variable. The paper was presented by Professor Kifer in Professor Nuthall's absence.

Explanatory potential of teacher topic questionnaires
Curtis McKnight, University of Illinois

McKnight analysed the topic specific questionnaires with a view to identifying variables possessing explanatory power.

Test construction issues - A practical problem and a proposed solution
Richard M. Wolf, Teachers College, Columbia University

Wolf discussed the problem of needing a large number of items to sample the proposed International Grid while having only a limited testing time available in schools. He examined alternative solutions and advocated multiple matrix sampling as a viable solution, outlining the advantages and disadvantages of the method.

Structural Equations and multivariate least squares; fitting different methods to different questions

W. Schmidt, Michigan State University

Schmidt contrasted multivariate least squares techniques, including repeated measures models, and structural equations models based on maximum likelihood procedures. An explanation of the two methods and examples of research questions posed in the study for which the methods would be appropriate were presented.

A general model of task learning

Richard D. Noonan, University of Stockholm

Noonan presented a task learning model devised to attempt to solve the problem: given a data set including a variety of student background, affective and cognitive measures, as well as teacher and school measures at two points in time (i.e., beginning and end of school year), how can the data be analysed to shed light on the process of cognitive growth?

Mathematics achievement in longitudinal models using ridge regression estimation procedures

J. W. Bulcock and W.F. Lee, Memorial University, Newfoundland, and W.S. Luk, Simon Fraser University, Canada.

In the paper presented, the limitations of least squares regression and two stage least squares procedures are examined. Simple ridge regression is examined as a possible solution but it is found that the minimum mean square error is not a proper criterion. However, the variance normalization criterion is found to correct all the limitations of simple ridge regression with the former critierion. When simple ridge regression using the variance normalization criterion is used in conjunction with two stage least squares, new insights into the reciprocal structure of the cognitive, affective and conative outcomes of schooling are gained.

Several implications of these findings - both substantive and methodological - for the Second IEA Mathematics Study were discussed.

The IEA Longitudinal analysis from a multilevel perspective: disentangling between-class and within-class relationships

Leigh Burstein, University of California, Los Angeles and Robert L. Linn, University of Illinois

The authors discussed how the multilevel character of the data should influence the longitudinal analysis in the Second IEA Mathematics Study and described how a multilevel perspective in the specification of substantive questions and in the analysis strategies employed can potentially clarify the ways in which instructional practices affect student performance. They outlined the way in which a multilevel analysis of the Study data might proceed and presented both the general questions that the multilevel analysis can address and specific features of a possible analytical framework.

Discussants for the symposium were:
Annagret Harnischfeger (CEMREL)
David Berliner (University of Arizona)
Andrew Porter (Michigan State University)

Australian Council for Education Research Seminar: February 19-20, 1979

The seminar referred to many of the papaers delivered at the Michigan State meeting and received presentations by the following:

Rosier, Malcolm. "Planned analyses for the Second IEA Mathematics Study-Cross-sectional Study in Australia." February 1979.

"Report of the IEA Second International Mathematics Study Data Analysis Seminar." Melbourne, ACER, February 1979.

Keeves, J.P. and R. Lewis. "Teachers, classrooms and student outcomes."

Rosier, M. "Hypotheses for the Australian National Study of Mathematics Achievement."

Capitalizing upon the visits to Australia of Neville Postlethwaite, Roy Phillipps, Robert Garden, Ian Livingstone, and Roslyn Slemint, extensive work was done on the sampling manual and administrative manuals for the reduced (cross-sectional) study. An impressive amount of work was done on those documents during a short period of time, and drafts were sent for national comment in March-April, 1979.

The International Mathematics Committee benefitted greatly from the proceedings of both seminars, and is very appreciative of the progress made on the manuals, as well.

2.2 IEA General Assembly, Paris, September 17-21, 1979

This important meeting provided a second opportunity for the IMC to interact directly with representatives from each country planning to take part in the Study (the first opportunity was at the Tokyo Assembly in January 1978). Activities during the week-long meeting included:

2.2.1 Review of the cross-sectional and longitudinal aspects of the Study

2.2.2 Consideration of planning for the Curriculum Symposium, January 1980.

2.2.3 Consultation with Dr. Keeves of the International Sampling Committee on a country-by-country basis.

Individual problems concerning sampling were discussed and recommendations for proceeding were offered. The IMC acknowledges with gratitude the tremendous commitment of time and energy offered by Dr. Keeves to this important task.

2.2.4 Critique of all draft final instruments and manuals

One important outcome of this activity was the formation of a subcommittee consisting of representatives of the National Mathematics Committees from France, Luxembourg, the Netherlands and French and Flemish Belgium, for the purpose of recommending the inclusion in the cognitive tests for both Populations of additional items. The items, in the I_s category (important in some countries) would serve to more accurately reflect significant diversity and emphases in the curriculum, particularly with respect to geometry. As a result, it was agreed by the IMC to increment the rotated forms by two items per form with I_s items. Thus, 8 items were added at Population A and 16 items were added at Population B.

2.2.5 Election of chairman of Mathematics Project Council

Mr. Roy W. Phillipps was elected unanimously to a second term as chairman of the Council.

2.3 Meetings of the International Mathematics Committee

2.3.1 Michigan State Meeting: January 27–February 3, 1979

This meeting took place during the depths of winter in the mid-western United States and featured such unplanned events as the "airlifting" of Roy Phillipps and A.I. Weinzweig from a small airfield in Chicago during a record blizzard by one of the University of Illinois' private airplanes. All in all, the meeting was productive, and owes its success largely to the attention paid by Jack Schwille to every detail. The IMC expresses its sincere thanks to him and to Michigan State University for hosting this meeting. Those attending: E. Kifer; S. Hilding; R. Garden; J. Wilson; K. Travers; R. Phillipps. Invited guests included F. van der Blij, University of Utrecht, the Netherlands; J. Schwille; and occasional visits from other scholars in the Michigan State University community. Gerard Pollock was unable to attend due to illness.

It was a distinct pleasure to welcome Dr. van der Blij to the meeting. His probing questions and lively commentary contributed much to the substance and enjoyment of the meeting.

Major accomplishments of the meeting included:

2.3.1.1 Preparation of final forms of the cognitive instruments for the reduced (cross-sectional) study. These were sent out for national comment in March 1979.

2.3.1.2 Preparation of draft background questionnaires for school, teacher and student for the reduced study.

2.3.1.3. Review of recommended final forms of the attitude scales for the reduced study.

2.3.1.4. Review of plans for the curriculum analysis, including the Bielefeld symposium on the curriculum.

2.3.1.5. Development of strategies for coordinating the various aspects of the study (reduced or cross-sectional component, full or longitudinal component with respect to both Populations). These details are elaborated upon in section 9 below. Here, it is stated briefly that it appeared realistic to proceed with Population B, reduced study, on the schedule as announced in Bulletin 3. That is, data collection could proceed as early as May 1980. The anticipated work on the manuals helped to give confidence in that projection. However, Population A was more problematic. Major issues relating to the cognitive instrument and the classroom process questionnaires were yet to be resolved. Funding for instrument development continued to be a severe problem. The IMC agreed that in view of these anticipated difficulties, countries should, where possible, delay data collection until 1981. For those few countries which must proceed with data collection for the reduced study, Population A, in 1980, the Committee would strive to provide interim instruments and manuals by the end of 1979.

2.3.1.?. Consideration of methodological issues raised at the Michigan State University Symposium.

The Committee met for half a day following the symposium to review the deliberations of the symposium and outline the next steps to attempt to resolve the many issues raised. Many of the subsequent sections of this Bulletin reflect the decisions made at this meeting.

2.3.2 Budapest Meeting: September 24-28, 1979

This meeting was hosted by the Orszagos Pedagogiai Intezet, Miklosvari Sandor, Director. Local arrangements were by Zoltan Bathory, the IEA Council Member for Hungary. The IMC is grateful to both gentlemen, and to Dr. Julia Szendrai, Dr. Judith Kadar-Fulop, and the many other individuals at the research institute who did so much to make the meeting productive and enjoyable. Outcomes of this meeting included:

2.3.2.1 Finalizing of the instruments (cross-sectional study)
Version III of the draft final forms (dated July 1979) of the Population A cognitive test was adopted for use in the cross-sectional study, wth the addition of the supplementary items as noted under 2.2.4. For the Population B test, it was noted that for psychometric reasons it is preferable to have items assigned to forms randomly with stratification on content and behavior. The version of the Population B test described in Bulletin 3 (page 29) is now available as an international option for the Longitudinal Study.
The final form of the attitude scales was recommended in a paper by E. Kifer of the IMC and was published separately in December 1979. A short scale on computers and mathematics has been added.
The questionnaires for school, teacher and student were finalized. Robert Garden of New Zealand and David Robitaille, Canada, contributed much to this work. Their assistance is greatly appreciated.

2.3.2.2 Editing of the manuals
Considerable re-working of the manuals was required in the light of developments since the work sessions in Australia in February of 1979. Much of this work at the Budapest meeting was done by Jack Schwille of Michigan State University and Richard Wolfe, Ontario Institute for Studies in Education. Neville Postlethwaite, Chairman of IEA, was also able to attend the meeting, for several days, and assist in the task. To all of these gentlemen, the IMC is greatly indebted for their dedicated and skillful efforts.

2.4 Update: Latin American activities

The Caracas meeting, held in May 1978, and reported in Bulletin 3, page 3, documented the existence of a broad basis of interest in the Study on the part of mathematics educators in Latin America. A follow-up meeting to consider the next steps for participation in the Study was held subsequent to the Fifth InterAmerican Conference on Mathematics Education, held in Brazil in February 1979. Limited travel monies for this meeting were made available by the Organization of American States through the office of Dr. Raul Allard, Director of Education.

Under grant from the Organization of American States, Mr. Peter Staples, a graduate research assistant in mathematics education at the University of Illinois, was enabled to spend one month in Brazil assisting with preparatory work for the Study. Mr. Staples also visited the Dominican Republic, where Dr. Eduardo Luna is actively involved in pilot testing the classroom process instruments.

Chile, with long experience in IEA, is well organzed for participation in the Study. Other countries, notably Brazil, Costa Rica and the Dominican Republic, have made considerable progress toward participation. Presently, mechanisms are being explored for promoting communication within Latin America and providing technical consulting services as needed. However, external funding is likely to be needed for realizing these mechanisms.

2.5 International Funding

Maintenance of the Office of the International Coordinator is provided by the Department of Education, Wellington, New Zealand. Support for meetings of the International Mathematics Committee and for the office of the Chairman of the International Mathematics Committee continues to be provided by the National Institute of Education, U.S.A. During 1978-1979, a large portion of the costs of developing the classroom processes instruments for Population A were borne by an NIE grant to the U.S. National Mathematics Committee.

The Federal Republic of Germany has provided a grant to Dr. Hans Steiner, Institute of Mathematical Didactics, Bielefeld, for the Curriculum Symposium scheduled for January 7-11, 1980.

The Ford Foundation provided a small grant to enable the publication of a brochure describing the Second Mathematics Study in general terms. This grant also allowed the International Coordinator to attend a meeting of the IEA Standing Committee and to visit national centers in Hong Kong, Federal Republic of Germany, Spain, Ireland and the United States.

Two major needs essential for the completion of the Study have yet to be met: (1) support for training of the national research coordinators, (2) support for international data processing in New Zealand and the United States. Dr. Neville Postlethwaite is seeking funding for these aspects of the Study.

3. Projected activities for 1980

3.1 Curriculum Symposium, January 7-11, 1980

This symposium, an essential component of the curriculum analysis component of the Study, will be hosted by the Institute of Mathematical Didactics, University of Bielefeld, Federal Republic of Germany. Dr. Hans Steiner is coordinator of the symposium and in charge of local arrangements. Planning for the meeting has been done by the Curriculum Analysis Group.* Details are given under Section 6, below.

3.2 Research Coordinators Meetings

3.2.1 Cross-sectional study, January 1980

It was hoped that a training session for countries participating in the cross-sectional study could be held in January, 1980, the week following the curriculum symposium. Dr. Steiner kindly agreed to make the local arrangements at the Institute of Mathematical Didactics, Bielefeld. However, at the time of writing, it was not known whether funding would be available to make this meeting possible.

3.2.2 Longitudinal Study, December 1980

An invitation has been extended to IEA by the Institute National de Recherche Pedagogique, Paris, for such a session. The International Mathematics Committee is most appreciative of this continued expression of interest and support on the part of M. Jaquenod, Director, and M. Daniel Robin. It is expected that the support will provide per diem expenses for invited participants. It will be necessary for individual countries to provide their own travel costs to this meeting.

3.3 IEA General Assembly, Finland, August 4-8, 1980

This meeting, which provides the opportunity for IEA members to deliberate upon general concerns of the Association, also serves as an extremely important mechanism for providing interaction between members of the National Mathematics Committees and the International Mathematics Committee.

A key item on the agenda of the Council meeting will be a consideration of results of pilot testing of the classroom process instruments in these countries: France, Japan, Scotland, Thailand, United States,

---

*The Curriculum Analysis Group, based at the University of Illinois, Urbana, consists of James Hirstein, Horacio Porta and Ian Westbury, of the University of Illinois at Urbana-Champaign, augmented at different times by Alan Purves of the University of Illinois at Urbana-Champaign; A.I. Weinzweig of the University of Illinois at Chicago Circle; Hans-Georg Steiner of the University of Bielefeld, FRG; Ulf Lundgren of the Stockholm Institute for Education, Sweden; Ed Jacobsen, Paris, France; and Roy W. Phillipps, Department of Education, New Zealand. Kenneth Travers serves as liaison between the Curriculum Analysis Group and the International Mathematics Committee.

together with India and the Dominican Republic through individual arrangements with the latter two countries.

3.4 Fourth International Congress on Mathematics Education, Berkeley, California, USA, August 10-16, 1980

This important international meeting is held quadrennially. Aspects of the Second Study will be dealt with on the program. One session is planned on the Curriculum Analysis and another is scheduled to report on results of pilot studies of the classroom process instruments in the seven countries referred to in 3.3.

3.5 Other forthcoming international meetings of special interest to mathematics educators.

The First International Congress on the Teaching of Statistics, August 1982; the Sixth Inter-American Conference on Mathematics Education, Latin America, 1983; the Fifth International Congress on Mathematics Education, 1984, in Australia.

3.6 Meetings of the International Mathematics Committee

It is expected that the Committee will meet at least twice in 1980, although funding may not permit more than one meeting in addition to those made possible by other Study activities, such as the Curriculum Symposium in West Germany in January 1980 and the IEA Assembly in Finland in August 1980.

A main consideration of the IMC in 1980 will be the developmental work required in preparation for the longitudinal study for which data collection is scheduled to begin in 1981.

4. Purposes and benefits of the Study

The Second International Mathematics Study is an investigation of the teaching and learning of mathematics in schools. The Study has three main components: a curriculum analysis, which is an examination of the curricular contexts in which teaching and learning take place; a study of classroom processes, intended to provide information on what takes place in the classroom as mathematics is taught; and an analysis of student outcomes (cognitive and affective) in the light of the nature of the curriculum and instructional practice. Two populations are targeted for study:

> Population A: All students in the grade enrolling the modal number of students which have attained the age of 13.0 - 13.11 by the middle of the school year.

> Population B: All students who are in the normally accepted terminal grade of the secondary educational system and who are studying mathematics as a substantial part (approximately five hours per week) of their academic program.

Instruction, even in one subject such as mathematics, is both complex and of crucial importance. Such a system is a highly complex network of many components and relationships. The system is dedicated to the implementation and actualization of the country's educational intentions in the subject matter area. If the subject matter is important to the country and its citizens, and hence if its educational intentions are important to the country, then the instructional delivery system becomes crucially important as the vehicle for realizing those intentions.

Thus the nature of the system, its structure, efficiency and power, become pressing issues. The conceptualization of the issues involved will depend upon the position from which one perceives the system. For the teacher or school official the issues include the adequacy of resources provided, the feasibility of tasks assigned, the nature of the environment made possible, and the concrete criteria of the achievement and growth of children and the quality of educational events. For the researcher the issues are the basis for (and partially the motivation for) the scientific task of authenticating a valid model of such a system, exploring the cause and effect relationships of the model and the effect of variation in key components of it, and, finally, determining the structural properties and dynamics of the class of models exemplified by the given instructional delivery system. For the curriculum specialist and for national (or international) educational decision makers there are policy questions such as the adequacy of outputs of the system in terms of national goals, the maximization of the efficacy of the system, and the isolation of parameters that enhance control and use of the system.

One element common to all of these tasks and issues is the instructional delivery system that is the focus of each. Involved in each of those tasks is the portrayal (model) of that system and its components, relationships and dynamics in at least a reasonable facsimile of their overall complexity. The level of descriptive detail needed in the portrait is established by policy needs, by the authenticity needed for a valid object of research, and by practical needs. The extent to which the portrayal is sufficiently complex to have explanatory power through the relations and dynamics captured is an index of the utility of the model in deciding the policy and research issues raised by the instructional delivery system.

The Second IEA Mathematics Study was conceived in the arena of such issues. It seeks to assist participating countries in their own mathematics instructional delivery systems, whether from a perspective of policy, research or practice, and it also seeks to provide resources that will help those concerned with policy or research to more broadly explore issues involving comparisons of alternative systems.

The First IEA Mathematics Study was conceived in a similar arena and faced a similar task of portraying mathematics instructional delivery systems for a group of countries primarily for policy purposes. That study chose to focus its descriptive portrait on a careful picture of the system's outputs of student achievement and to focus its explanatory efforts on key background and instructional variables.

The Second IEA Mathematics Study, building on the work of that prior study, seeks to enhance the usefulness of results both internationally and to each of the participating countries in terms of their own national concerns. It seeks to enhance the usefulness of results for those who approach instructional delivery systems from a policy perspective as well as for those with concerns for research and practice.

It is desirable and useful to provide a clear and detailed portrait of the outputs of the various national systems, i.e., the actual achievement and attitudes of students. However, it is also important to greatly enhance the portrait of mathematics instructional delivery systems, to move beyond the factors identified by the First Study. Toward this end, the Second Study conceives of an instructional delivery system as involving three key elements, as follows:

1. The System Specialists, who form national educational intentions into a curriculum (embodied variously), train personnel and distribute resources for implementing this intended curriculum, and provide mechanisms for monitoring and guiding that implementation;

2. The Instructional Agents, the teachers who provide the instruction that implements the intended curriculum in the schools and classrooms.

3. The Instructional Recipients, the students who receive the curricular impulse started by the intentions dealt with by the System Specialists and transmitted by the teachers as instructional agents by which the impulse travels through the delivery system from originators to recipients.

This conceptualization of intructional delivery systems leads to three levels of focus for the Second Study: the educational system (generating the Intended Curriculum), schools, classrooms and teachers (generating the Implemented Curriculum), and students (embodying the Attained Curriculum). These three levels of focus lead to the three key components of the Second Study described.

This multi-level conceptualization offers the advantage of providing detailed portraits of two aspects of instructional delivery systems (the Intended and Implemented Curriculums) not portrayed in such detail before. It also offers the advantage of delineating policy and research questions about dynamics and relations between inputs (the Intended Curriculum) and outputs (the Attained Curriculum) into two sets of related questions using the Implemented Curriculum as a critical intervening variable. That is, policy and research questions of interest may be explored by considering first the impact of the Intended Curriculum on the Implemented Curriculum and then the impact of the Implemented Curriculum on the Attained Curriculum. Thus, it seems probable that both the descriptive and explanatory aspects of the portrayal of mathematics instructional delivery systems will be enhanced by the multi-level conceptualization of the Second Study, both between and within countries, and for policy, research and practice perspectives as well.

The investigation will include:

4.1 An analysis of the mathematics curriculum in order to determine the nature of the curriculum today to specify the changes which have taken place in the curriculum in the past twenty years, and to identify those factors which have contributed to its current character.

Benefits of the Curriculum Analysis

For most IEA countries, the past twenty years have been a period of considerable activity in curriculum development. What have been the results of this enormous investment of time, talent and financial resources? For example, has the curriculum become much more heterogeneous, signifying more responsiveness to national goals and needs? Or, in spite of (or, as the results of) the activity and efforts, does school mathematics appear across countries as essentially a monolithic structure?

Detailed information about the curriculum can also be of importance in viewing the findings of the other two phases of the study: the investigation of instructional practice and the analysis of student outcomes. "Curriculum," as Griffiths and Howson have noted (London: Cambridge University Press, 1974, page 156) involves not only statements about goals and content, but includes either explicit or implicit notions of pedagogical method and of evaluation. Hence, a knowledge of the cirriculum of a country should help shed light upon teaching methods utilized to implement the content of the curriculum and should also be of assistance in understanding student outcomes as measured by the international tests and attitude scales.

The curriculum analysis, therefore, serves two purposes which in some senses are independent of each other. It provides much needed information about the curriculum in each country within the context of knowledge about the curriculum across some two dozen countries. The curriculum analysis also serves to help understand and interpret the data to be collected at the classroom level (both teacher and pupil).

4.2 An investigation of classroom processes (Population A)

A series of unique, detailed questionnaires has been devised for the purpose of obtaining information on what teachers do as they teach selected topics in the Population A curriculum. These topics have been chosen on the basis of an international consensus on what is important subject matter for that Population.

Benefits of the classroom process study

Very little detailed information is available on what instructional strategies are employed by teachers as they go about teaching. Yet, since the classroom is the heart of the educational process (at least as education is most commonly practiced today), it is essential that we have more information about what students encounter as they study in the mathematics classroom.

This information is also needed as the pre-service and in-service needs of teachers in a country are assessed. What aspects of teaching practice seem to be most common? What are the desirable aspects of this practice, judged on the basis of current professional wisdom about pedagogy? What aspects of instructional practice need to be improved? For example, do teachers use a variety of instructional techniques, or, instead, do they tend to have an approach which is applied to all topics, and to students of all ability levels? Are there "national profiles" of teaching behavior which can characterize a country?

An attempt will also be made to relate instructional practice to student learning. What categories of instructional practice are related to student achievement, where this achievement is a measure of growth on a particular topic in that teacher's classroom during the school year?

4.3 An analysis of student outcomes

Achievement and attitude instruments have been devised to reflect emphases and concerns in mathematics education for. An international item pool of several hundred items for each population has been developed. Drafts of the instruments were proposed by the International Mathematics Committee, and reactions invited from the National Mathematics Committee in each participating country. The cognitive instruments will provide information at the item and subscore level on mathematical content and behavioral process dimensions of achievement based on an international grid for each population (see Tables 12.1 and 12.2). The attitude instruments will measure these aspects of mathematics-related affect: mathematics in school; mathematics and self; mathematics as a process; mathematics and society; and a scale devised to measure student attitude toward computers.

Benefits of the student outcome component

This aspect of the Study will enable the level of mathematical achievement to be assessed in the light of curricular emphases and instructional practice. What is the level of computational skills in each country? To what extent are students able to solve mathematical problems? What are the attitudes of students about studying mathematics, and about its role in society? How do comparable ability groups (say, the top 5% of the students) compare in problem-solving skills across countries?

In planning for this Study, the IMC has emphasized the importance of utilizing information from the research instruments at the item level. Two reports, by Peaker and by Postlethwaite, illustrate the sort of analyses which are expected to be useful.

Peaker (International Review of Education, Volume XV, 1969, pages 222-228) examined the findings of the First Mathematics Study from the perspective of gleaning those results of greatest interest to classroom teachers of mathematics. His analyses of item level information are particularly rich in implications for instruction. For example, he examines country performance on an integration item for Population 3a (approximately our Population B), (Test 9, Item 4) and concludes that the striking differences in performance on this item (.04 in the Netherlands to .68 in England) cannot be ascribed to the retentivity of the system (which is 5% for both countries) but should, rather, be attributed to differences in curriculum and instruction. An analysis of performance at the subscore level produces similar conclusions. While two countries may have total scores which are nearly identical, the contributions to this score come from different sources. Again, in Population 3a, Belgium and England had total scores of 65 and 66 respectively. But, observes Peaker, "Belgium is strong on new mathematics, algebra, analysis and sets. England is strong on geometry, analytical geometry and particularly on calculus." (page 225).

Postlethwaite ["Item scores as feedback to curriculum planners: A simple case from the Swedish Comprehensive School and a more general model." Scandinavian Journal of Educational Research (15:3, pp. 123-136)] has suggested that item level analyses in conjunction with an examination of the textbooks and syllabi may yield important information for helping explain variation in achievement at the educational system level.

Ralph W. Tyler, in a recent paper entitled "American Education in the perspective of education in other nations", reviewed the state of education in the U.S. on the basis of the previous IEA studies. Many of his conclusions may be of interest to those in all countries who are concerned with the education of their youth. He viewed the IEA data from the point of view of "identifying factors that explain the variance in educational achievement both within and among industrialized nations." He goes on to conclude, "For the top five or ten percent of young people in all these nations, the two main factors are opportunity to learn the subject and the emphasis given the subject by those adults who are respected by these young people. For the majority of the children the factors identified are the opportunity to learn the subject, the education of the parents, the public attitude toward the subject and the time devoted to learning it."

4.4 An examination of changes in mathematics education since the early 1960s.

The past two decades have been marked by considerable interest and activity in the curriculum, particularly in mathematics and the sciences. In many countries, the 1960s especially were characterized as a period of curricular reform, and even revolution. Two events approximately twenty years ago now serve as important benchmarks in mathematics education. In 1958, the Organization for European Economic Cooperation conducted a survey of mathematics education in member countries which culminated in the Royaumont Seminar and the report, New Thinking in School Mathematics (OEEC, 1961). This survey provided information on practices and trends in school mathematics and rather detailed information on the mathematical content of the curriculum. The second event, the First IEA Study of Mathematics, 1964, provided empirical data on characteristics of schools, teachers and students as well as achievement and attitude data for students in twelve countries. Of those twelve, eleven are currently planning for participation in the Second Study.

With information from these two surveys, supplemented by information currently being requested of the some twenty-four countries which have expressed interest in the Second Study, it will be possible to chronicle many of the changes which have occurred in mathematics education.

Typical questions which could be addressed by an examination of changes in mathematics education include: What lasting changes in the curriculum have taken place? What factors have influenced these changes? How does curricular change take place in different countries? With respect to student outcomes, questions relating to current concern for "declining standards" can be addressed. In what aspects of mathematics achievement have there been declines and in what aspects have there been gains over the past twenty years? What evidence is there of the impact of technology, in particular computers and calculators, on classroom practice and student achievement and attitudes? Are there strikingly different patterns of student gains (losses) across countries? If so, what other factors can help account for these changes?

General questions relating to student achievement as suggested by Tyler (op cit) will be responded to in the study. "Wide publicity needs to be given to the negative effects of the erosion of the out-of-school learning environment. The average child now has fewer hours with parents, and fewer hours in youth-serving institutions than in the past. The average child from 10 to 14 years of age spends 1,500 hours per year watching television and only 1,100 hours per year in school. Hence, rebuilding the eroded out-of-school learning environment, is likely to bring greater results than any other single strategy."

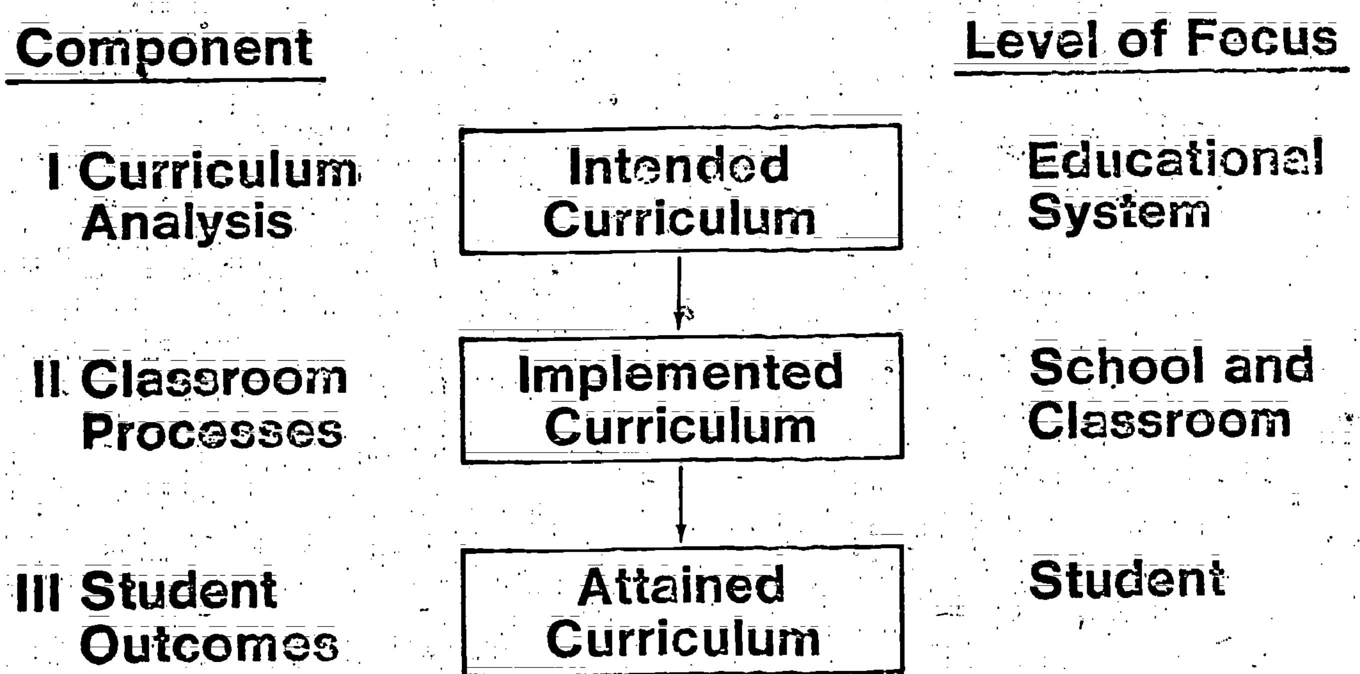
5. Model for the Study

Figure 1 illustrates a conceptualization of the Study. Three levels are objects of investigation, one corresponding to each of the three components of the study. The curriculum analysis focuses upon the educational system as an entity and is intended to portray the context in which mathematics education takes place (school organization, selectivity and educational goals, for example) as well as to review the status of mathematics education in that system. The investigation of instructional practice focuses upon the classroom level. Here is determined both the extent to which the "intended curriculum," as formulated at the system level, is actually implemented in the classroom, and the variety of instructional methods employed as the implementation takes place. The third focus is upon student attainment of curricular and instructional goals. What is the nature and extent of mathematical achievement and attitudes of the students in the two target populations of the Study?

An Expanded Model of the Study

Each of the components of the Study may be viewed in more detail with the aid of the expanded model (originally proposed by McKnight), as shown in Figure 2. The figure is useful in highlighting certain of the interrelationships which are to be examined.

FIGURE 1. CONCEPTUALIZATION OF THE STUDY



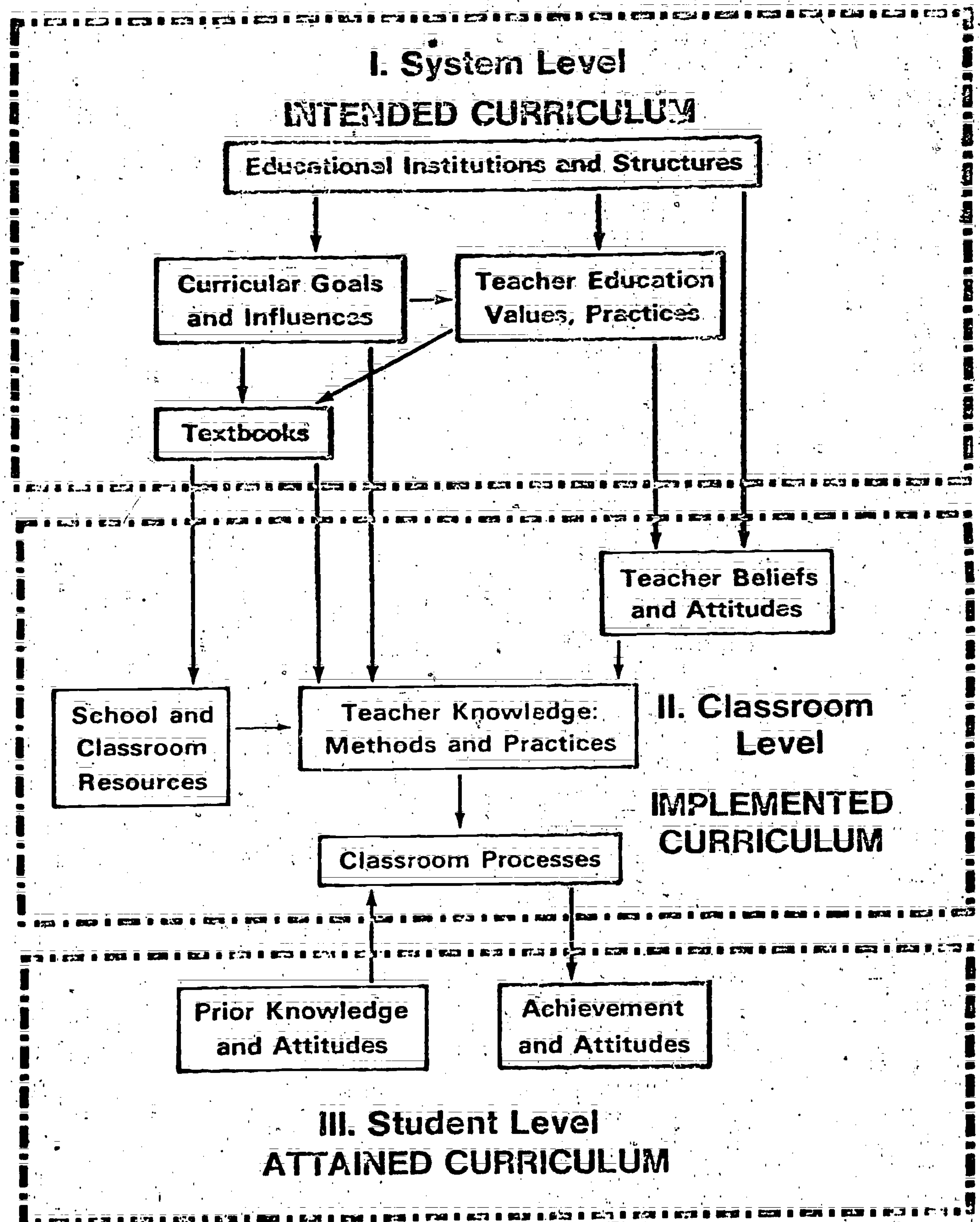


FIGURE 2. EXPANDED MODEL OF STUDY

6. Component I: Curriculum Analysis (The Intended Curriculum)

The curriculum analysis is to be conducted in order to provide a backdrop for viewing and interpreting the findings of the classroom processes and student outcomes components of the Study. The curriculum analysis will undertake two probes into the state of mathematics curricula in the participating countries and organizational contexts surrounding these curricula. These two probes, discussed below, are referred to as a Contexts Survey and a Status Survey.

6.1 The Contexts Survey

It is the goal of this component of the curriculum analysis to provide a framework within which the findings of the status study can be viewed. Some of the data on the contexts for school mathematics is already available in such investigations as the Royaumont Study and the First IEA Mathematics Study. These data will be updated where possible, and fresh data will also be sought. Explicit probes will be made in the following areas:

- 6.1.1. Societal contexts for schooling with particular reference to the occupational structure and national demography.
- 6.1.2. Institutional contexts of school mathematics:
  - 6.1.2.1 Articulation between secondary school and the subsequent careers of students.
  - 6.1.2.2. enrollment, tracking and school organization and overall curriculum with the implications for course enrollments and perceptions of mathematics.
- 6.1.3. The curriculum development system. Its character and form and its participants at the formal level, and its articulation with a curriculum research and development system, if any.
- 6.1.4. The character and form of the control systems and coordination surrounding mathematics, i.e.,
  - 6.1.4.1. examination and grading systems
  - 6.1.4.2. inspection systems
  - 6.1.4.3. text and materials development systems

Particular attention will be given to the jurisdictional settings of these systems and their articulation with the formal structures of school organization.

- 6.1.5 The legal and certification structure of the human resource system: supply and demand of teachers.

A questionnaire has been devised by the Curriculum Analysis Group in order to formulate a description of the societal and organizational contexts of mathematics education with a focus on occupational structure and national demography, articulation between secondary school and subsequent careers of students, school organization, curriculum development systems, control and coordinating systems (examinations, inspectorates, etc.) and the legal and credentialing structures of the system.

6.2 The Status Survey

In 1959, the Organization for European Economic Cooperation (OEEC) held a seminar in Royaumont, France, for the purpose of reviewing the current status of the school mathematics curriculum for the 21 member countries of OEEC and to consider the "production of a sound mathematics programme in harmony with modern thinking in mathematics." (New Thinking in School Mathematics. OEEC, 1961, page 7.)

The Curriculum Analysis Group has prepared a questionnaire based on the Royaumont survey instrument, but modified in order to take account of mathematics topics which have since been "legitimized" in the curriculum and new topics, such as computer or information science which are candidates for "legitimacy", to varying degrees, in the curriculums of the IEA countries.

There is significant overlap between countries participating in the current Second IEA Mathematics Study and those participating in the OEEC Royaumont Study (where possible, a picture of the curriculum as it was in 1960 will be collected from those IEA countries which were not members of the OEEC at the time of the survey. The countries participating in the Royaumont Study were: Austria, Belgium, Canada, Denmark, France, Germany, Greece, Ireland, Italy, Luxembourg, The Netherlands, Norway, Sweden, Switzerland, Turkey, United Kingdom, United States, Yugoslavia.) The expectation is that these two census-like surveys will provide, for a significant number of countries, useful and important data on both stability and change in the mathematics curricula in the participating countries. There have been major changes in the structures of many school systems since 1960 and it will be of considerable importance to attempt to assess the significance of these changes for the character of the mathematics taught in the schools.

6.3 National Case Studies

In 1977 an advisory committee of UNESCO, sponsored by the International Commission on Mathematics Instruction (ICMI), issued a call to various countries for reports on changes in mathematics education which have occured since the late 1950s. The call included the following outline, published in Educational Studies in Mathematics, to be used as a guide to the authors of the national reports as they described the various developments in their countries:

Changes in:
- subject matter
- teaching method
- attitude toward mathematics
- relations between mathematics and other subjects.

Change as a process:
- The leading ideas --
  - how did they develop in the course of the years?

how did they manifest themselves in plans, projects, proposals?
how were they actually realized?

The main forces --
general educational, social, cultural developments;
increasing mathematical literacy;
increasing mathematical excellence;
adult education;
applicability of mathematics.

The designers --
commissions, committees, curriculum developers, textbook writers, test designers.

The performers --
administrations, schools, teachers, students, parents.

The tools --
training, retraining, textbooks, syllabi, teacher's manuals.

Cooperation and resistance --
by administrations, teachers, parents, public.

Lessons learned in the past to the benefit of the future.

Changes on different levels:
school, teacher training and retraining, and guidance in their inter-relatedness.

Reactions to change:
in other areas;
by higher learning;
by parents;
by the public.

Reports were received from the following countries (authors of the reports are given in parentheses) and published in Educational Studies in Mathematics, Volume 9, Numbers 2 and 3. Countries planning to participate in the Second IEA Mathematics Study are marked by *.

Australia* (A.L. Blakers)
Bangladesh (S.M. Sharfuddin)
France* (A. Revuz)
Great Britain* (A.G. Howson)
Hungary* (Maria Halmos & Tamas Varga)
India (J.N. Kapur)
Iran (Badiollah Rostami)
Nigeria* (R.O. Ohuche)
Poland (A. Ehrenfeucht)
Sierra Leone (A. Williams)
Sri Lanka* (A.J. Gunawardena)
Sudan (M. El Sawi)
Thailand* (O. Purakam)
United States* (J.T. Fey)
West Indies (B.J. Wilson)
The Netherlands* (H. Freudenthal)

Subsequently, reports from an additional half dozen countries were received and are scheduled to be published by UNESCO in Summer 1979.

These reports will be examined by a committee of three persons (Frederik van der Blij, the Netherlands; Sven Hilding, Sweden; and A.I. Weinzweig, the United States) which will prepare a summary paper synthesizing what is reported by the countries. Professor van der Blij is serving as chairman of this committee. Robert Garden also assisted greatly in the work of the committee.

6.4 Case Studies in Teaching Mathematics

The final component of the curriculum analysis will deal with three mathematical topics and how these topics are articulated in the curriculum and classroom. For illustrative purposes, three countries will be asked to prepare national case studies on the teaching of one of the three topics: geometry, introductory algebra, and probability and statistics.

6.5 Publications of the Curriculum Analysis

The following publications, scheduled to appear in mid-1980, are expected to emerge from the symposium, and hence be the first reports of the findings of the Second International Mathematics Study.

6.6.1 Report of the proceedings of the Symposium

6.6.2 National reports on the status of mathematics education and the synthesis report

6.6.3 Implications of the curriculum analysis for interpreting the findings of the classroom process and student outcomes phases of the Study

7. Component II: Classroom Processes (The Implemented Curriculum)

Previous IEA studies have sought to provide detailed portraits of student achievement that represented the curricula of participating countries as attained by the students in the targeted populations. These detailed portraits of national student achievement, set in an international context of portraits of student attainment in other countries, had a variety of policy implications for the participating countries. The potential policy implications were enhanced by studies of a number of broad background variables concerning the students involved and a few key global characteristics of the instructional delivery systems of participating countries.

7.1 Description and Explanation

The goal of the Classroom Processes component of the Study is to provide detailed descriptions of national curricula as implemented in the schools and classrooms of the IEA countries. The practical complexities of school and classroom life lead to a transformation of any intended curriculum (such as those portrayed in the Curriculum Analysis component of the Study). It is possible to judge something of the extent and nature of this transformation by examining those processes and activities characteristic of instruction in the targeted classrooms, and by examining the methods and practices commonly used by the teachers providing instruction to the target classes.

7.1.1 Detailed Descriptions of Practice. Providing a portrait of classrooms involves, at the most fundamental level, a detailed description of typical practice in the classrooms in which instruction is delivered to the target population. Such description is prior to any study of relationships among aspects of that classroom practice. However, even descriptive studies at this level should begin to provide some information relevant to the "why's" of student outcomes. Without the information provided by such studies much is missing of the context in which student outcomes must be explained.

7.1.2 Relational Descriptions. To enhance the explanatory potential (and policy implications) of the classroom processes descriptions, it is necessary to go beyond the simple portrayal of discrete classroom practice elements to a portrayal of the relationships between these elements. An important aspect of the data analyses made possible by the Classroom Processes component of the Second Study is the examination of key interrelationships between the classroom practice elements captured by the instruments used. What should emerge is a description both of discrete elements of classroom practices and of interrelationships in classroom practices that make more global characterizations of those practices possible.

7.1.3. Explanatory Analyses. The curriculum as implemented in actual schools and classrooms is not only important as context for interpreting student outcomes and seeking reasons for them, but is also important as a source of potential explanatory variables. Classroom effects and teacher practices are essential intervening variables in describing any national instructional delivery system. In most cases, it is not realistic nor should it be particularly effective to use distant background variables and variables related to the official or intended curriculum of a country in order to explain the pattern of student achievement and attitude outcomes. The intended curriculum and the national system for directing the implementation of that curriculum must first have an impact on the schools and classrooms which are the site of instructional delivery. In return, this delivered instruction has an impact on actual student achievements and attitudes. Thus, classroom effects and teacher practice must be a central link in any detailed explanatory chain seeking to characterize a national instructional delivery system.

This status as source of intervening variables means that the data from the Classroom Processes component will enter into two other important kinds of analyses. First, classroom and teacher practices characterizing the implemented curriculum may be related to the intended curriculum and the national system for guiding implementation (both portrayed in the Curriculum Analysis component of the Study) to assess the extent and nature of transformation of the intended curriculum and to assess how such transformation is related to various kinds and aspects of national systems for guiding implementation.

Furthermore, aspects of the implemented curriculum may be related to the pattern of student achievements and attitudes. The design of the Study seeks to make such an analysis possible both within several specific mathematical topics and across those topics. It seeks also to analyze several different levels of classroom practice variables ranging from the very specific to the more general. This enhances the likelihood that this type of analysis, at some level of specificity, may reveal important relationships that might otherwise be hidden by some more global variable (such as "opportunity to learn") in a less detailed analysis. There is thus the potential for discovery of patterns of instructional practice that have demonstrated effectiveness in terms of student achievement and attitude outcomes.

7.1.4. Summary. The Classroom Processes component of the Second Study will potentially be able to provide a portrait of that aspect of instructional delivery that includes the following descriptions and analyses:

i) Descriptions of discrete classroom practice elements
ii) Descriptions of relations between classroom practice elements (including more global classroom practice characterizations)
iii) Analyses of relationships between intended and implemented curricula (and effect of various practices in guiding implementation)
iv) Analyses of relationships between instructional practice aspects of the implemented curriculum and the pattern of student achievement and attitude outcomes

7.2 Categories of the Implemented Curriculum

Section 5 of this Bulletin offered a model for this Study and Figure 5.2, is an expanded version of that model. The central "block" of that figure represents an analysis of the Implemented Curriculum in terms of four major categories, and seeks to put these four categories into a network of categories characterizing the other components of the Study.

7.2.1. Teacher Methods and Practices. Central to the implemented curriculum, to instruction as actually delivered in the classroom, is the category labeled "teacher knowledge: methods and practices." Individual teachers often act, when faced with similar instructional situations, in stereotypical ways that constitute a pattern or set of patterns characteristic of their teaching practice. In some cases for a teacher, these patterns are a reflection of consciously held and deliberately executed methods. In other cases, these patterns have evolved out of the teacher's experience and past practice. The result is, for any given mathematical topic and instructional situation, an array or repertoire of characteristic approaches to instruction in the various aspects of this topic. These practice patterns are quite complex but, in many cases, are relatively stable and available to the teacher for consideration (at least if probed by very specific questions in which demands for inferences and demands on memory are minimized).

A model has been constructed and is being refined which relates some of the major aspects of such teacher methods and practices, and which also integrates them with the other aspects of the implemented curriculum. A series of questionnaires for teachers has also been developed which probe (by means of very specific low-inference, low-memory questions) critical aspects of these kinds of practices in several topic areas. These questionnaires, along with the model, should provide both useful descriptions of discrete practice elements (as discussed earlier) and descriptions of relations between practice elements leading to more global characterizations of teacher practice (guided by the model).

7.2.2. Other Aspects. While "teacher methods and practices" are the central component of the model of the implemented curriculum they are not the only component. There are several categories that impact on teacher methods and practices, both to constrain and influence the selection of methods and also to affect the execution or performance of a given typical method or practice. One such major category is that of "school and classroom resources." The array of resources available in the school or a classroom within the school constrains the instructional choices of the teacher providing instruction in that setting. Resources act as both props and cues in instruction. Certain instructional activities can be carried out only if certain resources are available as props to be used in the activities. Further, the actual physical presence of certain resources may serve to cue or influence a teacher's decision to use certain activities.

A second major category is that of "teacher beliefs and attitudes." This category includes a variety of factors internal to the teacher that influence his or her interpretation of instructional situations and responses to them (in terms of methods selected and executed). These factors include such things as teacher beliefs or schemas of typical student behaviors, instructional situations, and teacher actions. They also include the teacher's internal representation of the mathematical subject matter; and basic attitudes toward mathematics, towards teaching, and towards specific classes and students. Teacher values relating to various kinds of activities, student responses, are also included. While this category is an incredibly complex one, certain aspects with demonstrated connections to practical instructional outcomes have been selected for probing through the questionnaires.

Two other major categories that directly affect the implemented curriculum and, in particular, teacher method and practices, are actual elements of the intended curriculum. This includes, first, the textbook or textbooks chosen and available to the teacher. This is one of the dominant influences (if not the dominant influence) on teacher choice of instructional activities and strategies. Secondly, there are various embodiments of curriculum goals and influences such as national or school syllabi, external examinations, etc., which affect the activities and strategies chosen not so much in the sense of by what they make possible (as a textbook does) but in the sense of by what they make desirable (which a textbook also does). Selected aspects related to both of these categories are also probed by the questionnaires.

Finally, there are the actual activities of the classroom. The previous set of four categories affects and constrains the selection and execution of teacher methods and practices, and the methods and practices so selected are realized in a set of instructional activities in (or related to) the classroom which constitutes the actual instruction for the target class.

7.3 Description of the Instruments

7.3.1 Instruments Involved. Instruments or elements of instruments related to the Classroom Processes component of the Study include the following:

1) Teacher Background Questionnaire.
   This seeks information on sex, age, years of experience, etc.
2) Teacher Attitude Scales
   Certain of the attitude scales [described elsewhere] will be taken by teachers as well as students.
3) Teacher Opportunity-to-Learn Items
   This construct was devised for use in the First IEA Mathematics Study as an index of the extent to which the intended curriculum was implemented by the teacher in the classroom. The measure has subsequently been refined and is currently being piloted for use as one aspect of the Classroom Process component of the Second Study. It is planned that both teachers and students will be asked questions concerning the extent to which opportunity has been provided in class to learn the various topics reflected in the items from the cognitive tests.
4) Classroom Process General Questionnaire
   Certain aspects of teacher and instructional practice are typically quite uniform across specific subject matter topics. A general questionnaire will be used, seeking information on

these more general practice elements (e.g., grouping practices, uses of instructional materials). Items from this questionnaire serve three purposes:

(1) They lead to variables powerful in their own right;
(2) They provide a context for interpreting more topic-specific information; and
(3) They make contact with variables used in other studies of general teaching behavior so that results of this study may be related to other studies.

5) Six Topic-Specific Classroom Process Questionnaires

Certain other aspects of teacher and instructional practice may or may not generalize across topics, but, even if they do, may still be best assessed through low inference and memory demanding questions about specific aspects of instruction related to specific topics.

A topic-specific questionnaire about instructional practice has been developed for each of six topics which are reported as important or very important on a consensus basis across countries. Thus, at least some of these instruments should be related to the year's work of any target class of any participating country. It is intended that each instrument will be completed soon after the majority of the year's instruction in the relevant topic.

The six topics are as follows:

(1) Ratio, Proportion and Percent
(2) Common and Decimal Fractions
(3) Geometry
(4) Measurement
(5) Formulas and Equations
(6) Integers

This array of instrumentation should be sufficient to generate a detailed portrait of the intended curriculum as implemented by the teachers involved in the Study.

7.3.2 Content of the Instruments. The six topic-specific questionnaires, which constitute the main data-gathering instruments of the Classroom Processes component of the Study, are designed to sample information from a rich conceptual domain (described in Section 7.2). The number and complexity of items scattered over the six questionnaires might seem, at first, quite daunting. Actually, however, all of the items can be grouped into 13 item types by content and the items of any given type (1) follow, for the most part, the same item format with only minor variations; and (2) cluster together in terms of the variables they represent and the higher-order variables to which they contribute.

The 13 item types fall into the following four major categories:

1) Teaching Methods (8 types)
2) Topic Profile (3 types)
3) Teacher Opinions (1 type)
4) Specific Area Questions (1 type)

Two of the categories are quite simply described. First, at the end of each questionnaire is embedded an "opinionnaire" that solicits a variety of teacher opinions on aspects of the instruction related to the topic involved. Much of this information is purely descriptive but some is used in characterizing teacher beliefs and values. Second, the "specific area questions" constitute a small set of items (7 in all) which seek purely descriptive information related to instruction in one or the other of the six topics, but which has no counterpart in the other topics (e.g., kind of geometry used).

The "topic profile" category involves three item types standardized across the six topics to provide a quick, global profile of instruction in that topic. The three item types deal with (1) allocated time, (2) sub-topics covered, and (3) program emphases.

The most complex, and certainly the most interesting, category of item types is that of "teaching methods" which involves 8 item types. This category has five major sub-categories. First, one sub-category relates to content dealing with "concepts and relationships." In this sub-category are two item types, one dealing with kind of representation used in the concept-oriented instruction and another concerning the kind of teaching techniques used in introducing and dealing with such conceptual representations.

A second major sub-category concerns "formulas, procedures and propositions." It also involves two item types, with one sampling the procedures taught and a second the techniques used in teaching those procedures.

A third sub-category deals with mathematical applications. Again, two item types are involved, one to sample the applications used and another to sample the sources from which these applications are drawn.

A fourth sub-category involves only one item type which samples the reasons for use and non-use of certain concept representations, procedures, applications and techniques from the above sub-categories. The fifth and final sub-category also involves one item type which relates to the number and kind of instructional aids or resources utilized.

This scheme of profiling topics and looking at teaching methods in terms of three kinds of content (concepts, procedures and applications), all of which can be realized through use of only about a dozen item types, seems quite simple. It is, in fact, this conceptual simplicity of design which allows the generation of a reasonable but small number of higher-order variables to characterize the rich variety that can be sampled by the many specific items of the few item types.

7.4 Classroom Process Variables: An Overview

7.4.1 Levels of Classroom Process Variables

The classroom process questionnaire will yield a wide variety of variables at several levels. At least five levels of variables can be identified as follows:

1) item level
2) basic item-type variables (most item variables may be aggregated with corresponding variables for other items of the same item type to yield a variable characteristic of some aspect of that item type)
3) homogeneous higher-order variables (further aggregates using conceptually-based clusterings of similar item types and variables)
4) heterogeneous higher-order variables (further aggregates using model-based clusterings of more heterogeneous item types and variables which have been related through the modeling of teacher methods and practices)
5) statistically-determined higher-order variables (aggregates of variables based on statistical clustering techniques rather than any a priori conceptual approach)

Item variables involve no clustering and hence no inference in their definition. Basic item-type variables involve minimal inference reflecting only the lowest level of clustering, i.e., that of identifying item types and obvious content similarities. These two stages yield a large number of variables since there are a large number of items and since this low level of clustering yields over 50 basic item type variables (not considering the fact that most yield indices for one or more of the specific topics as well as an overall across-topic score and also not considering that, in many cases, a number of different weighting schemes are available in aggregating by using frequency and/or emphasis information gathered by the instruments).

The next stages require some decisions. Two major approaches are available in seeking further, higher inference clustering. One approach simply uses statistical techniques (e.g., factor analysis) to identify, weight and aggregate variable clusters. Conceptual interpretation of aggregates created in this way takes place largely after clustering. A second approach is to use a priori some conceptual scheme to identify clusters before any empirical work and to base aggregation on such already conceptually meaningful clusters. A conservative versi n of this conceptual clustering involves still relatively low inference clusters of variables involving similar content. This is what was referred to above as (relatively) homogeneous higher-order variables. A more venturesome approach is to use a more elaborate conceptual scheme (e.g., a model that would generate profiles among sets of variables, with the profile becoming a new higher-order variable) to generate what were called above (relatively more) heterogeneous higher-order variables.

In the present situation of relatively limited information, it does not seem reasonable to choose the statistical "rather than the conceptual approach to higher-order clustering or to choose the more conservative rather than the more venturesome in conceptual clustering. Thus, the earlier list reflects the fact that all three strategies of generating useful higher-order variables will be followed at least through the stage of large scale piloting. It is hoped that a convergence between statistically generated and conceptually generated clusters will emerge during analysis of such data.

It is, of course, not possible to describe at this time clusters that may be statistically generated later. Work on model building and the more venturesome approach to conceptual clustering is ongoing at present but is somewhat tentative. It is appropriate, however, to go into more detail on the more conservative conceptual approach of generating homogeneous higher-order variables.

7.4.2 Homogeneous Higher-order Variables

The present set of basic item-type variables from the six topic-specific classroom process instruments reveal a number of clusters based on narrowly conceived, low inference similarities which are suitable for generating the next higher level of aggregated variables, i.e., those which were earlier called "homogeneous higher-order variables." Three clusters of this type are discussed here.

Representation Type. A major issue in education for some time (especially since elucidated by Bruner) has been that of which type of representation of subject matter is appropriate and differentially helpful in which type of instructional situation. Bruner used a typology involving three major types of content representation: the concrete, the iconic, and the abstract. This typology has been extensively discussed in the literature related to mathematics education and to cognitive growth.

A typology of instructional situations to be related to this typology of representation type is more difficult to identify from the literature. However, a broad typology is embedded in the Study's classroom process instruments. Instructional situations may be classified into three categories: concept-oriented instruction, rule- or procedure-oriented instruction, and applications-oriented instruction.

A crossing of these two typologies yields nine sub-categories of variables, one for each representation, instructional situation pair. The present instruments yield indices for each of these nine subcategories overall, i.e., across (or pooling) the six topic-specific instruments, and, in most cases, most of the six individual topics yield indices for the sub-categories for those topics separately. There is thus a network of variables able to portray in great detail representation type and its interaction with instructional situation type both within specific topics and across topics. The possibility of relating these variables to student achievement scores within specific topics and across topics suggests a rich potential for exploring the effect of representation type on student achievement.

It should be noted that the crossing or representation and instructional situation types to produce nine sub-categories that form a network of variables does not exhaust either the variables or indices that can be related to that crossing and that network. Something like "row sums" and "column sums" aggregates can be formed to get overall indices for a given representation type across instructional types, etc. In addition, the network of related indices also makes possible the development of variables related to profiles based on some set of subcategories from the nine. The portrait that can be used to study the effects of representation types is thus rich indeed.

Diversity. The issue addressed above was that of the effect of specific types or blends of subject matter representation in various instructional situations. A related, but separable, issue is that of the effect of diversity in instruction. Is it better, for example, to pick one central representation of the subject matter for large portions of the instruction or is it better to expose the student to a variety of representations (e.g., in teaching fractions, representing fractions through sets of concrete objects, through diagrams of shaded regions, through numerical ratios, etc.)? Does the value of such diversity depend on the type of instructional situation (e.g., diversity may be facilitating in concept development but inhibiting in learning computational techniques where selecting one algorithm for extensive drill may be the best approach)? The question of diversity (or as Dienes called it "mathematical variety") is clearly both an interesting question and one of practical importance.

Indices of diversity for various aspects of instruction essentially are based on measuring how many out of a set of alternatives related to that instructional aspect were typically used by a teacher in teaching the target class. In some cases (e.g., representation of subject matter as discussed earlier) a typology exists which partitions or classifies the members of that set of alternatives into a set of categories or types. In that case, diversity indices could either be based on the number of types utilized or on the actual number of alternatives utilized.

To the extent that diversity in instruction is facilitative or inhibitive, the simple, a-typological index should be most powerful. If, however, the typology captures some aspect of the instruction relevant to the effect of diversity, the typologically based index may be more powerful. For example, instruction which uses three representations of a concept passing from concrete to abstract to iconic may relate much more significantly to achievement than does instruction that simply uses three rather than one subject matter representation but all of the same type, e.g., abstract.

It is not possible a priori to specify whether the typological or the a-typological indices will be most powerful in any of the aspects of instruction which yield diversity indices and which also have available some potentially relevant typology. The approach here is to generate two separate types of diversity index in each situation: (1) variety, which is always a typologically based index, and (2) focus, which is always a simple, a-typological index.

The present analysis of variables has led to seven lowest-level variety indices, three related to representations used (in each of the three instructional situation types) and four to other aspects of instruction such as number of different factors indicated as utilized in making specified instruction choices (i.e., diversity in method selection or choice approach). There are, of course, seven lowest-level focus indices corresponding to these seven variety indices but there are also three other focus indices that have been identified in situations for shich no relevant typology seems available, bringing the total to ten lowest-level focus indices.

These indices have been called "lowest level" above because it is again possible to aggregate clusters of them into higher-level indices, although, in this case, such a cluster would involve more heterogeneity and would have to be either statistic--or model-based. An example of such a cluster would be combining focus indices for such things as diversity in factors considered in instructional decisions, in subject-matter sub-topics covered, in program emphases, etc., to get an overall diversity index related to the whole of a teacher's approach to instruction in a given topic. Actual specification of such clusters must wait either for data or further model development. For now it is enough to say that the array of seventeen types of diversity indices (many available for specific topics as well as across topics) offers the possibility for a rich portrait of the place of diversity in instruction and the opportunity to explore some quite detailed hypotheses about diversity effects.

Differentiation. A central issue in education generally, as well as mathematics education specifically, has been that of individualization or the adaptation of instructional approach to adequately respond to individual differences in the students taught. Issues relating to individualization have included various aspects of tailoring instruction, grouping practices, assigning different tasks to various ability groups. While information about attitudes and practices related to the first two of the preceding aspects is sought through a number of items on the general teacher classroom process questionnaire, the last area, i.e., task differentiation, is explored in more detail on the six topic-specific questionnaires.

As was stated earlier, the items of the six topic-specific instruments seek specific, detailed pieces of information that make possible a "descriptive mosaic" covering a variety of aspects related to a variety of specific instructional situations and instructional decision-making. Much of this mosaic relates to the subject matter (concepts, procedures, applications), representations, techniques and instructional strategies typically used in presenting various topics. Almost all of these aspects of instruction are potential candidates for differentiation decisions in order to respond to below average students, above average students or both.

In a large proportion of the items providing the information just described, additional information is sought about differentiation related to the thrust of each item, including the group (above average, below average, or both) towards which differentiation is directed and often some measure of the frequency of, or emphasis on, some particular difference.

This aspect of items makes possible differentiation indices for a large number of items, and it also makes possible more aggregated differentiation indices related to various aspects of instruction (e.g., type of representation used, subject-matter sub-topics included, etc.) both for specific topics and across topics. These indices are available in two forms: (1) differentiation indices, which use simple, unweighted aggregations of presence or absence of differentiation, and (2) intensity indices, which are weighted aggregates of simple differentiation indicators, with weightings determined by frequency and emphasis information.

Currently, in addition to item level and sub-category indices, there are seven higher-order differentiation indices and six intensity indices (available across topics and for many specific topics) related to most of the same aspects of instruction as were the diverstiy indices. This makes possible a detailed portrait of teacher differentiation practices to be related to the information gathered (through the general questionnaire) about the other aspects of individualization and to be related to student achievement and attitude. This also makes possible, as was discussed briefly earlier, the combination of differentiation information with diversity information and other types to provide more global characterizations of a teacher's practice through model-based, heterogeneous aggregations of variables. The characterizations of teacher practice that would emerge, with varying levels of aggregation, varying topics, relating to various aspects of instruction, would seem potentially to be both descriptively and explanatorily rich.

7.4.3 An Example

Perhaps additional insight can be provided through illustrations of how some typical items will be used. Shown on the next pages are some items from the Common and Decimal Fractions Questionnaire. The fate of these items in terms of the variables just described will be examined.

Essentially these items involve one central item securing key information surrounded by a cluster of other items seeking additional information. The central item examines teacher use of an array of eleven "interpretations" (representations) of fractions. This array or set of alternatives can be classified according to representation type (e.g., a is abstract, c is iconic, j is concrete). Although set of alternatives in this item is not well balanced in terms of representation type, the balance is much better when this item is aggregated with others of the same item type from the same questionnaire. (There are currently two other such items.) The central item will then contribute to indices related to representation type for conceptual instruction for the fractions topic (and through it to across-topic indices). The additional information sought on frequency and/or emphasis enhances the description and might be used for weighting the aggregate indices toward greater power.

Since the central item secures information relative to a set of alternatives on some instructional aspect (concept representations), it also will make a contribution to various diversity indices. Since a typology (of representations) is available, the item may contribute both to variety indices and focus indices. Specifically, it would aggregate with other items of the same type and topic to provide an index for the CV-F (Concept Variety, Fractions) and CF-F (Concept Focus, Fractions) variables and, through them, would contribute to other higher-order variables.

The cluster of related items seeks information about how this instance of concept representation is varied by the target class teacher to enhance the achievement of above average and/or below average stu-

COMMON FRACTIONS

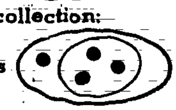
1. Various interpretations of fractions are depicted on the left below. For each interpretation (a–k), place a check in the box by the response that best describes your use of that interpretation.

a. Fractions as quotients:

$\frac{3}{4}$ means "3 divided by 4"

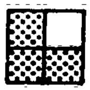
I do not use ☐	I use occasionally ☐	I use frequently ☐

b. Fractional parts of a collection:

$\frac{3}{4}$ means 

I do not use ☐	I use occasionally ☐	I use frequently ☐

c. Fractions as regions:

$\frac{3}{4}$ means 

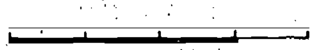
I do not use ☐	I use occasionally ☐	I use frequently ☐

d. Fractions as ratios:

$\frac{3}{4}$ means 

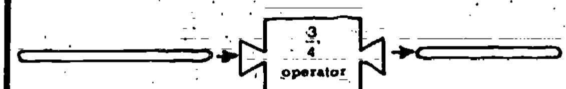
I do not use ☐	I use occasionally ☐	I use frequently ☐

e. Fractions as segments:

$\frac{3}{4}$ means 

I do not use ☐	I use occasionally ☐	I use frequently ☐

f. Fractions as operators:



I do not use ☐	I use occasionally ☐	I use frequently ☐

g. Fractions as repeated addition of the unit:

$\frac{3}{4} = \frac{1}{4} + \frac{1}{4} + \frac{1}{4}$

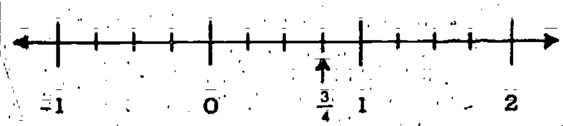
I do not use ☐	I use occasionally ☐	I use frequently ☐

h. Fractions as decimals:

$\frac{3}{4} = .75$

I do not use ☐	I use occasionally ☐	I use frequently ☐

i. Fractions as points on the number line:

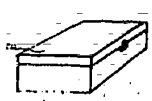


I do not use ☐	I use occasionally ☐	I use frequently ☐

j. Fractions as measurements:

this container holds $\frac{3}{4}$ ℓ 

or

this box weighs $\frac{3}{4}$ kg 

or

this stick is $\frac{3}{4}$ m 

I do not use ☐	I use occasionally ☐	I use frequently ☐

k. Fractions as number pairs:

three fourths as (3,4)

I do not use ☐	I use occasionally ☐	I use frequently ☐

dents. This additional information thus makes possible a contribution to a differentiation index, RD-F (Concept Representation Differentiation, Fractions), for the topic and through it a contribution to the across-topics version of the index and to other, higher-order variables. Since frequency and/or emphasis information is also available, this additional information makes possible contributions to intensity (weighted differentiation) indices as well.

Still other items in the cluster of related items secure information on the reasons for use and non-use of the interpretations involved. Utilization of this information relies on a facet model for the item type dealing with reasons for use/non-use. One version of that model is presented in the next pages. It uses four major categories of reasons with various sub-categories for each. Within each category two versions of reasons are available, positively stated reasons for use and negatively stated counterpart reasons for non-use. The sample statements of each category and version are exactly the choices for items of this type throughout the various questionnaires (there are currently eleven other such items).

The particular items from the example can thus be scored in terms of the categories (or even sub-categories should that level of analysis prove fruitful) of the model. This makes possible a contribution to a profile of the factors entering into instructional choices for this teacher. The model, in fact, provides a typology for this aspect of instruction and makes it possible for these items to contribute to diversity indices (both a variety index, COV, and a focus index, COF) for this aspect and, through them, to other, higher-order indices.

The fate of this example is typical. First, the central item and cluster of related items provide item level variables and add to a first-order descriptive portrait of elements of instructional practice. Second, they make contributions to a number of different aggregate indices (still of the more homogeneous type of cluster). Third, the aggregate indices that emerge can be further aggregated both across topics and/or to still higher-order statistical- or model-based clusters. Next, these indices at varying levels of aggregation can contribute to studying relationships within the portrait of instructional practice (i.e., within the Classroom Processes component) and, finally, they may contribute to analyses relating the three components of the study by connecting these sort of indices to topic-specific achievement subscores, to influences identified in the Curriculum Analysis (e.g., Does national policy affect differentation? Do national syllabi and examinations affect diversity?), etc. This combination of simple, specific items with contributions to a complex network of portraying variables would seem to be one of the major strengths of the classroom processes component of this Study.

REASONS FOR USE	REASONS FOR NON-USE
1. Student: Cognitive	**1. Student: Cognitive**
1.1 Meaning (1) Allows student to associate meaning to symbols ____. (2) Provides a meaningful context for ____. (3) Concrete interpretations are meaningful.	1.1 Meaning (1) Not an appropriate interpretation. (2) Not appropriate for the formula studied.
1.2 Understanding Knowing more than one method aids understanding.	1.2 Understanding (1) Learning rules does not aid understanding. (2) Premature learning of rules interferes with understanding.
1.3 Structure (1) Allows student to distinguish ____ from ____. (2) Allows student to see similarities of ____ to ____. (3) Allows student to relate ____ to ____. (4) Allows student to use previously acquired knowledge to learn ____.	
2. Student: Affective	**2. Student: Affective**
(1) Interesting (and/or appealing) to students. (2) Easiest for students to understand. (3) Easier for students to have definite rules.	(1) Though students would dislike it. (2) Might confuse students.
3. Subject Matter	**3. Subject Matter**
3.1 Effectiveness (1) Felt it would be effective. (2) Most effective technique. (3) Powerful technique for mental arithmetic	3.1 Effectiveness (1) It is ineffective. (2) It is too time consuming* *(or did they mean to teach rather than to use; if so, then 4.4)

(Cont'd.)

3.2 Justification (1) Mathematic justifications are important. (2) Provides mathematical justifications for the steps of ________.	
3.3 Applications (1) Application students likely to see in the future. (2) Numerous interpretations facilitate applications. (3) Students need practice in describing natural phenomena through mathematic symbols.	
4. School/Teacher System	4. School/Teacher System
4.1 Textbook (1) It's in the textbook (2) It's emphasized or recommended in the textbook.	4.1 Textbook (1) Not emphasized or recommended in the textbook.
4.2 Syllabus (1) It's required in this school	4.2 Syllabus
4.3 Teacher	4.3 Teacher (1) Not familiar with the interpretation. (2) Did not think of using it.
4.4 System Demands	4.4 System Demands (1) No time to present all ________. (2) Required background students did not have.

8. Component III: Student Outcomes (The Attained Curriculum)

As has already been stated and illustrated, the Study probes the curriculum at three levels. The curriculum analysis seeks to portray the curriculum as intended by the educational system; the classroom process component portrays the ways in which the curriculum is implemented by the teacher in the classroom. The third component, the subject of this section, explores the nature and extent of curricular goals as they are demonstrated by student attainment or achievement measures.

Two broad classes of measures are utilized: cognitive and affective. Proposals analyzing the data at the item and subscore levels are given in Section 14.

It is envisaged that <u>limited</u> supplementary data from students will be sought in the areas of general background information (sex, age, breadwinner occupation, etc.); and classroom process (including opportunity to learn and uses of hand calculators).

The following aspects of student outcomes are of particular interest in the Study: benchmark comparisons; growth during the academic year; linkage with classroom process; sex differences.

8.1 Benchmark Comparisons

In many IEA countries, a major concern is a comparison of the status of mathematics education now with that of fifteen years ago, the date of the first survey. Although many changes have taken place in mathematics education in the past decade and a half, there are little empirical data to document the extent and magnitude of these changes from an international perspective.

<u>Cognitive Measures</u>. "Anchor Items," that is, items used in the first survey, will be selected for the purpose of comparing student achievement then and now. Criteria for selecting items from the first survey will include representation of V cells or current grid, and satisfactory psychometric properties. Items will be sought that can be classified as follows:

		Low Level	High Level
Population A:	Arithmetic		
	Algebra		
	Geometry		
	Measurement		
	Probability/Statistics		
Population B:	Number		
	Algebra		
	Geometry		
	Analysis		
	Statistics		

Appendix C provides further information on the anchor items.

Affective Measures. The change of focus for the affective responses has implications for how the new Study will dovetail with the old one. A narrowing of focus implies the need to generate new scales rather than simply re-administer the old ones. Still, it is desirable to have items common to both surveys to provide a basis for comparisons between what exists now and what was found fifteen years ago. Although the old scales have a different focus, there are items in those scales that can be utilized in the new survey. A goal in the construction of affective scales is to include those items that have functioned well in the past but to supplement them with items and scales that more closely fit the goals of the new survey.

8.2 Growth During Academic Year

A growth measure for classes during the academic year is planned internationally for Population A and as an international option for Population B. The primary justification for a growth measure resides in the concern of the Study for the classroom. In order to focus upon the classroom, what happens there and what students learn, it is essential that data which references the curriculum as attained by the student entering the class be available when end of year measures are obtained. Hence, a pre-test achievement and attitude measure at the classroom level will be sought. Some have proposed using such measures as an index of school "effects": emphasis on education; qualifications of teaching staff; impact of supplementary resources, etc. The primary intent, however, is to use pre-test measures on classrooms as a covariate in isolating learnings which have occurred in the classroom during the school year.

8.3 Sex Differences (as related to issue of between-student analyses)

"Women and (or in) mathematics" is a topic receiving considerable attention in several countries at this time, according to preliminary information received by the Curriculum Analysis Group. While this topic is only illustrative, it is important that some between-student analyses be possible even though the main thrust of the Study is at the classroom level.

9. Versions of the Study

As work on the Study has proceeded, certain modifications in the Study have emerged. Feedback from national committees has pointed to the need for further work on the selection of items for the cognitive test, and probably for a more complex structure of the instrument for the longitudinal study (see Section 12). The demands which the proposed study will likely make upon national centers and schools, in terms of administrative details, in terms of teacher time for responding to the classroom process questionnaries throughout the school year, and in terms of maintaining close contacts between schools and national centers all point to two needs: (1) a "feasibility run" of the full Study in at least one country, and preferably several, should be conducted to demonstrate the manageability of what is proposed (2) every country intending to participate in the full study should plan to undertake a full scale dry run of all instruments and data collection procedures. This should take place in 1980-1981. Finally, problems in funding have continued to impose further delays on the time schedule.

The International Mathematics Committee therefore agreed, at its January 1979 meeting (see Section 2.3.2) to view the Study as having two versions, or elements. One is a reduced, or cross-sectional study, which makes fewer demands upon the countries in terms of data collection, testing time (testing takes place only once during the school year) and costs (administrative, printing, etc.). This version of the study consists, essentially, of Components I (the curriculum analysis) and III (the student outcomes survey) and eliminates the investigation of instructional practices (the classroom processes component) for <u>Population A</u>.

For Population B the version of the Study remains for all countries as announced in Bulletin 3. The implications of these plans for timetabling can be summarized as follows:

i. Population A (cross-sectional): proceed on timetable announced in Bulletin 3 and reproduced in Section 15 of this Bulletin.

ii. Population A (longitudinal): conduct full scale dry run during 1980-1981 and collection data <u>one year later</u> than indicated in the timetable.

iii. Population B: Since a longitudinal study is a <u>national option only</u>, countries doing the cross-sectional study will proceed on the schedule indicated in the timetable. Countries interested in the longitudinal study may wish to get in touch with the U.S. National Coordinating Center, University of Illinois, Urbana (attention: Kenneth Travers) for copies of the classroom process instruments being developed for the calculus and other topics at that level). Recommended dates for longitudinal study, population B, are 1980-1981.

9.1 <u>Population A Cross-sectional Version</u>

Some countries have indicated that they wish to take part in only a cross-sectional study at the Population A level. At least two of these countries have indicated that they must collect data in the first half of

1980. The International Mathematics Committee therefore made efforts to have instruments ready for these countries by December 1979.

Cognitive Tests: Knowledge of Mathematics

For each population, final instruments were devised to address the needs of the cross-sectional study.

Population A Test

This test consists of a core form and four rotated forms (A, B, C, D). The core test is devised to provide a relatively comprehensive measure of mathematics achievement with the balance of items from the V and I cells of the international grid assigned by content and behavior stratification to the rotated forms. The number of items in the 5 forms reflects the weightings of the V and I cells in the grid and includes about 30 items from the First Study (these are the "anchor items").

Population B test

As this test was originally structured (see, for example, Bulletin 3, page 29) it consisted of 8 forms of 15 items each. The seven forms were constructed by random assignment of items from the international pool with stratification on content and behavior. The eighth form consisted entirely of items on the calculus.

Those seven forms were constructed using the international grid as a blueprint, and for some countries provide an adequate reflection of curricular emphasis. For other countries, however, those seven forms do not provide sufficient emphasis on the calculus. Therefore, for those countries all eight forms were to be used.

Two concerns caused the IMC to propose a new structure for test. First, the use of a test form defined by content (as opposed to random assignment) introduces methodological problems. For example, the sampling errors of the estimates of item difficulty on a form where there is no random assignment will have inflated item intercorrelations. Thus, a substantial amount of imprecision, and perhaps bias, would be introduced into the estimates of item and subscore parameters.

An additional concern had to do with administrative practicalities. The calculus form is, of course, quite different from the other form and could cause confusion or concern in the classroom. It is also likely that the calculus form would take longer to answer than the other forms.

The Population B test has therefore been restructured by the IMC and presented to the countries for consideration. This restructured version is proposed for use by all countries participating in the cross-sectional study. Those countries which, as a national option, engage in the longitudinal study, will have available either the version of the Population B cognitive test presented in Bulletin 3 or some other structure which responds to their interests and needs. For example, a core test which yields subscores on such topics as trigonometry, systems of equations and elementary functions may be appropriate. The IMC will assist in the exchange of information between countries concerning plans for the Population B longitudinal study.

The new international version of the test consists of eight forms of 17 items each. Each form is a stratified (by content and behavior) random sample of the item pool which was devised on the basis of the international grid.

Attitude Scales

Four scales in their recommended final form are as follows:
Mathematics in School
Mathematics and Society
Mathematics as a Process
Mathematics and Myself
Computers and Mathematics

Details on these scales, together with results of the pilot testing of the items, are available in the IMC document, Report of the Attitude Trials, December, 1979.

Questionnaires

During 1979, National Centers had two opportunities to comment on the school, teacher and student questionnaires for the cross-sectional version. These instruments were finalized in September, 1979.

Opportunity-to-learn ratings

It is planned to gather these ratings from the students and their teachers.

Sampling

A sampling manual for the cross-sectional version has been sent to each country including a timetable for the negotiations with the sampling referee. This timetable recommends a one-year lead time before data collection to allow for drawing the samples, securing permission from schools, and making other arrangements.

Administrative manuals

Drafts of these manuals were sent to National Centers in early 1979. Finalizing took place during the September 1979 meetings.

Data Processing

Despite the continuing lack of funding, contingency plans have been drawn up by the International Coordinator. Dr. Larry Nelson, University of Otago, New Zealand, has been contracted to provide guidance for the data management. One of the Research Officer positions within the Coordinating Unit has been redesignated to allow for the appointment of an experienced programmer. Budgetary provision has been made for the necessary computer time and negotiations are continuing with the Department of Education for the installation of a visual display unit and hard copy terminal in the Department of Education. Difficulties are being experienced in finding a source of funding to purchase Osiris IV as an editing and file building package. Once this is solved, work will begin on establishing the necessary routines with dummy data which are already prepared. Efforts will be made to coordinate the work of this unit with that of the Data Processing Centre for the longitudinal study at the University of Illinois.

9.2 Longitudinal Version

Developmental work on the various instruments will proceed in 1980.

a) Background questionnaires (School, Teacher, Student) and attitude scales
The questionnaires as produced for the cross-sectional study will form a substantial basis for the full study.

b) Classroom process questionnaires
These will be pilot tested in about six countries with an attempt to sample diverse instructional practice. As a result, considerable "internationalizing" of the instruments in order to account for differences in instruction should emerge.

c) Cognitive instruments
Suggestions for further development are outline in Section 12. The IMC will strive to produce instrumentation in time for dry run beginning in September 1980.

For countries which, of necessity, must proceed with data collection for Population A during 1980, interim forms of the instruments (reduced study) will be available by December 31, 1979. It is expected that the overlap, in terms of items, between these instruments and the final instruments will be maximal. That is, at least 90% of the items will be common to both sets of instruments.

10. Research Hypotheses

10.1 Level I: Educational System

The data from both the Contexts and Status Surveys of the Curriculum Analysis will make possible the examination in a descriptive sense of relationships illustrated by such research hypotheses as:

1. It is expected that the existence of national examinations is accomplished by minimal variation in curricular patterns and characteristics.

2. It is expected that curricular change takes place more readily where such change is supported, and to a considerable extent directed, by external examinations.

3. It is expected that more variation in curricula exists where there is a number of discrete examining bodies with different foci and constituencies as contrasted with jurisdictions in which there is one examining authority or no system-based examining authority.

10.2 Relationships between Level I and Level II

1. It is expected that the more overt and explicit the control in the educational system, as indicated by the presence of a national curriculum, a centralized inspectorate and a system of external examinations and/or grading systems, the less variation will be found between teacher opinions on curricular issues, justification for and use of instructional strategies and both within and between classroom use of instructional strategies and resources.

2. It is expected that more limited resources, in terms of institutions and facilities for, say, teacher education, will lead to more common socialization of teachers, hence less variation in teacher attitudes, opinions, expectations and instructional behaviors.

One aspect of the linkage between the curriculum at the system or national level and the classroom is that of teacher coverage or "opportunity-to-learn". One view of such a measure is to regard it as an index of the curricular validity of the international tests (the greater the extent to which the subject matter has actually been taught in a country's classrooms, the more appropriate is that test as a measure of student achievement in that country and as a corollary, the more meaningful are cross-national comparisons on that test). As in past IEA studies, we would expect that teacher coverage will be a relatively powerful variable in terms of accounting for variance in student achievement.

There is another direction in which interpretation of the teacher coverage index could move. One could speak of congruence between the intended and implemented curriculum as some sort of measure of teacher competence or effectiveness, and contemplate some sort of ranking of national systems of education on the basis of the degree to which the systems' teachers implement the national curriculum.

Such an interpretation is fraught with pitfalls, however. A fundamental problem is that teaching is not a mechanistic process whose success depends upon the degree to which desired information can be delivered, or "laid on" students. The essential dynamics of a classroom, the sensitive interaction between teacher and student, the judicious selection of appropriate subject matter and instructional approach for a given day are among the factors which are to be taken into account in an educational system. Teacher coverage is an important measure, but is only one of a vast complex of critical variables which enter into the mix for effective teaching and learning.

10.3 Hypotheses Linking System and Student Outcomes

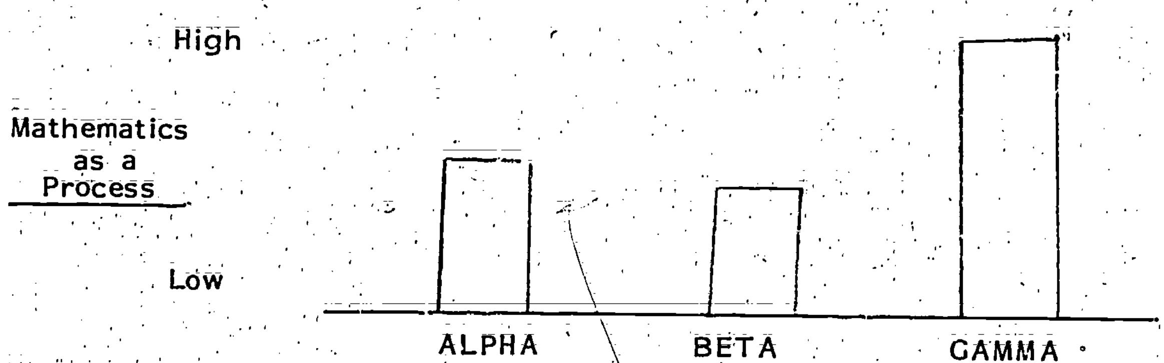
This class of comparisons investigates possible relationships between characteristics of the curriculum and accompanying student outcomes.

Attitudinal Dimensions

There is considerable overlap between the attitude scales of the first and second mathematics study. This should permit useful benchmark comparisons for countries which participated in the first study. In particular, countries in which student attitudes were manifestly negative in the first study may wish to identify possible shifts on these measures. The Mathematics as a process scale, designed to measure one's view of the nature of mathematics may also be examined in the light of curricular emphasis in the various countries. One might hypothesize, for example, that curricular emphases in Alpha and Gamma, combined with the above projected profiles of classroom process in those countries, would be accompanied by Mathematics as a Process outcomes such as suggested below.

Mathematics as a Process

This scale is intended to obtain a measure of the student's perceptions about the nature of mathematics. It is hypothesized to reflect differences in curriculums and in classroom processes. For example, emphases on mathematics as a creative enterprise, and classroom approaches that encourage open-ended investigations on the part of students, are expected to impart a view of mathematics that is dynamic and creative.



Cognitive Dimensions

The most popularized aspects of this sort of study are, undoubtedly, the student outcome measures, and particularly the cognitive aspects. In view of the considerable curricular activity over the past 20 years, one should expect attention to focus upon comparisons between student achievement "then and now." The results of the National Longitudinal Study of Mathematical Abilities, which examined student outcomes over a five-year period beginning in 1962 for curriculums based on "conventional" and "modern" mathematics textbooks, suggest the sort of findings one might expect for the United States, and, assuming reasonably comparable definitions of "conventional" and "modern", and accompanying implementation in the classrooms (bold assumptions, to be sure!) one might expect that countries which have undergone extensive "modernizing" of the curriculum will exhibit student outcome measures which have increased on the upper levels of cognitive behavior while at the same time some loss of ground on the lower levels might be anticipated. On the other hand, countries which have experienced less "modernizing" of the curriculum might expect less shifting across behavioral levels and (again, other factors being equal, which they never are) more congruence of student outcomes between the first and second studies.

The figures on the following page present mean achievement scores for students in "modern" (Textbook T_1) and "traditional" (Textbook T_2) textbook groups, transformed to standardized scores having a mean of 50 and standard deviation of 10. The circles in each figure represent textbook groups. Therefore, the figures present profiles illustrating the relative position of the various groups on 37 cognitive scales used in the survey. In Figure 3 the darkened dot represents a "modern" text, the SMSG program. Notice the relative position of these groups on the lower cognitive level as to the higher levels. In the second figure (Figure 4) the darkened dot represents a conventional text. The relative position of this group is in many respects a striking contrast to that of the modern text.

(The following two figures are from E.G. Begle and J.W. Wilson, "Evaluation of mathematics programs," Mathematics Education, Sixty-ninth Yearbook of the National Society for the Study of Education, 1970, pages 396 and 398, respectively.)

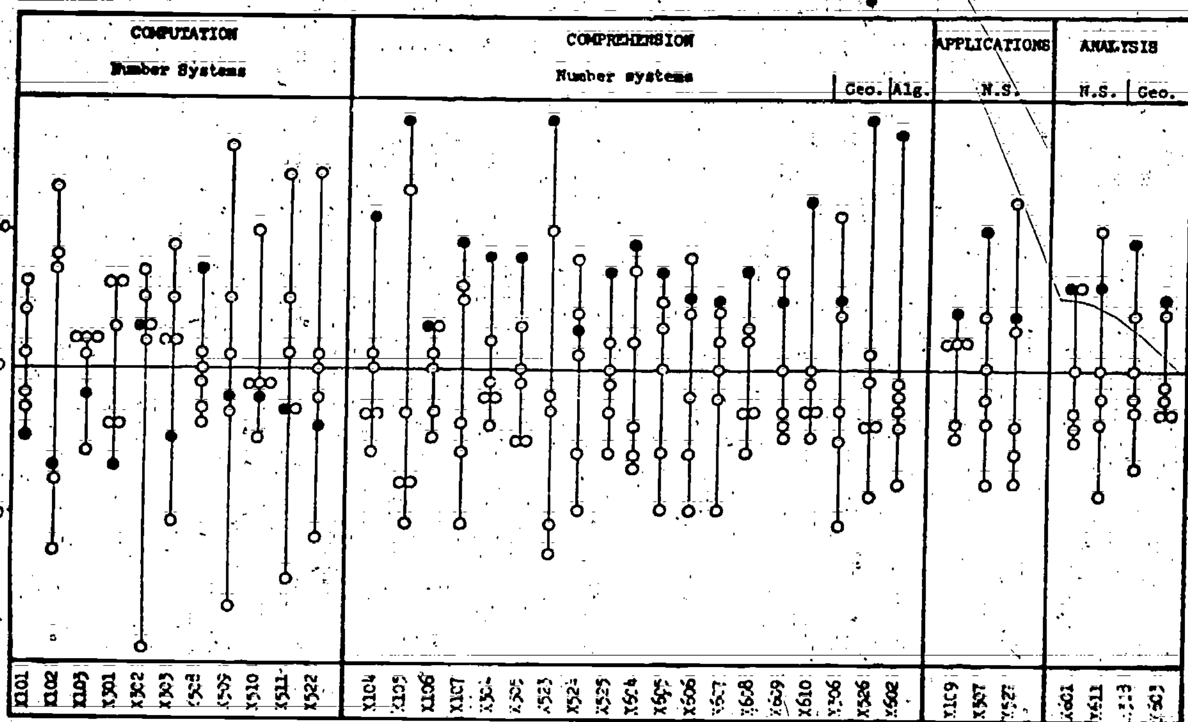


Figure 3 -- Profile for Textbook Group T_1
(X-Population; Grades 4, 5, 6; Years 1, 2, 3)

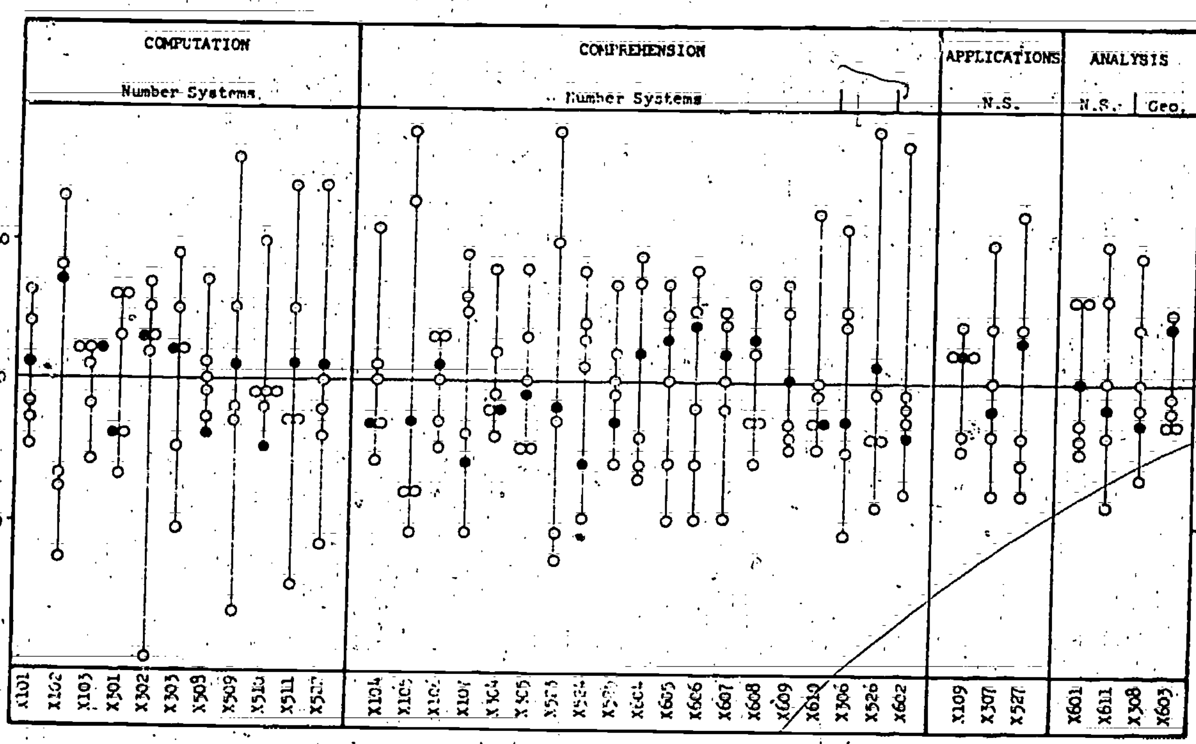
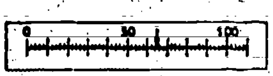


Figure 4 -- Profile for Textbook T_2
(X-Population; Grades 4, 5, 6; Years 1, 2, 3)

Item Level Information

As has already been stated in this bulletin (Section 4.3), and elsewhere (see, for example, Bulletin 3, pages 36-37), the International Mathematics Committee wishes to focus upon simple description of outcomes at the subscore or item level in an attempt to present findings which are most easily understood and communicated. There is also the intention, however, to preserve the data in a format which enables analyses based upon a variety of models.

One of the potentially most valuable sources of information out of the first mathematics study, which still remains to be fully mined and is reported in Appendix II to the report of that study, is the item data indicating those countries in which each item was hardest and in which it was easiest. A sample from the Population 1 (our Population A) results is provided.

Item	Target Population	Average Difficulty	Average Discrimination	Order of Error Frequency	Easier in	Harder in
7. A box has a volume of 100 cc. Another box is twice as long, twice as wide and twice as high. How many cc is the volume of the second box? *Ans. 800*	1a	.21	.39		—	Australia, Sweden, United States
	1b	.24	.40		Belgium, Israel	Australia, Sweden, United States
8. On the scale to the right, the reading indicated by the arrow is between  A. 51 and 52; B. 57 and 58; C. 60 and 62; D. 62 and 64; *E. 64 and 66	1a	.42	.45	B, D	England, Japan	Belgium, France
	1b	.44	.44	B, D	England, Japan	Belgium, France, Israel

One can offer many hypotheses for explaining the differential difficulty of these items in various countries. If it can be assumed that Item 8 has more of a "real world" flavor than Item 7, what is it about the curriculum in England and Japan which might account for the relative ease of the scale reading task? Is the applied nature of mathematics much more evident in the curriculum of those countries than in Belgium and France? Item 7, on the other hand, has its own intriguing characteristics. Does success on this item reflect understanding of volume in any deep sense or the ability to use formulas? And again, what in the curriculum might account for the differences?

Certainly, instructional effects must be taken into account as well. The kinds of experiences in which children are engaged as learning takes place would surely be reflected in performance on items such as these. Children who have actually stacked blocks and counted as they learned the concept of volume could be expected, generally, to be more successful at solving problems like the one posed in Item 7. But, on the other hand, prior experience in applying mathematical formulas might over-rule and account for the success of, say, the Belgians.

The following hypotheses have been proposed by the Curriculum Analysis Group for the attitude scales.

Concerning relationships within measures:

1. Strong positive correlations are expected between:
   a. Mathematics in School (ease, liking) and Mathematics and Self;
   b. Mathematics in School (importance) and Mathematics and Society;
   c. Mathematics in School (like) and Mathematics and Self;
   d. Mathematics as a Process and Computers and Mathematics
2. The strongest correlate of growth in achievement will be Mathematics and Self.

Concerning within country relationships:

a. Characteristic national profiles on the attitude score will be found.
b. These profiles will help provide a context for national achievement scores.
c. Congruence between teacher and student profiles for the common measures (Mathematics in School and Mathematics as a Process) will be found within countries.

Concerning within class relationships:

a. Characteristic profiles of classroom attitudes will be identified, and be found to vary according to:
   (1) type of school
   (2) type of program
   (3) teacher characteristics (age, sex, training, methods used in class, attitudes)
   (4) type of curriculum

Concerning between-student relationships (note: attitude items are common to all students).

Individual patterns of attitude will vary according to:

a. achievement
b. sex
c. socio-economic level and parental education
d. expected education
e. expected occupation
f. birth order and family size

The rationale for these hypotheses comes from a sense that group attitudes will result from curriculum and instruction but that individual variation from the group will result from individual characteristics.

11. Sampling Issues

The technical aspects of the international sampling design and guidelines for implementing the plan in each country are to be found in the sampling manuals. Persons requiring information on procedures to be followed in drawing samples of their own countries, or on other aspects of sampling should refer to those documents. The manual for the cross-sectional version of the study was prepared by the International Sampling Committee and circulated to all countries for initial commentary in March 1979. The final version was circulated in June 1979. The International Mathematics Committee is indebted to the Sampling Committee for the work represented by the sampling manual.

The present section summarizes certain of the main features of the sampling design as enunciated in documents from the Australian Council of Educational Research (notably, M. J. Rosier, K. N. Ross and J. P. Keeves, "The Sampling Design," 1978 in Sections 11.1-11.9). Comments on the design, particularly as they relate to the full (longitudinal) study, are drawn from a paper by R. Wolfe (1979), Sections 11.8-11.13. It is intended that the issues raised here help highlight the particular problems of sampling which attend the aims and design of the longitudinal study.

11.1 Purposes of the Study

The purposes of the Study are outlined in Section 4 above. As indicated there, data to address these purposes will be obtained through questionnaires and written documents at the national or system level, from teachers and coordinators of mathematics within schools and through tests, attitudes scales and questionnaires from students.

11.2 Procedures

The statistics used to summarize the data collected by the Study will, in the main, be proportions responded to in a specified way to items in tests, attitudes scales or questionnaires, or will be mean values of responses on a subtest or scale. Analysis of the classroom process data will involve some explanatory model, such as regression for use in the main report of the Study in seeking to account for the effects of curriculums and instruction on learning. The technical volume will exhibit a variety of models used to explain the classroom process and student output data.

There will be an emphasis on the descriptive and comparative aspects of the data collected in the preparation of profiles for cognitive test items, attitude scale items, student and teacher views and classroom and school practices. It is intended to present these findings as simply as possible so that the reporting will be meaningful to an audience of mathematics educators, curriculum co-ordinators and teachers. However, there will also be an emphasis on the explanation of change in student performance over a school year. This will require the development of statistical models to account for change, as well as the development of appropriate analytical procedures for the examination of the models.

To the extent that this study seeks to make comparisons with respect to performance in mathematics within a country between 1964 and 1980 or comparisons between different countries in 1980 it will be necessary to maintain comparability in definition of target populations and in measures of student performance both between countries and between occasions.

These many constraints on the design of the study and the sampling plan, in particular, will be difficult to satisfy. Consequently, it will be necessary to allow each of the participating countries some latitude to select options provided in the design of the study and incorporated in the sampling plans to answer the specific research questions of greatest relevance to them and to conduct the study within the financial provision available to them. Some of the options available are outlined in subsequent sections.

11.3 Sample Sizes

Previous IEA research has indicated very convincingly that in any multivariate analyses serious difficulties are encountered if the number of schools in a between-school analysis falls below a level of about 70 schools. Furthermore IEA experience suggests that consistent and meaningful findings are only obtained when about 100 to 200 cases are used in any analysis in which more than a very limited number of predictor variables are introduced into a regression equation. This experience is in agreement with the advice given by Kerlinger and Pehazer (1973, p. 46). Moreover, experience indicates that to have less than 15 students from any school or classroom yields unstable results and to include more than 30 students from any school or classroom provides redundant information. Thus an intact class group of 25 to 30 students, with the loss of cases between the pre-testing and the post-testing not exceeding 10 students to yield a final group size of 15 to 20 students, would appear a sound basis on which to collect data.

Nevertheless, it is important to recognize that different National Centers have different levels of financial support available to them. As a consequence it would be important to maintain a degree of flexibility in the design of samples, so that no centre would be prevented from participation on the grounds that financial support would not be available for testing a large enough body of students for full analyses to be carried out. Where only a sample of limited size can be tested, certain types of analyses of the data may not be possible but other important results can clearly be obtained. Yet it would also be important to caution National Centres that in the design of a sample, the use of too large a sample involves not only unnecessary expense but collects data that are, in the main, redundant. The sample sizes discussed in this section are those seen to be the optimal sizes for carrying out an efficient and unrestricted investigation. Table 11.1 records what is seen to be the optimal range within which any sampling of schools, classrooms and students should take place together with a recommended sample size.

Recommended Sample Sizes

Table 11.1

	Lower Bound* (R)	Lower Bound* (F)	Recommended	Upper Bound	Recommended Total
Schools	70	70	100	150	100
Classrooms within schools	1	2	2	3	200
Students within schools					
Pre-test	20	20	25	30	5,000
Post-test	15	20	20	25	4,000
Total students in analysis	1,050	2,030	4,000	11,250	
Total classrooms in analysis	70	140	200	450	

*Lower Bound (R): For reduced study. Does not provide for disentangling school and classroom effects.

Lower Bound (F): For full study. Enables isolation of school and classroom effects.

In the design of a sample, a National Center faced with limited resources should consider carefully which aspects of the analyses it wished to forego and thus whether it wished to maintain an adequate number of schools for an effective, between schools analysis, an adequate number of classrooms for a between classrooms analysis or an adequate number of students for a sound estimation of national statistics. If a National Center were not concerned with distinguishing between the effects due to schools and those due to classrooms within schools it would restrict its sample to only one classroom per school. If, however, a National Center sought to study fully the effects of classroom and teacher curricular differences within schools, it should increase the number of classrooms selected within a school to two or three. It should be noted that the practice of forming pseudo-schools as a composite of two or three classrooms from different schools or of forming pseudo-classrooms of two or more smaller classrooms while beneficial for making national estimates would not be meaningful in a study of classroom and school practices. If a country had a substantial number of small rural schools with less than 25 students per class or with less than two classes in the target population, a separate stratum of such schools could be formed. This would enable such schools to be included in the calculation of national estimates and in the analysis of change in performance over a school year.

It is increasingly common to find schools and classrooms organized on an open plan with fifty or more students in the classroom group. In such a situation it would clearly be redundant to collect information on student performance for more than 25 or 30 of these students, and some sampling of students within classrooms would be desirable.

11.4 A Sampling Plan

In the development of a sampling plan each National Center would obtain a list of all schools within the regions in which the study was being conducted. Thus in the United States it might be considered desirable to select certain school districts or certain States for involvement in the investigation, while in Australia it might be considered important to include all States and Territories, but to draw separate samples from within each State. The listing of schools would be stratified by appropriate school stratifying variables (e.g., school type), which from previous research studies were known to be related to student achievement and teaching practice. In addition, an estimate of the number of students within the target population should, if possible, be obtained for each school, together with an estimate of the average size of class so that the number of class groups within each school could be estimated. In the first stage of sampling schools would be randomly selected with a probability proportional to size, estimated in terms of either the number of students in the target population within the school or the number of class groups within the school. (Stratification of schools by size will usually be essential.) In the second stage of sampling, one, two or three class groups would be randomly selected from within each chosen school for participation in the study, again with a probability proportional to size. In those classes within schools where there was an adequate degree of uniformity in class size, it would be unnecessary to sample students from within classrooms, provided between 25 to 30 students were present within each class. If, however, classrooms contained substantially more than 25 to 30 students it would be more efficient to sample students from within classrooms.

Since three, and in certain cases four, stages of sampling are envisaged in this plan the design is necessarily a complex one and any accurate estimation of sampling error cannot be undertaken by formula but only by an examination of the data collected in the study. Two procedures are available for the calculation of sampling errors, namely 'balanced repeated replications' and 'jacknifing.' Where it is necessary for a large number of strata to be formed it would be important that at least two schools were drawn from each stratum so that balanced repeated replication procedures could be used. Where relatively few strata were included in the sampling frame it would be important to have at least five schools drawn from each stratum so that jacknifing procedures could be used most efficiently.

Alternative procedures are available for obtaining a sample of schools and classrooms within schools where selection is made with a probability proportional to size. The procedures depend on the available information associated with the number of students and the number of classrooms within each school. The selection of an appropriate procedure should be made according to the nature of the information most readily available. Further details of alternative procedures are given in the Sampling Manual (IEA [MATHS-NZ]/A/149). It is, however, important that random selections are made at each stage of sampling.

11.5 Units of Analysis

The data which are to be gathered for the study may be analysed at different levels of aggregation. Table 11.2 describes the level of aggregation for four possible sources of data (students, teachers, classrooms, schools) when applied to four possible units of analysis (students, classrooms, schools, countries).

Table 11.2
Level of Aggregation of Data for Different Units of Analysis

Source of data	Unit of Analysis			
	Students	Classrooms	Schools	Countries
Student data	Same	Aggregate	Aggregate	Aggregate
Teacher data	Disaggregate	Same (Wtd)	Aggregate	Aggregate
Classroom data	Disaggregate	Same	Aggregate	Aggregate
School data	Disaggregate	Disaggregate	Same	Aggregate

The development of descriptive profiles could be undertaken by using students as the unit of analysis for student data, and classrooms as the unit of analysis for classroom data and teacher data. In sample designs where more than one classroom per school is to be selected then the descriptive profiles for school data could be undertaken by using schools as the unit of analysis.

The consideration of factors which explain change in mathematics performance will focus on the classroom and the school as the unit of analysis. In some schools the teaching of a class group is shared by two or more teachers either within an open plan or team teaching situation or merely to maintain a balanced workload for each teacher. In such circumstances it will be necessary to combine the information composite of those teachers providing a learning experience for the classroom group. The basis of weighting should be the time spent with the classroom group.

11.6 Estimates of Error

When using data from different sources, namely: students, teachers, classrooms and schools different levels of error are present for the different statistics likely to be employed in this investigation. In estimating, prior to the collection of data, the errors which are involved in the presentation of the results of this study it is necessary to make guesses of the probable values of the design effects to be found.

Table 3 records estimates of the standard errors for several different statistics: proportions, means, and regression coefficients. In making these estimates we have assumed an intra-class correlation for students within schools of 0.1, for students within classrooms of 0.2, and average sizes of school and classroom groups of 40 and 20 students respectively. These assumed values are consistent with estimates for these statistics made in earlier IEA studies, although considerable variation has been found between countries according to the nature of the school and curriculums existing within those countries.

Table 11.3 Estimates of Standard Error for Different Statistics Using Recommended Sample Sizes

Statistic	Source of Data			
	Students	Classrooms	Schools	Countries
Proportions	0.02	0.04	0.05	-
Means[a]	$0.04s_{ST}$	$0.08s_C$	$0.10s_S$	-
Regression Coefficients	0.04	0.08	0.10	
Recommended Number	4000	200	100	1
Simple Equivalent Sample	800	140	100	

s_{ST}, s_C, and s_S are the standard deviations for students, classrooms and schools respectively.

Since the testing program will involve the administration of a core test with four rotated tests being administered to sub-samples of the students some test items will be taken by only 1000 students. This reduction in the number of students taking certain items will only increase marginally the standard errors for proportions and mean values recorded in Table 3 for data collected from students, provided all four rotated tests are administered in each classroom.

It should be noted that much published research has usefully employed regression coefficients which have magnitudes much less than 0.16. Consequently, it would seem that the use of only 100 schools, or a simple equivalent sample of only 140 classrooms would run the risk of failing to detect as significant some potentially interesting school and classroom variables. For a full examination of school and classroom effects it would clearly be desirable to approach as near as financial resources would allow, the upper bound of sample sizes recommended in Table 11.1.

11.7 Level of Response

During the IEA Six Subject Survey, which was limited to the collection of data at one point in time, the sampling losses in the execution of the sample design were such that in 10 out of 20 countries the response rate was less than 80 percent of the students in the designed sample and in seven of these ten countries the response rate was less than 60 percent of the students in the designed sample. Therefore, it is important to recognize that in this study we are attempting an extremely ambitious data gathering operation by collecting data at two points in time, from, in general, two or more classrooms in a school with a substantial demand on the teachers during the school year in the completion of questionnaires. While the sampling plan may be carefully designed and the level of response in the initial stages of the investigation high, the important evidence for the study is being collected at the final stages and during the school year. It will be necessary for efforts to be made to reduce the administrative load on schools and teachers in order to maintain as high a level of response as possible so that sound analyses of the data can be carried out and useful generalizations made.

Commentary on the Sampling Plan

11.8 Facilitating National Analysis

It is likely that the Second Mathematics Study will effectively preempt the energy and resources for high-quality international surveys in mathematics for a long time, certainly a decade or more. It is even more important to keep in mind that for many countries, participation in the Second Mathematics Study will absorb the local energy and resources for extensive national surveys in mathematics for an equally long time. It can be argued that, ultimately, most of the users and users of the study will be for national analysis, interpretation and planning.

For national interests, therefore, the sample designs must be arranged to facilitate within-country analysis. This may be more important than provision of national summaries or international comparison. Two aspects of the sampling process are critical to facilitating national analysis.

First, national centres need to be directed, encouraged, and assisted in determining substantively relevant major stratification for their studies. Certainly in all countries there are important regions, types of communities, types of schools, or subpopulations of students or teachers for which nationally important educational policy questions need to be answered. A process at the national level of discussion and debate needs to be carried out to decide what the critical strata are. In one country these may be regional, while in another they may be according to school stream.

11.9 Disproportionate Sampling: A National Concern

While sampling proportional to size (pps) yields probability samples at the national level, this method may be disadvantageous to some countries. One cannot assume that sub populations of critical national interest (such as minority ethnic groups or experimental technical schools) are present in equal proportions.

In all practical cases, national averages will be sufficiently accurate for international or longitudinal comparisons no matter what scheme is used within country. But if proportional sampling is used, small but critical groups of schools, teachers, and students will likely not be represented sufficiently in a sample to allow separate analysis, and that could diminish national interest in analysis.

The prospect of disproportionate sampling could mean reduced national accuracy, but if sufficient accuracy is obtained within each major stratum, it will be obtained overall as well. There will be cost, perhaps, in increased sample sizes, and the payoff will be in improvements just in national analyses. All national and international analyses will require weighting, since samples will not be self-adjusting. This means that the central data processing and analysis and the subsequent national work will have to program in weights. This does not seem to be an overwhelming problem and, in fact, the necessity of using weighting means that other sampling methodologies, such as some kinds of classroom sampling, which would also require weighting, need not be avoided.

11.10 Practical Error Analysis

The analysis methodologies proposed for the Second Mathematics Survey are more sophisticated and complicated than in earlier studies. Also, it is planned that a variety of complex analyses will be carried out and reported from data. It should be hoped that the complexity and variety considered necessary for realistic international analysis will be made available for national analysis as well.

The notion of applying classical error analysis procedures to such a variety of fancy analytic results in such a large number of national and international circumstances under conditions of very complex sampling is obviously impractical. What is needed is a general methodology for error analysis which can be applied immediately and consistently to all analytic results for the surveys.

The national sample designs should be arranged with proper balancing and replications to allow interpenetrating subsamples so that jackknife or balanced-replicate sampling formulas can be universally applied to statistical results. Perhaps the US-NAEP rule should be adopted: no statistic should go out without having an honest estimate of its accuracy. The advantage, of course, of subsampling-comparison procedures is that reasonable estimates of standard errors can be derived by simply repeating analyses on subsamples. Incremental computer costs (relative to set-up costs) are minor.

There is a need for explicit direction and guidelines to national centres to ensure replication arrangements when national sampling proposals are reviewed. It will also be important to encourage good replication arrangements when national sampling proposals are reviewed.

11.11 School Size and Sampling Classrooms

The Cross-Sectional Study to some extent and the Longitudinal Study fundamentally are concerned with what happens in the mathematics classroom and how that affects achievement. This means that the sampling of classrooms and the linkage of students and teachers are critical aspects of the design.

It is impossible to deal realistically with sampling classrooms (and therefore students or schools) without taking into account the fact that in many (most?) school systems, students are assigned to classrooms explicitly according to their mathematics achievements and interests and according to the instruction they are to receive. A sampling plan which does not take into account the streaming of students by ability or the differentiation by classroom of content of instruction is going to suffer terribly in terms of accuracy and potential for analysis.

The issue is connected to an issue of sampling according to school size. Small schools are likely to have different numbers and kinds of classroom arrangements. Perhaps this suggests that stratification by size is desirable for substantive reasons. In any case, stratification by size will be necessary for the technical reason of making rational, careful selections of classrooms.

It is desirable that schools first be stratified by size (of the population of interest, of course). Within size strata, there seems to be no reasonable alternative to careful and planned selection of classrooms according to the kinds of classrooms which appear. For the Longitudinal Study certainly and for any part of the Cross-Sectional Study which presumes to relate instruction to achievement, classrooms rather than students must be sampled. But if there are, say, two low and two high streams in a school, the sample must include one of each kind. If there were two low and one high, the sample would still require at least one of each kind.

This obviously complicates matters. The sampling and administration plans will have to contemplate further information collection, more extensive decision processes, and perhaps some in-field sampling. The data collection, processing, and analysis may have to provide for in-school weighting.

11.12 Selecting the Longitudinal Part of the Sample

One major variable in the study is "growth in mathematics" over the year. Analysis of this obviously requires representative samples and assessments of accuracy. The hypothesis that relational studies of variables can be carried out on judgment samples is certainly restricted to very good judgments or very special relationships. Preferably, probability samples will be employed, however.

The ideal approach to selection for the longitudinal study is random sampling. Careful, substantive stratification, probably with disproportionate sampling rates, should be encouraged. If general sampling is not feasible, there are a number of alternatives without eliminating the randomization. First, the scope of the survey could be restricted, for example, by taking a random sample within a convenient region or set of regions. Second, only schools of certain types, such as public schools, might be chosen. If non-random samples become absolutely necessary--that is, after attempts at drawing random samples fail--then it will be essential that the best non-random procedures be employed. For example, an explicit quota system might be used after fine-grain stratification. In this case, the national centres are going to need more guidance and assistance in defining the sampling.

11.13 Sampling of Teachers

The sampling plan calls for a national probability sample of classes. However, this does not automatically yield a concurrent sample of teachers associated with these classes which is random. In particular, the sample will be biased in the favor of those teachers whose teaching assignment includes classes in the target population. It is important to keep in mind, therefore, that the data associated with the sampled classes is a probability sample of the instructional environment in which the national probability sample of students is embedded. It might be found, for example, that 30 percent of the teaching of addition of fractions at Population A involved the use of number line. In national profiles of teachers (as opposed to teaching) were desired (What percent of teachers at the Population A level use the number line in teaching addition of fractions?) weighting would need to be employed. The school questionnaires will provide the needed information for the weighting.

12. Structure of the cognitive tests

The Cognitive tests are subject to many demands and constraints. For both populations, the demands include

12.1 Cross-national curricular validity

Meaningful comparisons across countries require measures which have validity in the countries involved. The effort made here is the familiar one of devising a test which is, as far as possible, equally "fair" or "unfair" to all countries, with their individual programs and educational goals. See Peaker (1969, pages 229-237) for a technical discussion of this issue, with specific reference to First Study.

The following procedures were followed in developing cognitive instruments which meet the curricular validity criterion. (These comments are drawn from a memo prepared by Robert Garden.)

12.1.1 The International Grid

In responding to Working Paper I, National Centers indicated the relative importance in the mathematics curriculum of their countries of each topic and behavior on comprehensive lists for each population. From these responses International Grids were constructed in which the elements common to most countries' curricula were included. These grids, later modified in response to further replies to Working Paper I, formed the basis for blueprints for drawing up test forms. (See Tables 12.1 and 12.2.) It should be noted that the grids reflect the importance with which National Centers regarded topics and behaviors for the Population A and B level students whether they taught at those or at earlier levels of schooling.

12.1.2 The Item Pools

Trial items were drawn from:

(i) items used in the First IEA Mathematics Survey
(ii) items sent by countries proposing to take part in the study
(iii) items constructed from tests, item banks and examinations sent by countries
(iv) items constructed for those cells in the grid for which items were not available from other sources.

In assembling the pools of items the major factors taken into account were:

(i) the need to sample the International Grid adequately with items as free as possible from cultural bias (e.g., currency)
(ii) the need to include sufficient "anchor" items from the First IEA Mathematics Survey to allow for comparisons to be made where this seemed appropriate
(iii) the need to include items on particular fields of interest of the International Mathematics Committee.

For example, the Committee saw it as essential that items including calculations with money be included for Population A, even though this would cause some translations problems in one or two countries. Items appropriate for hand calculators were also included.

The classification of items by behavior is regarded only as suggestive of levels of cognitive complexity. Some items will have a different classification from country to country depending on factors such as prior instruction.

TABLE 12.1

Population A: Importance For Instrument Construction Of Content Topics And Behavioral Categories

	Content Topics	Behavioral Categories[a]: Computation	Comprehension	Application	Analysis
000	Arithmetic				
001	Natural numbers and whole numbers	V	V	V	I
002	Common fractions	V	V	I	I
003	Decimal fractions	V	V	V	I
004	Ratio, proportion, percentage	V	V	I	I
005	Number theory	I	I	–	–
006	Powers and exponents	I	I	–	–
007	Other numeration systems	–	–	–	–
008	Square roots	I	I	–	–
009	Dimensional analysis	I	I	–	–
100	Algebra				
101	Integers	V	V	I	I
102	Rationals	I	I	I	I
103	Integer exponents	I	–	–	–
104	Formulas and algebraic expressions	I_s	I	I	I
105	Polynomials and rational expressions	I	I	–	–
106	Equations and inequations (linear only)	V	I_s	I	I
107	Relations and functions	I	I	I	$–_s$
108	Systems of linear equations	–	–	–	–
109	Finite systems	–	–	–	–
110	Finite sets	I	I	I	–
111	Flowcharts and programming	–	–	–	–
112	Real numbers	–	–	–	–
200	Geometry				
201	Classification of plane figures	I	V	I	I
202	Properties of plane figures	I	V	I	I_s
203	Congruence of plane figures	I	I	I	I
204	Similarity of plane figures	I	I	I	I_s
205	Geometric constructions	I	I	I	$–_s$
206	Pythagorean triangles	I_s	I_s	I_s	–
207	Coordinates	I_s	I_s	I_s	I
208	Simple deductions	I_s	I	I	I_s

[a] The following rating scale has been used: V = very important; I = important; I_s = important for some countries. A dash (–) indicates that the topic was not considered important enough to warrant trial items being found or constructed.

TABLE 12.1 (Continued)

	Content Topics	Behavioral Categories[a]: Computation	Comprehension	Application	Analysis
200	Geometry (Cont'd.)				
209	Informal transformations in geometry. . . .	I	I	I	–
210	Relationships between lines and planes in space	–	–	–	–
211	Solids (symmetry properties)	I	I	I	–
212	Spatial visualization and representation . .	$-_s$	I_s	I_s	–
213	Orientation (spatial).	–	I_s	$-_s$	–
214	Decomposition of figures.	–	$-_s$	–	–
215	Transformational geometry	I_s	I_s	I_s	
300	Probability and statistics				
301	Data collection	I	I	I	–
302	Organization of data.	I_s	I	I	I
303	Representation of data	I	I	I	I_s
304	Interpretation of data (mean, median, mode)	I	I	I	$-_s$
305	Combinatorics	–	–	–	–
306	Outcomes, sample spaces and events. . . .	I_s	–	–	–
307	Counting of sets, (P(A B), P(A B), independent events	–	–	–	–
308	Mutually exclusive events	–	–	–	–
309	Complementary events.	–	–	–	–
400	Measurement				
401	Standard units of measure.	V	V	V	–
402	Estimation.	I	I	I	–
403	Approximation	I	I	I	–
404	Determination of measures: areas, volumes, etc.	V	V	I	I

TABLE 12.2

Population B: Importance For Instrument Construction Of Content Topics and Behavioral Categories

Content Topics	Behavioral Categories[a] Computation	Comprehension	Application	Analysis
1 Sets, relations and functions				
1.1 Set notation	I	I	-	-
1.2 Set operations (e.g., union, inclusion)	I	I	-	-
1.3 Relations	-	-	-	-
1.4 Functions	V	V	V	I
1.5 Infinite sets, cardinality and cardinal algebra (rationals and reals)	-	-	-	-
2 Number systems				
2.1 Common laws for number systems	I	I	I	-
2.2 Natural numbers	I	I	I	I
2.3 Decimals	I	I	I	I
2.4 Real numbers	I	I	I	-
2.5 Complex numbers	V	I	I	I
3 Algebra				
3.1 Polynomials (over)	V	V	V	I
3.2 Quotients of polynomials	I	I	I	-
3.3 Roots and radicals	V	V	I	-
3.4 Equations and inequalities	V	V	V	I
3.5 System of equations and inequalities	V	V	V	I
3.6 Matrices	I	I	I	I
3.7 Groups, rings and fields	I_s	I_s	I_s	I_s
4 Geometry[b]				
4.1 Euclidean (synthetic) geometry	I	I	-	-
4.2 Affine and projective geometry in the plane	-	-	-	-
4.3 Analytic (coordinate) geometry in the plane	I	I	V	I
4.4 Three-dimensional coordinate geometry	-	-	-	-
4.5 Vector methods	I	I	I	I
4.6 Trigonometry	V	V	V	I

[a]The following rating scale has been used: V = very important; I = important; I_s = important for some countries. A dash (-) indicates that the topic was not considered important enough to warrant trial items being found or constructed.

[b]This section is currently being modified to take into account curricular emphases in some European countries.

TABLE 12.2 (Continued)

	Content Topics	Computation	Comprehension	Application	Analysis
		Behavioral Categories[a]			
4	Geometry (Cont'd.)				
	4.7 Finite geometries	–	–	–	–
	4.8 Elements of topology	–	–	–	–
5	Analysis				
	5.1 Elementary functions	V	V	V	V
	5.2 Properties of functions	V	V	V	I
	5.3 Limits and continuity	I	I	I	–
	5.4 Differentiation	V	V	I	I
	5.5 Applications of the derivative	V	V	V	I
	5.6 Integration	V	V	V	I
	5.7 Techniques of integration	V	V	I	I
	5.8 Applications of integration.	V	V	V	I
	5.9 Differential equations	I_s	I_s	I_s	I_s
	5.10 Sequences and series of functions	–	–	–	–
6	Probability and statistics				
	6.1 Probability	V	V	I	–
	6.2 Statistics	I	I	I	–
	6.3 Distributions	I	I	I	–
	6.4 Statistical inference	I_s	I_s	–	–
	6.5 Bivariate statistics	–	–	–	–
7	Finite mathematics				
	7.1 Combinatorics	I	I	I	–
8	Computer science	I_s	I_s	I	–
9	Logic	–	–	–	–

12.1.3 Item Trials

From the assembled pools, 10 collections of 40 items each at Population A level and 13 collections of 20 items each at Population B level were field tested in several countries. In one country a further 2 collections of 40 items at Population A level were field tested. The Committee examined the item statistics and comments by National Centers resulting from this field testing, paying particular attention to mathematical appropriateness, psychometric properties and cultural bias. Careful interpretation was necessary as comparatively few countries had taken part in the trials and some had administered the items before the end of the school year or with students at higher grade level than Population A. Inevitably, some measure of compromise was undertaken in meeting the needs of widely differing curriculums and instructional practices.

Further items were constructed for cells in the International Grids to supplement the pools and replace deleted items. The International Grids were modified slightly as a result of additional replies received from countries in response to Working Paper I. The next round of trial testing included 12 collections of 40 items each at Population A level and 15 collections of 20 items each at Population B level. The IMC has selected those items which appear to be most appropriate for the purposes of the study.

At the IEA General Assembly in Paris, September 1979, a petition to the IMC was submitted to the IMC in behalf of several countries, including French and Flemish Belgium, France and Luxembourg concerning the lack of items which accurately reflected the distinctive nature of their curricula. The shortfall, it was noted, was particularly in geometry. As the result of several meetings with various members of the national committees, and further field testing in some countries, it was agreed to further modify the international grid and to increment the international tests.

Population A

Modification of Grid:

Add to geometry, Section 200, the category 215: Transformational Geometry. These are to be I_s cells at behavioral levels I - III.

Items to be Added to Test

Eight items were produced, 2 for each of the four rotated forms, as follows:

Content	Behavioral Level	Number of Items
103	II	1
205.3	III	1
207	II	2
215	I	1
215	II	2
215	III	1
	Total	8

Population B

Modification of Grid

Categories will have to be added or present categories slightly redefined, especially in Section 4, Geometry, to take account of the special characteristics of the curriculum in the countries requesting this modification.

Items to be Added to Test

Sixteen items were produced, two for each of the eight rotated forms. Since the grid is not yet revised to take the new subject matter into account, a precise tabulation is not available. The items are primarily in geometry.

12.1.4 Anchor Items

An objective of the Second Study of interest to many of the countries which participated in the First Study is that of examining differences in student outcomes since 1964, the date of the First Study. Evidence of such differences will be gained through the use of anchor items, that is, items used in both the First and Second Study.

For the cognitive instruments, the achor items raised special problem areas. The selection of items from the First Mathematics Study pool was difficult due to ambiguity in how the First Study item measured the stated objectives. It was also important to select items which would help address issues in mathematics education such as what changes in computational ability have occurred since the First Study.

It was regarded as desirable to preserve the original wording and format of the items as they appeared in the First Study. However, it was also recognized that preserving the item as it appeared in the First Study does not guarantee meaningful comparison. What is desired in measuring changes in student performances over a span of years is a measure of the same behaviors at the two points of time. However, schools and curriculums change. The instructional content changes. Hence, the IMC strove to take into account changes in usage, convention, terminology, and other factors, in attempts to provide measures of the same construct.

The result of these efforts is three categories of anchor items. One category is that of items which have remained identical from the first study. The second category is a set of items which have changed in terms of format, e.g., open-ended to multiple choice. The third category is a "middle group" of items which have undergone only minor changes in wording, and in the judgment of the IMC, sample for all practical purposes the same constructs in the two studies.

A summary of the anchor items appearing in the cognitive tests for the cross-sectional study is given in Appendix C.

12.2 Responsiveness to growth during academic year (Population A)

The longitudinal aspect of the Study requires as a dependent measure growth in those aspects of mathematics being examined in the classroom process questionnaires. This is indeed a stringent requirement, for it goes beyond the validity issue to narrowing the curricular emphasis to a common period of time cross-nationally (that is, between the administration of the pretest and posttest.)

The initial attempt of the International Mathematics Committee to produce such a measure was the proposed common core of items to serve as the growth measures (5 scales of 8 items each). The results of the pilot testing for Population A are summarized in Tables 12.3 (4 pages) and 12.4 (5 pages). The median difficulty of the items is, for most countries, within acceptable ranges for the end of the year. Japan and Hong Kong are notable exceptions in that the items are very easy for those countries.

Countries were also asked to rate the items according to when they are taught in the school year (important information for the classroom process questionnaire) and the appropriateness of the items to the target populations. Table 12.5 (3 pages) summarizes this information. Clearly, adjustments in the core tests are needed so that they better fit the curriculums of the countries. Much of the remainder of this section is addressed to this issue.

Two other analyses were done on the pilot data to help shed light upon the current structure of the Population A test. Table 12.6 shows scatterplots of between country correlations for various topics. New Zealand is chosen as the vertical scale since more of the items were pilot tested there than in other countries. Table 12.7 illustrates the relationship between the core and rotated form subtests by country median difficulty.

Table 12.8 depicts the between school differences on the core items for the United State and Canada. With the exception of a few items that were tried two different times with two different samples (in Japan and New Zealand), this represents the only information available to the IMC about within country variation at levels other than between students. The between school variation found in the U.S. and Canada may not be representative of what will be found in the majority of countries. If not, then this table may be disregarded. But if so, this table may contain important information for those who are structuring the cognitive instruments.

The majority of the discussion of targeting the difficulty of the cognitive test has had as a context the problem of between country differences. Concerns have been expressed that the cognitive items may be too easy for one or two of the countries and too difficult for a few of the others. Seldom has the notion of looking at within country differences entered these discussions. But, if there are substantial within country differences at the school or classroom level an overall adjustment may not be the most efficient. Since the focus of the study is on the classroom, an aim should be to eliminate ceiling and floor effects at those levels. It may be necessary, therefore, to get estimates of within country differences at, at least, the school level, so that a cognitive instrument can be structured that is varied enough to be applicable to most of the classrooms within a country.

While the response of the national committees to our proposed items was all in all encouraging, it is clear obtaining sufficient variation in the measures to yield interesting analysis will require scales which are more responsive to the individual curriculums. What follows is a description of suggested structural changes for the cognitive test and procedures for allowing countries to construct the best possible cognitive test for their purposes.

TABLE 12.3. Item Difficulties by Country for Population A - Cognitive Test (June 1979 version)

CORE TEST

TOPIC

FRACTIONS

	1	2	3	4	5	6	7	8
Belgium (Fl)	67		75				85	70
Belgium (Fr)		87	31				50	46
Can/USA		79		74	60	63		
France				57		38		
Hong Kong			85	97				
Hungary		63			60	57		
Ireland		57						
Japan		93		95	83	90	93	65
Netherlands	64	76	46	85	62	69	91	54
N. Zealand	54	71	53	62				
Scotland				77		72		
Sweden								
Median	64	76	53	77	61	66	88	60

RATIO, PROPORTION, PERCENT

	1	2	3	4	5	6	7	8
Belgium (Fl)	56	71	41		71		81	80
Belgium (Fr)	39	51	10	38		16		
Can/USA				54	73	40		72
France								
Hong Kong				77			95	
Hungary					35	37		61
Ireland								
Japan	57	61	57		73	84	84	89
Netherlands	49	52	28	61	54	64	73	77
N. Zealand								
Scotland								
Sweden								
Median	53	57	34	58	71	40	83	77

ALGEBRA

	1	2	3	4	5	6	7	8
Belgium (Fl)	73					56	40	57
Belgium (Fr)		91					42	
Can/USA	49	27	38		33			31
France	39	61	62		27			37
Hong Kong			88	76				
Hungary				23	40	33		
Ireland		61			52			42
Japan				81	84	70	63	
Netherlands	64	60	79	57	55	37	29	46
N. Zealand		36	63	21		47		54
Scotland		63			52			40
Sweden								
Median	57	61	63	57	52	47	41	42

GEOMETRY

	1	2	3	4	5	6	7	8
Belgium (Fl)	78	46		36		40	54	
Belgium (Fr)	78	42		11				14
Can/USA					44		47	22
France							53	
Hong Kong		24	96	41				63
Hungary			51		59			
Ireland					52			
Japan	94		96		94			84
Netherlands	73	46	79	27	54	65	66	23
N. Zealand		31	65	29		77	67	35
Scotland								
Sweden								
Median	78	42	79	29	54	65	54	29

MEASUREMENT

	1	2	3	4	5	6	7	8
Belgium (Fl)	86		64	47	26	70	33	
Belgium (Fr)		32		42	81	83	27	
Can/USA	52	33						
France	74							
Hong Kong		64	80	67	97	72	43	89
Hungary								23
Ireland		51						
Japan		80	83					93
Netherlands	92	49	59	42	81	69	37	74
N. Zealand	64	42	51	37	60	65	40	43
Scotland	83	58						
Sweden								
Median	79	50	64	42	81	70	37	74

TABLE 12.3 - Continued

ROTATED FORMS

TOPIC

RATIO, PROPORTION, PERCENT

	1	2	3	4	5
(Fl)	72		62		90
(Fr)	67	70			
Can/USA		56		33	
France		39		38	
Hong Kong	87			91	96
Hungary					
Ireland		58			
Japan				86	82
Netherlands				63	67
N. Zealand	49	50	35	38	59
Scotland		69		60	
Sweden	65	56	13	48	65
Median	67	56	35	54	75

MEASUREMENT

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
(Fl)	83	73	33	46	87	76	40	89		30	55	60			46
(Fr)		73	17	49		71	16			13		38		22	23
Can/USA	85				82				27					28	
France									31						
Hong Kong		80		42			38	51					74	66	
Hungary	49				15				52				4		
Ireland												36		25	25
Japan	96		72		93	79		96	79	68			93	75	
Netherlands						46	39								
N. Zealand	84	66		23	38		21	28	56		52	27	50	31	35
Scotland															
Sweden	82	40	9	50	34	17	25	25	82	77	16	15	68	32	13
Median	84	73	25	46	60	71	32	51	54	49	52	36	68	31	25

GEOMETRY

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
(Fl)					26	36				49				37		53	18	47
(Fr)				15		10			65	20	51		39				14	
Can/USA	29		11	21	30		31	37	39		47	29	35	47		35		
France	19		12					35				28				40		
Hong Kong		89	68				72	64	70		77	51			21		42	
Hungary	24	38		35	12									15	23			
Ireland				32	8		39		61	38	41		43				24	
Japan	79	96	86	78	78	70	92	93	77		85	65		84	91			
Netherlands																		
N. Zealand		63	27	37			56	61	52	56	55	32	37	40	34	22		46
Scotland	28		45		49			80				52	56			14	56	
Sweden																		
Median	28	76	36	34	28	36	56	63	63	43	53	42	39	40	30	35	24	46

TABLE 12.3 - Continued

ROTATED FORMS

TOPIC

ALGEBRA

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Belgium (Fl)		88			76			91	88	73		67			66	93	56	80	43			90	88			47	82	71
Belgium (Fr)		71						56		55	67	59	92			65		62		31	63	45	87	54				
Can/USA	34		48	20	60	61	47				37		38		47				16	75				20	51			64
France	53		63	22		79	67								17				29						21			
Hong Kong			94						82				82	90				85		93	75				76	61	92	
Hungary					35	81	49			91	48			69										31				53
Ireland											38	32	61							44			76	35				
Japan		87	94		56	88	81	91	94		88		93	83		94				61		95		81		73	86	88
Netherlands																												
N. Zealand	46		62	29	43				54	51	51	30	33	35	18		46	62	14	53	52		78	24		41	66	59
Scotland			78			83	49								42													
Sweden																												
Median	46	87	71	22	56	81	49	91	85	64	50	46	72	76	42	93	51	71	23	57	63	90	83	33	51	54	84	64

FRACTIONS

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Belgium (Fl)			69	86							74	37	75	67			61	52	85				
Belgium (Fr)			49	92	91		60	93			70	56	54						79			57	
Can/USA	92	61			48			69		61				56	86	47		40		76	58	48	45
France	70	33			33		44			34					91	49				85	38		65
Hong Kong								95	72	92	93	60	77	42		75	54			95	74	79	83
Hungary	77	66					56		67									29					
Ireland					56	67	66	56									22					40	
Japan	96	90	71	77				95	87	84				83		84		63		88	73	72	86
Netherlands																							
N. Zealand					71	54	70	46	43	52	62	58	26	20	67	49	27	14	69	70	19	37	65
Scotland	97	63			58		75			70						71				83	35		79
Sweden																							
Median	92	63	69	86	57	61	63	81	70	65	72	57	65	56	86	60	41	40	79	84	48	52	72

TABLE 12.3 - Continued

REMAINING ITEMS ON ROTATED FORMS

TOPIC

WHOLE NUMBERS

	1	2	3	4	5	6	7	8	9	10
Belgium (Fl)	32			91				75	92	64
Belgium (Fr)		86			58	82			86	
Can/USA	29	82	92		86	72		58		41
France			91					76		56
Hong Kong		89		97		95	79			
Hungary	32						62			
Ireland		84			60	73			60	
Japan	76	93		93		84	81			
Netherlands	90		79		66					
N. Zealand	23	86	86	73	81	81	51	66	81	71
Scotland								83		
Sweden	74	78	67	10	43	77	74	62		
Median	32	86	86	91	63	81	74	71	84	60

NUMBERS (OTHER)

	1	2	3	4	5	6	7
Belgium (Fl)			61	76		73	
Belgium (Fr)		57	16		79		
Can/USA	66	57		42	74		
France	61			66			
Hong Kong		81	36		86	75	61
Hungary	28						21
Ireland		63			75		
Japan	87	74			72	96	67
Netherlands			55			63	59
N. Zealand		58	17	53		26	21
Scotland				48	77		
Sweden	35	30	45	52	10	32	16
Median	61	58	41	53	75	68	40

PROBABILITY AND STATISTICS

	1	2	3	4	5	6	7	8	9
Belgium (Fl)			65		77		62		
Belgium (Fr)				7		30	34		
Can/USA	30	27	72	40	71	39		36	67
France	23	25						32	65
Hong Kong	40	40							87
Hungary			35	19	33	2		16	
Ireland				28		6	49		
Japan	49	70	92	83	83	81		92	
Netherlands					36	37		86	
N. Zealand	36	28	67	44	64	47	61		
Scotland	49	33						71	
Sweden	45	67	48	74	35	45	60	76	54
Median	40	33	66	40	64	37	60	71	67

Table 12.4. MEDIAN ITEM DIFFICULTIES & RANGES, POPULATION A
June 1979 Version of Test

(FRACTIONS)

	CORE			ROTATED FORMS		
	# of items	Med	Range	# of items	Med	Range
Belgium Flemish	4	72	67-85	9	69	37-85
Belgium French	4	48	31-87	11	70	49-93
Canada/U.S.	4	68	60-79	14	57	40-92
France	2	47	38-57	10	47	33-90
Hong Kong	2	91	85-97	12	76	54-95
Hungary	3	60	57-63	4	67	42-77
Ireland	1	57		5	56	22-66
Japan	6	91	65-95	14	84	63-96
New Zealand	4	58	53-71	19	52	14-71
Scotland	2	75	72-77	9	71	35-97
Total Possible	8			23		

Table 12.4 (Cont'd.)

(PERCENT, RATIO, PROPORTION)

	CORE			ROTATED FORMS		
	# of items	Med	Range	# of items	Med	Range
Belgium Flemish	6	71	41-80	3	72	62-90
Belgium French	5	38	10-51	2	78	67-90
Canada/U.S.	4	63	40-72	2	44	33-56
France	-			2	38	38-39
Hong Kong	2	86	77-95	3	91	87-96
Hungary	3	37	35-61	-		
Ireland	1	56	-	1	58	-
Japan	8	67	23-88	2	84	82-86
New Zealand	5	45	39-67	5	49	35-59
Scotland	-			2	65	60-69
Total Possible	8			5		

Table 12.4 (Cont'd.)

(ALGEBRA)

	CORE			ROTATED FORMS		
	# of items	Med	Range	# of items	Med	Range
Belgium Flemish	4	57	40-73	15	80	47-93
Belgium French	2	66	42-91	13	65	31-92
Canada/U.S.	6	32	16-49	13	47	20-75
France	5	39	27-62	7	53	17-79
Hong Kong	2	82	76-88	10	83	61-94
Hungary	3	33	23-40	8	48	31-91
Ireland	2	52	42-61	6	41	32-76
Japan	4	75	63-84	17	88	61-95
New Zealand	5	47	21-63	21	46	14-78
Scotland	3	52	40-63	4	63	42-83
Total Possible	8			28		

Table 12.4 (Cont'd.)

(GEOMETRY)

	CORE			ROTATED FORMS		
	# of items	Med	Range	# of items	Med	Range
Belgium Flemish	5	46	36-78	7	37	18-53
Belgium French	4	28	11-78	7	20	16-65
Canada/U.S.	3	44	22-47	12	31	11-47
France	1	53	-	6	31	12-40
Hong Kong	4	52	24-96	9	68	21-89
Hungary	2	55	51-59	6	23	12-38
Ireland	1	52	-	8	32	8-61
Japan	4	94	83-96	13	84	65-96
New Zealand	6	70	29-77	14	43	22-63
Scotland	-			8	51	14-80
Total Possible	8			18		

Table 12.4 (Cont'd.)

(MEASUREMENT)

	CORE			ROTATED FORMS		
	# of items	Med	Range	# of items	Med	Range
Belgium Flemish	6	55	26-86	12	57	30-89
Belgium French	5	42	27-83	9	23	13-73
Canada/U.S.	2	43	33-52	4	55	27-85
France	1	74	-	1	31	-
Hong King	7	72	43-97	6	58	38-80
Hungary	1	23	-	4	32	4-52
Ireland	1	51	-	3	25	25-36
Japan	3	83	80-93	9	79	68-96
New Zealand	8	47	33-64	12	43	21-84
Scotland	2	70	58-83	-		
Total Possible	8			15		

Table 12.5

SUMMARY OF QUALITATIVE JUDGMENT OF COUNTRIES
(Population A)

Core Test

1. Fractions

	When Taught[1] (% of Items) P	T	S	Appropriateness[1]
Belgium (Fl)	50	50	0	1.5
Japan	100	0	0	1.1
US	0	100	0	2.0
Canada	0	100	0	2.0
Australia	38	*62	0	1.9
Spain	75	25	0	1.6
Chile	88	12	0	2.0

X=1.72 SD=.35

2. Percent, Ratio and Proportion

	P	T	S	Appropriateness
Belgium (Fl)	50	--	50	.5
Japan	100	--	0	1.0
US	--	100	0	2.0
Canada	--	100	0	2.0
Australia	12	88*	0	1.9
Spain	100	--	0	1.4
Chile	--	100	0	2.0

X=1.54 SD=.59

3. Algebra

	P	T	S	Appropriateness
Belgium (Fl)	--	100	0	2.0
Japan	--	75	25	1.1
US	--	100	0	1.1
Canada	--	88	12	1.6
Australia	--	*100	0	2.0
Spain	--	88	12	1.6
Chile	--	12	88	2.0

X=1.63 SD=.40

[1] For key, see bottom next page.

Table 12.5 (Cont'd.)

4. Geometry

	When Taught[1] (% of Items)				Appropriateness[1]
	P	T	S	N	
Belgium (Fl)	--	--	100		0
Japan	12	38	50		1.0
US	--	88*	*88	12	.6
Canada	--	25	.75		.4
Australia	12	62	25		1.5
Spain	37	62	0		.6
Chile	--	62	37		1.6
				$\bar{X}$=.81	SD=.58

5. Measurement

	P	T	S	N	Appropriateness
Belgium (Fl)	100	--			.12
Japan	87	13			1.3
US	--	100			2.0
Canada	13	87			2.0
Australia	13	87			1.7
Spain	100	--			1.6
Chile	25	75			2.0
				$\bar{X}$=1.5	SD=.68

Grand Mean = 1.3
Grand Standard Deviation = .67

Key

When taught

- P Prior to this year (that is, taught at a lower level than 8th grade but not specifically taught in 8th grade).
- T Taught this year (that is, either introduced in 8th grade for the first time or reviewed or retaught in 8th grade).
- S Subsequent years (that is, not taught up to or including 8th grade but not taught in 9th or higher grades).
- N Not in curriculum (not in curriculum at any grade level).

Appropriateness

- 0 inappropriate
- 1 acceptable
- 2 highly appropriate

Note: If the item tests knowledge or skills taught up to or at the level of the target population and is likely to be moderately difficult or easy at the end of the year then it is judged to be appropriate.

Table 12.5 (Cont'd)

SUMMARY BY COUNTRY OF THOSE SUBTESTS THAT FIT A COUNTRY'S CURRICULUM

Country	Fractions	Percent, Ratio Proportion	Algebra	Geometry	Measurement
	TOPIC				
Belgium Flemish (2)	X	--	X	--	--
Japan (1)	--	--	X	--	--
U.S. (5)	X	X	X	X	X
Canada (4)	X	X	X	--	X
Australia (5)	X	X	X	X	X
Spain (2)	--	--	X	X	--
Chile (3)	--	X	--	X	X
	(4)	(4)	(6)	(4)	(4)

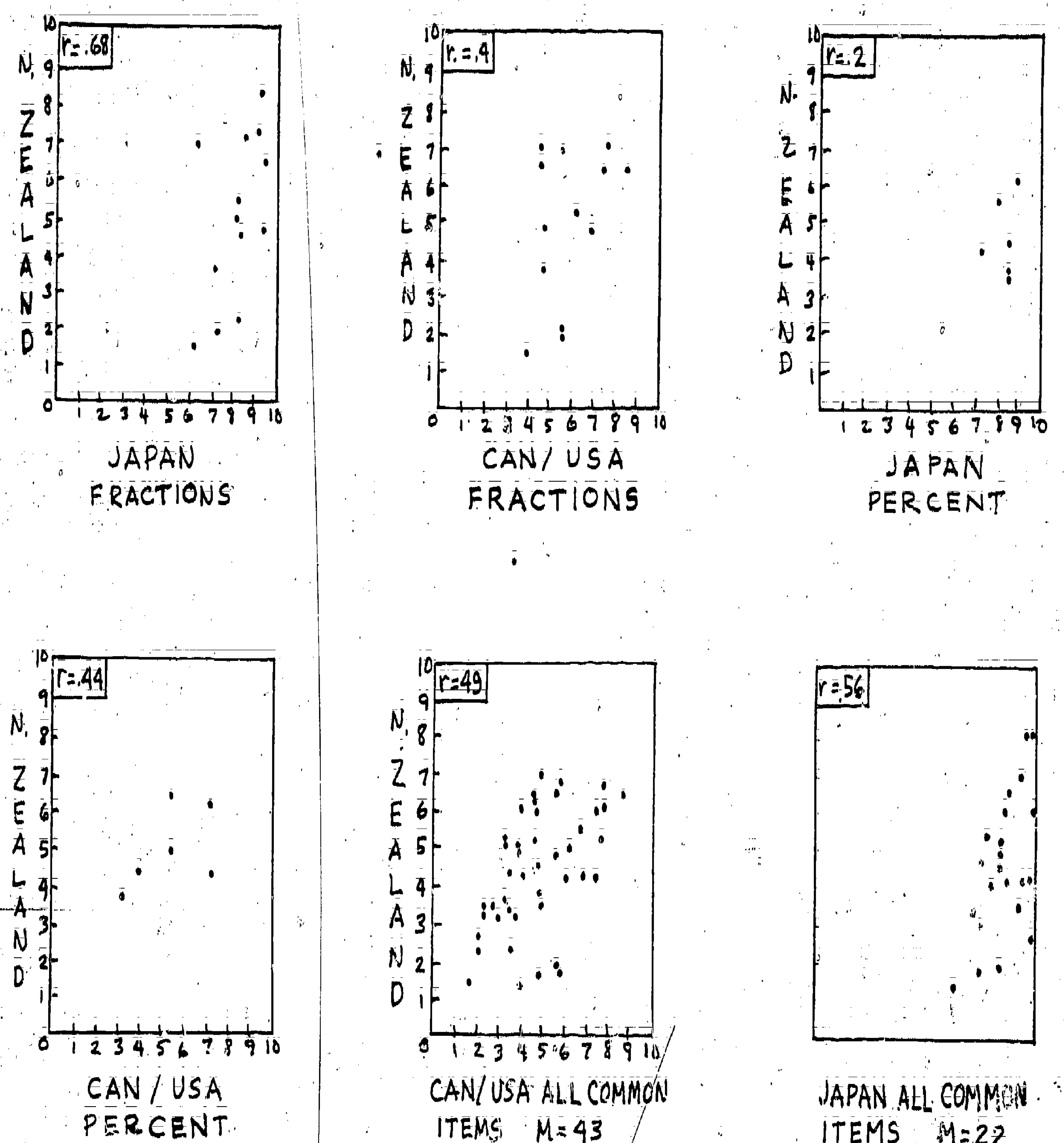


TABLE 12.6

SCATTERPLOTS OF ITEM DIFFICULTIES (BETWEEN COUNTRY CORRELATIONS)

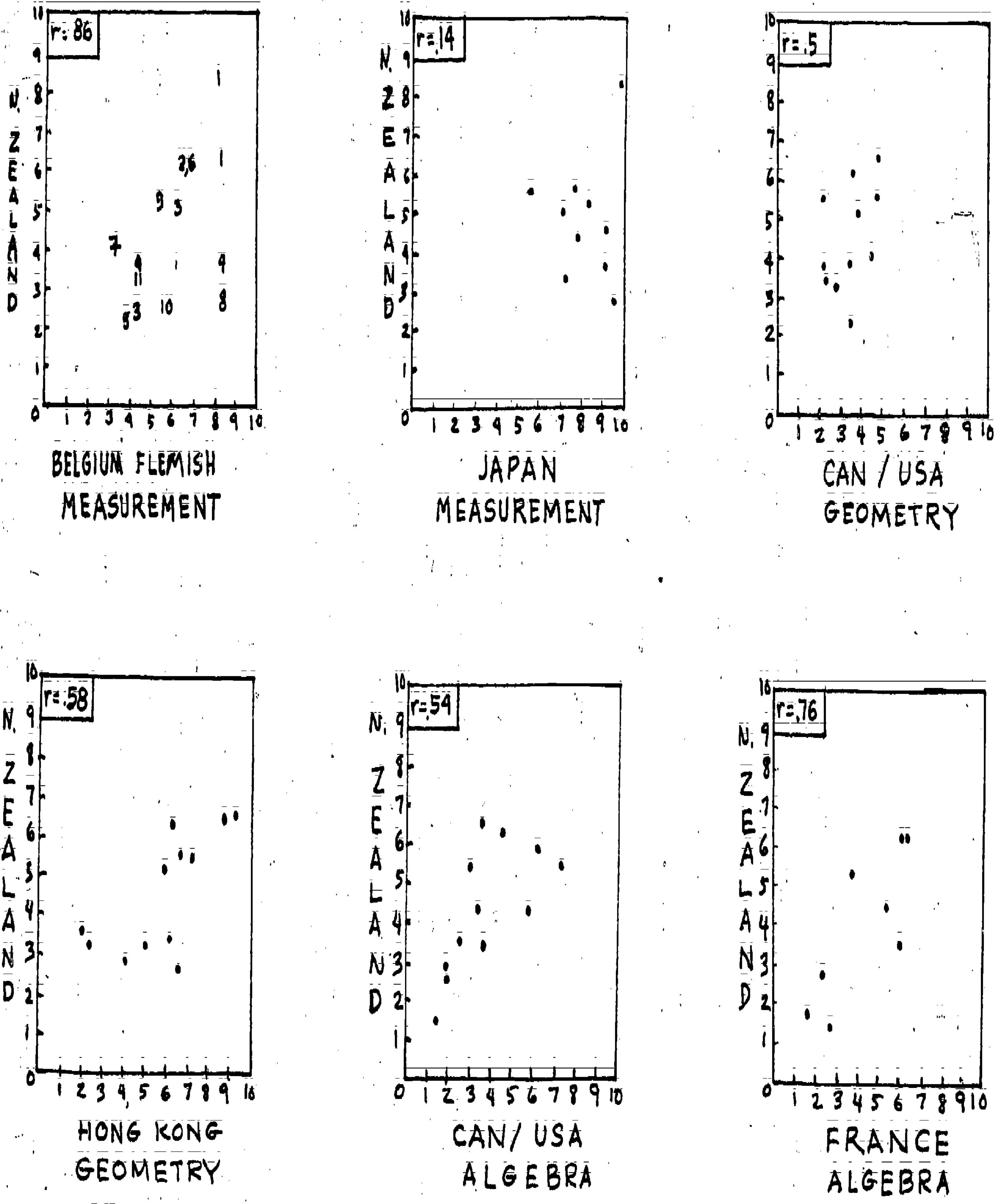

r: .86
N. ZEALAND
BELGIUM FLEMISH
MEASUREMENT
r=.14
N. ZEALAND
JAPAN
MEASUREMENT
r=.5
CAN / USA
GEOMETRY
r=.58
N. ZEALAND
HONG KONG
GEOMETRY
r=.54
N. ZEALAND
CAN/ USA
ALGEBRA
r=.76
N. ZEALAND
FRANCE
ALGEBRA

TABLE 12.6

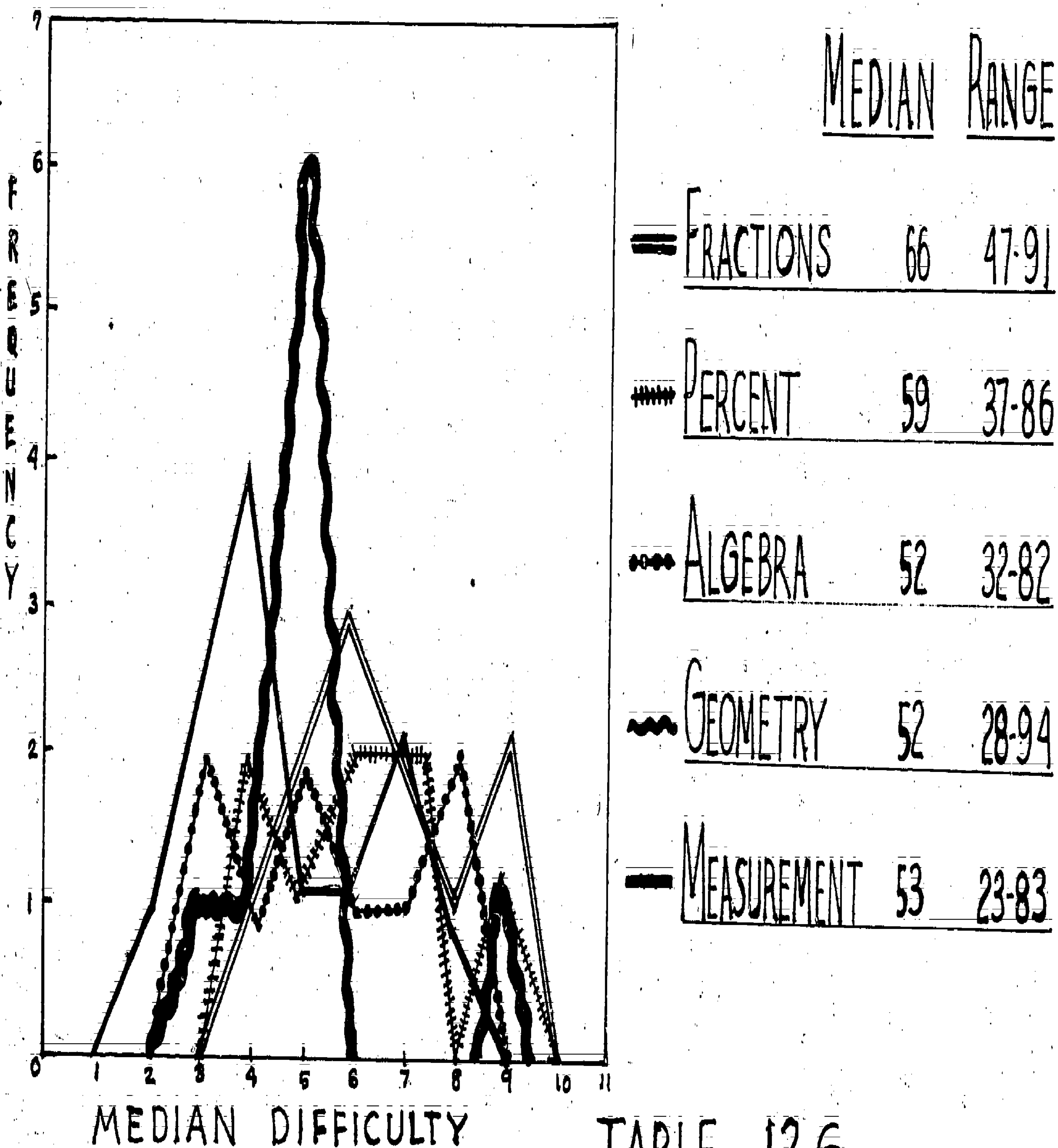
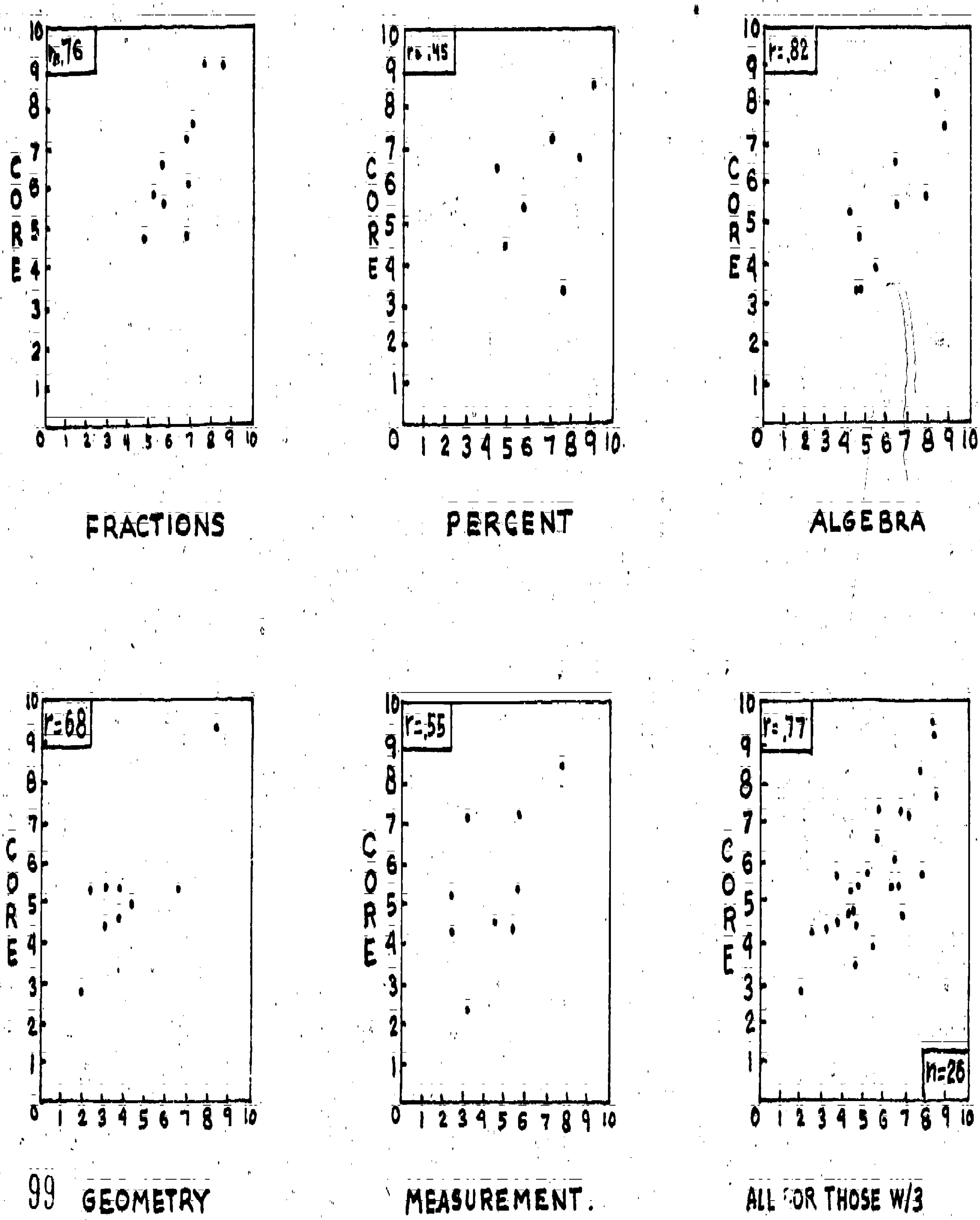


TABLE 12.6

FREQUENCY DISTRIBUTION MEDIAN ITEM DIFFICULTY (COUNTRY)



98

TABLE 12.7

CORRELATIONS BETWEEN CORE AND ROTATED FORMS ([illegible]OUNTRY MEDIANS)

TABLE 12.8. BETWEEN SCHOOL VARIATION (Item Proportions & Subtests)

CANADA/USA

N = 15 Schools N = 381 Students

FRACTIONS

Item	Item Statistics		Between School Statistics			
	$\bar{X}$	SD	Lo	Hi	SD	ETA^2
1.	.54	.49	.34	1.0	.25	.17
2.	.72	.45	.31	1.0	.21	.12
3.	.69	.46	.15	1.0	.24	.21
4.	.72	.45	.31	1.0	.18	.10
5.	.45	.49	.07	1.0	.26	.16
6.	.79	.40	.38	1.0	.18	.14
7.	.38	.49	0	.8	.22	.10
Totals	4.3	1.9	1.5	6.8	1.4	.34

PERCENT, RATIO, PROPORTION

Item	Item Statistics		Between School Statistics			
	$\bar{X}$	SD	Lo	Hi	SD	ETA^2
1.	.34	.47	0	1.0	.29	.21
2.	.60	.49	.31	1.0	.21	.11
3.	.25	.43	0	.6	.16	.07
4	.46	.49	.28	1.0	.21	.08
Totals	1.3	1.0	.83	3.6	.8	.24

MEASUREMENT

Item	Item Statistics		Between School Statistics			
	$\bar{X}$	SD	Lo	Hi	SD	ETA^2
1.	.74	.43	.54	1.0	.16	.11
2.	.45	.49	.19	.9	.20	.10
3.	.18	.39	.02	.4	.14	.10
4.	.62	.48	.22	.93	.25	.23
5.	.18	.39	0	.6	.13	.03
Totals	2.2	1.2	1.1	3.4	.69	.24

GEOMETRY

Item	Item Statistics		Between School Statistics			
	$\bar{X}$	SD	Lo	Hi	SD	ETA^2
1.	.70	.46	.46	1.0	.18	.11
2.	.41	.49	.14	.89	.22	.12
3.	.25	.43	.16	.5	.16	.04
4.	.17	.38	0	.8	.21	.12
Totals	1.5	.97	1.0	3.0	.49	.14

ALGEGRA

Item	Item Statistics		Between School Statistics			
	$\bar{X}$	SD	Lo	Hi	SD	ETA^2
1.	.41	.49	.07	1.0	.25	.12
2.	.57	.49	.23	.9	.16	.06
3.	.70	.46	.38	1.0	.21	.15
4.	.49	.50	.08	1.0	.26	.17
5.	.61	.49	.08	1.0	.26	.19
Totals	2.7	1.3	1.2	4.9	.87	.32

12.3 Proposed structure of the Cognitive Test (Population A: Longitudinal)

Table 12.9 shows the present structure of the Population A test. This structure differs from what is given in Bulletin 3 (pages 28-29) in that the present design calls from stratified random assignment of items across the rotation forms where the strata are subject matter classifications from Table 12.1. The anticipated requirements for testing time for the present structure are as reported in Bulletin 3, page 34.

A proposed revised structure is given in Table 12.10. The new features are:

(i) The required common core for all countries has been reduced from 8 items per scale to 4 items. This set of items, the International Core, is comprised of those which are taught to the target class during the school year (and therefore, presumably, hold promise of reflecting growth).

(ii) The remaining 4 items in the core are chosen by the countries to best reflect their curriculum or to otherwise respond to national interests, specifically as they relate to classroom process. This set of items is the National Core.

(iii) The present rotated forms (4 with 32 items each) may be augmented by the countries from the international pool of items.

(iv) In order to provide a comprehensive growth measure, all items should be administered in the pretest and posttest.

"Existing" (A/ 124)

Topics	Core Test # of Items	Rotated Forms	# of items
Fractions	8	Rotated Form 1	32
Percent	8	Form 2	32
Algebra	8	Form 3	32
Geometry	8	Form 4	32
Measurement	8		
Subtotals	40		128

Table 12.9 TOTAL ITEMS = 168

Sequence of Testing: Core - Pre and Post; Rotated Forms - Post

"Proposed"

Topics	International Core	Floating or National Core	Rotated Forms
Fractions	4	varied on	#of forms and
Percent	4	content and	length of forms
Algebra	4	difficulty	depends on coun-
Geometry	4	depending on	try's choice of
Measurement	4	country choices	"sampling" plan
Table 12.10 Subtotals	20	20	120 or more

Sequence of Testing: International Core and other items given both Pre and Post.

12.4 Demands of student time vs. options for analysis

From the point of view of the researcher, it is desirable to have as many students as possible respond to as many items as possible. If all students repond to all items, items are said to be completely crossed with students. The completely crossed design permits maximal flexibility for analyses: between students, between classes, between schools. However, this option would require an estimated 4 hours for an international pool of 160 items. For the vast majority of countries, this is undoubtedly an unrealistic expectation.

From the point of view of minimal intrusion into the school (minimal administration time), the preferred test design is that of matrix sampling, one version of which is rotated forms. This is the design proposed for Population B for two reasons: (1) a growth measure is not required, except as a national option (2) for many countries, testing time appears to be more difficult to obtain at the advanced grade levels. The analyses which such a design permits are those at the classroom level.

Under the rotating form scheme, each item in the pool is taken by a random sample of students. For example, if there are four rotated forms, each item is taken by ¼ of any sampled class.

The proposed structure for Population A is an attempt to acknowledge the many constraints upon the instrument as well as the demands of the objectives of the Study.

The remainder of this section presents a proposed procedure for designing the cognitive test for the longitudinal study, Population A, in the light of comments from the various countries and the empirical information on the items which has been received to date.

The proposals in summary are as follows:

1. The main cognitive test forms should be constructed as stratified random samples of items from the whole pool.

2. The number of forms taken by a student--that is, the density of item sampling--should be a national option.
3. There should be a national core of items taken by all students in a given country, but this core will differ over countries.
4. There should be an international core taken by all students in all countries.
5. The pre-test instrumentation for the longitudinal study should be a complete copy of the post-test instrumentation, with all items being used both times.

These proposals and the discussion points which follow them are aimed primarily at the design of the longitudinal, Population-A part of the study, although they are perhaps applicable to the other parts as well. In a number of places, one is led to the conclusion that really good design would require preliminary estimates of difficulty levels of items and of difficulty differences over a school year. Ideally, the overall study plan should leave time to obtain such data, analyze them, and feed the results into a final design. The plan also places considerable responsibilities on National Centers in terms of coding, administration, reporting of data and so forth.

PROPOSALS

Proposal 1. Stratified random instruments

It is clear that in most countries, it will not be practical to have all sampled students respond to all the cognitive items. Some kind of matrix sampling is required. It has been suggested that in certain parts of the Study the test forms be defined along content lines, that is, there would be a form for ratio, one for geometry, etc. However, for a number of reasons, stratified random assignment should be used. The major substantive categories of the item pool would define the strata. Where possible, further blocking would be made on estimated or approximate item difficulty. Each test form would then contain a stratified random sample of the pool. That is, the forms would be made up by stratified randomization.

One immediate effect is that the sampling errors of the estimates of item difficulty within a content area become less correlated. The determination of the distribution of subtest scores is potentially much more accurate.

Stratified random instrumentation is essential for relational studies. No core test can have sufficient content variability to be a realistic measure of achievement. The subsamples of students taking different content-based forms are difficult to combine in a single analysis, and the sample size for any one form would likely be too small for relational analysis. Even classroom means might be too inaccurate, because the sample of students would be so small in a classroom.

But stratified random instruments obtain a little information about each content category from each student in a classroom, while getting the same amount of item data per classroom. Total and subpool means can be calculated for each student and they will be comparable over forms. Consequently, integrative analyses can be carried out in which all student data are used simultaneously and the dependent variable (cognitive achievement in the whole pool or a stratum of the pool) includes form differences just as measurement error. Technical accuracy can be improved in such relational analyses by cross-calibrating the different forms. A simple method is to use form as a controlling variable in the analyses. In any case, the quality of the information with stratified random instrumentation is higher and so we should feel obliged to accept any computational complications.

Proposal 2. Flexible arrangements

The practicalities of administration vary across countries. In some, the success of the study will depend greatly on keeping the response burden on individuals to a minimum. In others, there would be no problem in having every student answer every cognitive item, in fact, this might be preferred.

Because of this variation, the study design should allow flexibility in the instrumentation arrangements, perhaps by defining three levels of matrix sampling--light, medium and complete response (no sampling). The light sampling is defined by having each student take one of the stratified random forms, the medium by two (or more?) forms, and the complete by each student's taking all forms. Attention in the administration procedures would have to be given to assuring rotation of the forms, and the data processign would have to anticipate varying amounts of data. No difficult problems can be seen.

The sampling referrees will, however, have to watch out for reduction in effective sample sizes when each student is providing more information. They should not allow the number of schools, for example, to be much reduced.

Proposal 3. National cores

The statistical accuracy of the analytic results, both the national and subnational summary item and subtest statistics and the relational coefficients, can be improved by incorporating a core set of items which is given to all students in addition to the stratified random forms.

This is proposed partly as a technical device. Under the item sampling plan, any particular item will be given to a sample of students spread over many classrooms. It will not be possible to restratify the students in each classroom, so the sample within classroom for an item will be simple random. But if there is a core test given to all students in the classroom, then the item statistics can be derived from regression formulas and ratio estimates which will have much more accuracy than conventional estimates. The core can also be used to cross-calibrate the test and subtest scales of the stratified random forms, and this will improve the quality of the pooling of the forms in relational analysis.

The determination of this core should be made partly on technical grounds: one wants a test which will correlate highly with most content areas, which discriminates well between students, and which does not have floor/ceiling problems. It is unlikely that a single core will have appropriate technical quality in all countries, especially with respect to difficulty level. Consequently, a separate core of items should be adopted by each country. When the cores are used for statistical adjustment of responses to the main item pool, the lack of international correspondence will not matter. A stratified set of items would be set aside for core use, and in each country, the local core would be selected on the basis of estimated item difficulty, content coverage, etc. Preliminary item statistics will help in choosing the cores.

Another reason for including national cores is to allow each country to obtain extensive data, especially pre/post data, in areas of content which are of more interest locally and which may not be covered in sufficient detail in the item pool and the international core.

Proposal 4. International core

The original concept of a core test was to provide a way to do international comparisons. If some countries are at the ceiling and others at the floor of a core, then the comparisons are limited and the core will be useless for national analyses in the countries at the extremes.

Nevertheless, a small international core is desirable. It would provide a quick way to check the statistics on the complete data, it would allow some quick initial analysis, and it would be a way to have some results which do not require too much explanation. The international core should be considered as a separate issue from the floating, national core.

Proposal 5. Pre-test/post-test schemes

In the longitudinal study, it is strongly advised that the cognitive instrumentation for the pre-test be a complete copy of the post-test instrumentation. That is, the same items should be used.

One reason is to be able to describe cognitive growth in detail, that is, for each item and subtest. A principal goal of the study is to find out how many items, which items, and what kind of items a student learns over the school year. A core test if defined internationally would be inappropriate for each nation separately.

A second purpose of the pre-test is to provide a control measure for assessing the effects of classroom processes. But in this case, the test had better measure student knowledge relative to what is taught during the year. Again, an international core will be insufficient for this purpose.

GENERAL DISCUSSION POINTS

A. Item and subtest analysis

The primary analytic outputs of the Study will be estimates of the difficulties of the cognitive items (each of the items) and estimates of the distributions of subtests from the item pool. The estimates are required at the national level for purposes of international comparisons. For within-nation analyses, it will be necessary to provide breakdowns of item difficulty and subtest performance according to educational or demographic stratifications of local interest. In the longitudinal analysis for Population A, the item statistics and subtest summaries need to be calculated at the beginning and end of the school year, presumably with equivalent detail (see Proposal 5), to evaluate growth.

While these item and subtest analyses may seem elementary, a great deal of statistical and computational labour will be devoted to them. Regression estimates based on core tests can greatly improve the quality of the results, and jackknife estimates of standard errors will probably be necessary. It would be valuable to ensure that each subtest had built-in replication. The design considerations for sampling items and students should take into account the likely within-country breakdowns as well as whole-country statistics.

B. Relational analysis

A major intent of the longitudinal study is to relate student background and classroom process with cognitive achievement. Inevitably, the cross-sectional studies will also attempt relational analyses. At the least, there will be item and subtest breakdowns by type of student (e.g., sex) and type of school or classroom (e.g., track).

It is important to keep in mind that the analytic goal of relational analysis is to estimate relational parameters, such as the regression coefficient relating mathematics achievement to student sex or mean classroom achievement to opportunity to learn. Theoretically, the cognitive measure for the individual or the classroom should be defined in terms of the whole pool of items or of some complete subpool. The analysis has to work from the incomplete response data, based on joint item and student sampling, to estimate the relationships among the theoretical variables. While this certainly will involve calculating averages and partial scores, the accuracy and significance of those intermediate calculations are only critical in how they affect the final results. For example, one is not fundamentally interested in the cognitive test score mean for a classroom, one wants instead to know the regression of true mean on classroom characteristics. If the items and students are too sparsely sampled, there will be technical difficulties in estimating the regression. If the item pool is substantively restricted--for example, by taking a limited core--then the regression analysis is inherently doomed.

C. Criteria for instrumentation

The decisions about instrument arrangement should be technically based and given as much hard analysis as are the decisions about student sampling. What can we say or guess about the inter-country or inter-content variability in item difficulty? Optimization of the instrument arrangements involves consideration of data costs, student time, and administrative costs relative to analytic accuracy.

Several notions of accuracy are relevant here: (1) the completeness of the sampling of the item pool or subpool; (2) the accuracy of that sampling, and whether the accuracy can be estimated and adjusted for; (3) the accuracy of the student and school samples for items and test forms; and (4) the efficiency of the statistical analysis.

D. Methodological principles

The problem of instrument arrangement is connected to the problem of student sampling, and a general principle is to spread the content as wide as possible. That is, each item or subtest should be administered to as broad as possible a sample of students. The matrix sampling is a great help.

A corollary is that the instrument arrangements should be designed to facilitate pooling of response data over forms or item samples. Relational studies should have as much content variability as possible included in their dependent measures, and this means, in particular, that analysis based just on core tests will be deficient (see Proposal 1).

12.5 Administration of the Cognitive Test

The IMC also proposes options for test administration. Which options a country chooses will depend, of course, on the amount of testing time availale and the number of items that are to be administered.

Minimum Expectation: At both pre- and post-test times a country administers the International Core (20 items) and the rotated forms (120 items).

a) With heavy sampling--all students take all items that would mean each student would take 140 items or about 3 hours of testing time.

b) With light sampling--students take the International Core plus one rotated form--each student would take 50 items at both pre- and post-test time.

Example of one option: At both pre- and post-test times a country chooses to administer the International Core plus a National Core plus the rotated forms plus additional items.

a) With heavy sampling this could mean that students would take up to 250 items at both the pre- and post-test sessons.

b) With light sampling a student could taken the International Core (20), the national core (2), a rotated form (50) and effectively take 80 items at both pre- and post-test times. (This assumes four rotated forms of 50 items each, 20 of which are chosen specifically by the country.)

12.6 Procedural Steps to Insure the Quality of the Cognitive Instruments

To be able to provide a valid item pool upon which countries can make valid choices it is necessary to gather additional evidence about the suitability of the cognitive items. What is needed is both universal quantitative information (item difficulties, estimates of amount of growth) and more qualitative information (appropriateness of item for growth, and when and to whom the content and behavior upon which the item is based is taught.

In order to get this information from countries, the IMC proposes the following set of procedures.

For each participating country

1. Selected items in the International Item Pool should be re-administered to two groups from students: those in the grade that contain the target population plus the appropriate grade. Since these data must be collected prior to January 1, 1980, this implies that

a. Northern Hemisphere countries should test students in the target grade and the next higher grade;

b. Southern Hemisphere countries should test the target grade and the grade level immediately prior to the target grade.

It is suggested that a minimum of 100 students at each grade level be administered each item; that the sample be a judgment sample that is representative of the population, and that there be documentation of the types of class and school in which the testing took place. The empirical results of this trial will provide the following information:

1. item difficulties for all items in the pool
2. an empirical estimate of the growth potential of the item
3. a means to estimate the between class and between school variability in item responses.

In addition to this quantitative information the IMC would like qualitative information about each item at the teacher level. This includes when the item is taught, to what proportion of the target population, and subjective judgments of the potential growth.

The IMC will organize the quantitative and qualitative information and provide it to countries as a basis upon which to "tailor" the cognitive instrument in line with the suggestions contained above.

The IMC recognizes the increased burden that countries are being asked to assume. We believe, however, the need to be responsive to national concerns and the complexity of developing a sound cognitive test that will produce valid international results dictate an increased knowledge of the properties of items in the International Pool. We appreciate your cooperation in the past and look forward to your comments on this proposal.

13. International Reporting

As stated in Bulletin No. 3 the international reporting of results will be through a series of volumes. One volume will focus on a Curriculum Analysis; one a comprehensive but not overly technical report of all the major findings of the study; one a technical volume that shows how a variety of statistical models of a highly technical nature can be used to analyze the data and; finally, a series of communications that highlights the most pertinent findings of the study, written in popular language and addressed to lay audience.

The schedule of completion for the international reporting follows, roughly, the timetable for the longitudinal study. That is to say, data are to be collected by mid-1982, analyzed by mid-1983 reports are to be completed by mid-1984.

Each of these reports will utilize the results of the study in different ways. Each requires analysis and presentation of data, but at different levels of technical and statistical sophistication. These requirements combined with the complexity of the design that is envisaged, make, as in previous studies, the calculation of statistics and dissemination of results a formidable task. Not only is it necessary to know how to calculate the appropriate statistics but also results must be provided quickly and efficiently to those who are responsible for reporting the results.

In addition to providing results to authors of various volumes, plus complete data tapes to authors of the technical volume, present plans call for the dissemination of procedures and software to facilitate within country analyses. The latter places additional importance on proper data analysis and appropriate statistical procedures. Although it will be complex technically to produce some of the desired results, it is essential that simple procedures and adequate software be provided so that countries can produce both analyses comparable to those done internationally and ones that are uniquely appropriate to them.

14. Statistics and data analysis

Tables 14.1 through 14.7 outline an initial attempt to delineate the variables in the study and to indicate the kinds of statistics that will be computed. Just as other tables tend to make the complex appear simple, so too for these. Given the design of the study some of these statistics are extremely difficult to compute. There are issues of the rotated forms, sampling decisions, weighting, levels of analysis, estimating standard errors, appropriate regression equations, and many others, embedded in the problems of what statistics should be computed and how they should be computed.

Below are some guidelines that have been discussed in relation to data analysis and the calculation of appropriate quantitative results. This list is not all inclusive and some guidelines have benefitted from more discussion than others. A thorough discussion of these issues and responses from the countries to the International Mathematics Committee (IMC) will, we hope, add to and clarify these general guidelines.

14.1. Measures of Central Tendency and Variability

The tables contain marks where measures of central tendency and variability are to be computed. It should go without saying that such measures will be chosen that best reflect the type of variable being analyzed and the kind of information to be generated.

14.2. Item Analysis

Though not specifically stated in the Tables, both the cognitive tests and the attitude scales will be analyzed for their psychometric properties. Such analyses include item discrimination parameters, factor analysis (of various sorts), latent trait analysis and other appropriate statistical descriptors and methods.

14.3. Levels of Analysis

The full study is based on a sampling plan that includes students within classrooms, classrooms within schools and schools within countries. Although the question one asks should always precede the statistics that are computed, it is important to keep in mind the different levels of analyses that are possible. Of particular importance in this regard, are the distributions of variables within classes. Present plans call for the calculation of a number of descriptors of these within classroom distributions.

14.4. Standard errors.

It has been suggested that as a general principle no statistic be disseminated without an honest standard error attached to it. This implies the jackknifing of all estimators in order to get appropriate standard errors.

14.5. Weighting

It is unlikely that samples will be self-weighting. This implies the necessity of using weighting schemes in the calculation of statistics. Such is especially the case where countries might choose to over sample a particular stratum in order to address an important question. On the other hand there are some analyses where weighting might not be important (e.g., between country analyses). In general the guideline is to weight appropriately where necessary and to present both weighted and unweighted results where they will provide more insightful interpretation of results.

14.6. Corrected Scores

Given that there are many corrections for guessing and none available given the design that will provide an appropriate correction, in general the results should be produced without corrections for guessing. Those countries or individuals who wish to correct would be free to use the correction they deem most appropriate.

14.7. Standard Scores

When presenting profiles of various scores and results it may be desirable to express scores in standard units. It is suggested that those standard units be chosen so as to eliminate negative and non-integral scores. For example, a within country metric might have a mean of 50 and standard deviation of 10; a between country metric might have a mean of 100 and standard deviation of 15.

What follows are tables that begin to delineate what kinds of statistics will be computed for the variables that are to be in the study. As the study becomes more refined, the variables may change and how they will be handled statistically could change also.

One assumption upon which the tables are based is that countries will use rotated forms to gather information about students' cognitive achievements. A second is that the sampling plan will be based on samples of classrooms within school, thereby making the level of analysis issue extremely important. As a consequence, many satistics must be computed for students within classrooms, classrooms within schools, and schools.

Although, given the design of the study, some of these statistics will be difficult technically to compute, the general goal is to produce estimates that are relative, common and straightforward.

STATISTICS AND DATA ANALYSIS

14.1. SCHOOL QUESTIONNAIRES (POPULATIONS A AND B)

VARIABLE	STATISTICS: Frequency Distribution	Measures of Central Tendency	Measures of Variation	Level
1. Community Served	X			3
2. Enrollment	X	X	X	3
3. Enrollment: Sampled Population	X	X	X	3
4. Enrollment: Sampled Population Mathematics	X	X	X	3
5. Number of Teachers	X			3
6. Number of Mathematics Teachers	X			3
7. Mathematics Preferred teaching subject (teachers)	X			3
8. Number of school days/year	X	X	X	3
9. Number of periods/day	X	X	X	3
10. Length of periods	X	X	X	3
11. Calculators: Encouraged	X			3
12. Staff meetings: teachers mathematics	X			3
13. Staff meetings: activities	X			3
14. Calculators: Policy, Four functions	X			3
15. Calculators: Policy, Programmable	X			3
16. Grouping: Policy	X			3
17. Gender: Policy	X			3

Legend:
Central Tendency: p = proportion, x = mean.
Variation: SE = standard error, SD = standard deviation, SS = both.
Level: 1 = student, 2 = classroom, 3 = school, 4 = all levels.

14.2. TEACHER QUESTIONNAIRES (POPULATIONS A AND B)

VARIABLE	STATISTICS: Frequency Distribution	Measures of Central Tendency	Variation	Level
1. Gender	X			2,3
2. Age	X	X	X	2,3
3. Teaching Experience - Total	X	X	X	2,3
4. Teaching Experience - Mathematics to Pop.	X	X	X	2,3
5. Preparation - Mathematics	X	X	X	2,3
6. Preparation - Pedagogy: Mathematics	X	X	X	2,3
7. Preparation - Pedagogy: General	X	X	X	2,3
8. Teaching Hours/Week: Total	X	X	X	2,3
9. Teaching Hours/Week: Mathematics	X	X	X	2,3
10. Additional Duties	X			
11. Teaching Schedule				
a) Pop. A or B - classes; hours	X	X	X	2,3
b) Lower - classes; hours	X	X	X	2,3
c) Higher - classes; hours	X	X	X	2,3
12. Target Class: Subjects Taught	X			
13. Target Class: Number of Teachers	X			
14. Target Class: Number of Students	X	X	X	2,3
15. Target Class: Periods Instruction/Week	X	X	X	2,3
16. Target Class: Length of Period	X	X	X	2,3
17. Target Class: Hours Instruction/Year	X	X	X	2,3
18. Target Class: Compared to others in School	X			2,3
19. Target Class: Ability Range	X			2,3
20. Target Class: Initial Mastery	X			2,3
21. Target Class: Subject Matter	X			2,3
22. Target Class: Compared to Country population	X			2,3
23. Target Class: Activities Teacher				
a) Last Week	X			2,3
b) Typical Week	X			2,3
24. Target Class: Activities Student				
a) Last Week	X			2,3
b) Typical Week				
25. Target Class: Students Respond to Questions	X			2,3
26. Target Class: Varying Assignments	X			2,3

TEACHER QUESTIONNAIRES (CONTINUED)

27. Target Class: Hours Homework Assigned				
a) Last Week	X	X	X	2,3
b) Typical Week	X	X	X	2,3
28. Target Class: Calculators - Access	X[1]	X[1]	X[1]	2,3
29. Target Class: Calculators - Use	X[2]	X[2]	X[2]	2,3
30. Target Class: Topics Covered				
a) Number of weeks	X	X	X	2,3
b) Spiral	X			2,3
31. Target Class: Materials	X			2,3
32. Target Class: Textbook	X			2,3

[1]This question yields at least two variables; frequency of use with what type of calculator

[2]This question also yields at least two variables; how used with what type of calculator

Possible Composites:

A. Allocated Time: Total-Var 15 x Var 16 (Var 17 provides a check)
B. Allocated Time: Total Instructional - Var 15 x Var 16 x ((Var 23b + Var 23c)/Var 23a + . . . + Var 23f)
C. Allocated Time: Instructional - Common Fractions Var 15 x Var 16 x ((Var 23b + Var 23 c)/Var 23a + . . . + Var 23f) x (Var 30a)
D. Allocated Time: Instructional - Decimal Fractions (as above except substitute Var 30b for Var 30a)
E. Allocated Time: Instructional - Ratio and Proportion (as above except)
F. Allocated Time: Instructional - Percentage (as above)
G. Allocated Time: Instructional - Measurement (as above)
H. Allocated Time: Instructional - Geometry (as above)
I. Allocated Time: Instructional - Formulae & Equations (as above)
J. Allocated Time: Instructional - Directed Numbers (as above)
K. Student Time - Var 24a + Var 24b + Var 24c + Var 24d
L. Materials: Variety - Var31a + . . . + Var 31g
M. Materials: Individualized - Var 31c + Var 31f + Var 31g

Possible Clustering Variables:

A. Target Class: One teacher - Var 11 combined with Var 12 and Var 13
B. Target Class: Ability - Var 18 combined with Var 19, Var 20, Var 21 and Var 22
C. Teacher: Mathematics Preparation: Var 5/(Var 6 + Var 7)
D. Teacher: Individualization: Var 24d/(Var 24a + Var 24b + Var 24c) combined with Var 25 and Var 26

14.3. STUDENT QUESTIONNAIRES (POPULATIONS A AND B)

VARIABLE	STATISTICS: Frequency Distribution	Central Tendency	Measures of Variation	Level
1. Gender				
2. Age	X	X	X	4
3. Father's Occupation (Scaled)	X	X	X	4
4. Mother's Occupation (Scaled)	X	X	X	4
5. Father's Education (Scaled)	X	X	X	4
6. Mother's Education (Scaled)	X	X	X	4
7. Language Spoken in Home	X			4
8. Planned Further Education	X	X	X	4
9. Hours Homework - Mathematics				
a) Last Week	X	X	X	4
b) Typical Week	X	X	X	4
10. Hours Homework - All Subjects	X	X	X	4
11. Hours Outside Tutoring - Mathematics				
a) Last Week	X	X	X	4
b) Typical Week	X	X	X	4
12. Parents Help with Mathematics: Frequency	X	X	X	4
13. Computational Aids: What Used	X[1]	X[1]	X[1]	4
14. Computational Aids: How Used	X[2]	X[2]	X[2]	4
15. Home Support for Mathematics (18 items)	X	X	X	4

[1] This question yields at least two variables; not only what computational aids are used but also where they are used.

[2] This question also yields at least two variables; how computational aids are used and which ones for what purposes.

Possible Composites:

A. Student Background - Var 1 + Var 2 + Var 3 + Var 4 (Weights for composite may be chosen in a number of ways)
B. Exposure to Mathematics Outside of School - Var 9 + Var 11 (Composite to be formed of either the "a's" or "b's")
C. Environmental Support for Mathematics - Var 12 + Var 15 (Scale to be formed; potentially there are 14 items)
D. Computational Aids: Access - (Var 13a + . . . + Var 13e)
E. Computational Aids: Use - (Var 14a + . . . + Var 14 e)

14.4. COGNITIVE TESTS (POPULATION A)

VARIABLE	STATISTICS: Frequency Distribution	Central Tendency	Variation	Level
For Pre and Posttest or Posttest only				
1. Item responses (about 160)	X	p	SE	1,2
2. Subtest: Fractions	X	X	SS	4
3. Subtest: Percent, Ratio, Proportion	X	X	SS	4
4. Subtest: Algebra	X	X	SS	4
5. Subtest: Geometry	X	X	SS	4
6. Subtest: Measurement	X	X	SS	4
7. Other Content Subtests as Specified (e.g., whole numbers)	X	X	SS	4
8. Subtest: Computation	X	X	SS	4
9. Subtest: Comprehension	X	X	SS	4
10. Subtest: Higher Level Behaviors	X	X	SS	4
11. Other behavioral level subsets (e.g., minimal competence)	X	X	SS	4
Growth Scores				
1. Item responses (Raw Gains)	X	[illegible]	SE	1,2
2. Subtest: Fractions (Raw Gains)	X	X	SS	4
3. Subtest: Percent, Ratio, Proportion	X	X	SS	4
4. Subtest: Algebra	X	X	SS	4
5. Subtest: Geometry	X	X	SS	4
6. Subtest: Measurement	X	X	SS	4
7. Subtest: Other content	X	X	SS	4
8. Subtest: Computation	X	X	SS	4
9. Subtest: Comprehension	X	X	SS	4
10. Subtest: Other behavioral	X	X	SS	4

COGNITIVE TESTS (POPULATION A) CONTINUED

VARIABLE	Frequency Distribution	Central Tendency	Variation	Level
Profiles within and between country (pre and posttests or posttests only)				
1. Item reponses		p	SS	1,2
2. Content Subtests		X	SS	1,2
3. Behavioral Subtests		X	SS	1,2
4. Items aggregated according to issue (e.g., role of applications)		X	SS	1,2
Profiles within and between country (growth)				
1. Item responses (Raw Gains)		p	SE	1,2
2. Content subtests		X	SS	1,2
3. Behavioral subtests		X	SS	1,2
4. Other subtests		X	SS	1,2
Anchor Items				
1. Differences at item level	X	p	SE	1
2. Differences at subtest level	X	X	SS	1
3. Profiles within and between countries	X	X	SS	1
4. Growth between pre and posttests on anchors	X	X	SS	4
Variance Components Analysis (Random Effects Analysis of Variance)				
1. Items by Students	To estimate components of variance due to students, classrooms, schools and interactions.			
2. Subtests by Students				

Including Fractions, Ratio, proportion and percent; Algebra, Geometry, Measurement and other content areas. Also, possibility of doing it for behavioral levels.

COGNITIVE TESTS POPULATION B

The statistics to be computed are similar to those for Population A

a) There will be no growth scores

b) The topic subscores will include:
   1) Arithmetic/number systems
   2) Algebra/polynominals; equations and inequations
   3) Geometry/trigonometry
   4) Analysis/functions; differentiation; integrations

c) Subscores within behavioral levels are also envisaged.

d) Item responses will be tabulated

e) Within and between country profiles will be generated

f) Analysis based on anchor items will be conducted

14.5. ATTITUDE MEASURES (BOTH POPULATIONS AND TEACHERS)

VARIABLE	Frequency Distribution	Central Tendency	Variation	Level
1. Mathematics in School				
a) Item responses	X			1,2
b) Subscale: Like (alpha)	X	X	SS	4
c) Subscale: Important	X	X	SS	4
d) Subscale: Difficult	X	X	SS	4
e) Whole Scale	X	X	SS	4
f) Subscale: Growth	X	X	SS	4
2. Mathematics as a Process				
a) Item responses	X			1,2
b) Full Scale	X	X	SS	4
3. Mathematics and Utility				
a) Item responses	X			1,2
b) Full Scale (alpha)	X	X	SS	4
4. Mathematics and Myself				
a) Item responses	X			1,2
b) Whole Scale (alpha)	X	X	SS	4
5. Mathematics Anxiety				
a) Item responses	X			1,2
b) Whole Scale (alpha)	X	X	SS	4
6. Calculators and Computers				
a) Item Responses	X			1,2
b) Whole Scale (alpha)	X	X	SS	4
7. Within and Between Country Profiles				
a) All attitude scales and subscales	X	X	SS	1,3

14.6. CLASSROOM PROCESSES AND OPPORTUNITY TO LEARN (POPULATION A)

VARIABLE	STATISTICS			
	Frequency Distribution	Central Tendency	Variation	Level
1. Opportunity to Learn				
a) Teacher's responses				
1) Item responses	X			2
2) Subscore: Fractions	X	X	SS	2
3) Subscore: Ratio, proportion, percent	X	X	SS	2
4) Subscore: Algebra	X	X	SS	2
5) Subscore: Geometry	X	X	SS	2
6) Subscore: Measurement	X	X	SS	2
b) Student's responses				
1) Item responses	X			1,2
2) Subscore: All above	X	X	SS	1,2
c) Within and Between Country Profiles	X	X	SS	4
2. Classroom Processes (variables in the process of being refined)				
a) Item responses (e.g., number of interpretations) Fractions Ratio, Proportion, Percent Algebra Geometry Measurement	X	X	SS	2
b) Composites (e.g., degree of individualization, teaching strategy) Fractions Ratio, Proportion, Percent Algebra Geometry Measurement	X	X	SS	2
c) Within and Between Country Profiles	X	X	SS	4

14.7. BIVARIATE STATISTICS

1. "Grand" correlation matrix
   a) At student, class, and school levels
   b) Among all continuous variables
   c) Including:
      1) Subtests of the Cognitive Examination
      2) Attitude Scales
      3) Teacher General characteristics (Including composites)
      4) Home background variables (Including composites)
      5) Home environmental variables (Including composites)
      6) Topic specific teaching practices (Including composites)
      7) Topic specific Opportunity to Learn
      8) School variables

2. Other appropriate "relational analyses"
   a) At student, class and school levels
   b) Among categorical and continuous variables
   c) Calculation of "statistics" such as
      1) Eta^2
      2) $Omega^2$
      3) Contingency Table coefficients
      4) Log linear models

3. One type of Multivariate Analysis (Growth)
   a) At the topic level
      1) Students Pooled Within Classrooms:
         Topic Specific Posttest regressed on Topic specific pretest, Background variables, Opportunity to Learn, Student level instructional variables
      2) Between classroom analysis
         Aggregated Posttest regressed on (Predicated posttest scores from pooled within classroom regression, classroom variables, school variables)
   b) Technical recommendations
      1) Use dummy variable for rotated form
      2) Use unstandardized coefficients
      3) Jackknife to get standard errors
      4) Should be properly weighted (depends on sampling plan, point of view about weighting)
      5) Should make assumption checks (Homogeneity of composite regression (from 1); robustness of regression)

15. TIMETABLE OF KEY DATES

The following timetable is reproduced for convenience from Bulletin 3 and applies to Population B. For Population A, (longitudinal) countries should add one year to the key dates for sampling plans, translation, administration, and so forth.

To ensure inclusion of results in the international report (see Section 13), countries are urged at this time to plan to have completed data collection by mid-1982.

TIMETABLE FOR SECOND IEA MATHEMATICS STUDY

	Start	Complete
Curriculum Analysis		
Preliminary analysis	October 1976	January 1977
National responses to international grid	January 1977	August 1978
IMC Meeting	August 1978 January 1979 September 1979 February 1980	September 1978 February 1979 September 1979 March 1980
Planning of Curriculum Analysis Model	May 1978	August 1978
Preparations for Curriculum Analysis Symposium-Committee to write up national statements from Working Papers I and VI, textbooks, examinations, etc.	September 1978	December 1978
National Centers to identify key national mathematics experts	September 1978	January 1979
Curriculum Analysis Symposium		August 1979
International Report: data analysis, editing of Symposium proceedings and papers	September 1979	January 1980
Publication of Volume I, Curriculum Analysis Report	January 1980	July 1980
Cognitive Test Construction		
Identify International Test Grid		September 1977
National responses to International Grid	September 1977	July 15, 1978
Field Trial of item collections	May 1978	January 1979
IMC and National Centers write new items to fill gaps in grid	May 1978	August 1978
Review trial data and extra items	May 1978	August 1978

	Start	Complete
IMC report on trial data and final chance for National Centers to contribute items for trial	September 1978	September 1978
Response to IMC report	September 1978	December 1978
Additional field trials as necessary	October 1978	December 1978
Review of field trial data and synthesis of item pool by New Zealand Coordinating Unit	January 1979	January 1979
Final draft of cognitive instruments by IMC	February 1979	February 1979
Initial review and comment on draft of cognitive instruments by National Centers	March 1979	March 1979
Preparation of manuals	September 1978	February 1979
Dry Run all instruments (includes translation and refereeing of national options)	April 1979	July 1979
Final instruments		October 1979
Printing and distribution of instruments	October 1979	December 1979
Administration of pretest		
Southern Hemisphere	Feburary 1980	April 1980
Northern Hemisphere	September 1980	October 1980
Administration of posttest		
Southern Hemisphere	October 1980	April 1981
Northern Hemisphere	April 1981	July 1981
<u>Classroom Processes Instruments</u>		
Initial development of combined opportunity-to-learn and classroom processes instrument	June 1977	August 1978
New Zealand pilot trial of growth scores and classroom instrument	March 1978	December 1978
Consultations on instrument	May 1978	August 1978
Draft of instrument		August 1978

	Start	Complete
Limited national trials	August 1978	December 1978
Data analysis of trials plus New Zealand data	January 1979	February 1979
International trial as part of Dry Run	April 1979	July 1979
Finalize instrument	July 1979	October 1979
Translation, refereeing, and printing	October 1979	January 1980
International Mathematics Committee meetings		August 1978 February 1979 October 1979
Manuals	April 1979	October 1979
Administration of classroom instrument		
Southern Hemisphere	February 1980	January 1981
Northern Hemisphere	September 1980	July 1981

Attitude Scales

	Start	Complete
Rationale, and identification and development of affectives scales	January 1977	June 1977
Pilot trial affective scales in USA	June 1977	October 1977
International trials of affective scales	March 1978	June 1978
Data analysis of trials	June 1978	August 1978
IMC report on field trials	September 1978	September 1978
Review of National Center comments	October 1978	October 1978
Additional field trials (if necessary)	October 1978	December 1978
Final draft of affective scales		February 1979
Translation, refereeing of problems	February 1979	March 1979
Dry Run	April 1979	July 1979

	Start	Complete
Final instruments		October 1979
Completion of manuals		October 1979
Printing and distribution of instruments	October 1979	December 1979
Administration of pretest		
Southern Hemisphere	February 1980	April 1980
Northern Hemisphere	September 1980	October 1980
Administration of posttest		
Southern Hemisphere	October 1980	January 1981
Northern Hemisphere	April 1981	July 1981
Student, Teacher, and School Questionnaires		
Draft questionnaire items for student, teacher, and school questionnaires	January 1977	July 1978
Draft questionnaires		August 1978
International Trial in conjunction with trial of additional cognitive items	September 1978	December 1978
Analysis of final data	January 1979	February 1979
IMC settle final draft instruments		February 1979
IMC report data to national centers	February 1979	March 1979
Questionnaires finalized		October 1979
Completion of manuals	February 1979	October 1979
Translation and refereeing	November 1979	December 1979
Administration of questionnaires		
Southern Hemisphere	February 1980	January 1981
Northern Hemisphere	September 1980	July 1981
Sampling		
Discussion of sampling specifications and consultation	January 1977	May 1978

	Start	Complete
International Sampling Committee prepare draft paper for IMC	May 1978	August 1978
Final sampling design settled	August 1978	September 1978
Sampling manual prepared	September 1978	February 1979
Manual approved IMC		February 1979
National Centers draw samples and consult International Sampling Committee	February 1979	October 1979
International Sampling Committee report to IMC		October 1979
Natonal Centers contact schools and replace refusals		
Southern Hemisphere	October 1979	December 1979
Northern Hemisphere	June 1980	July 1980
Data collection		
Southern Hemisphere	February 1980	January 1981
Northern Hemisphere	September 1980	July 1981
Data Collection Modes		
Consultations with National Centers on potential methods of data collection	January 1977	July 1978
Report to IMC by Dr. J. Schwille	July 1978	August 1978
IMC suggestions to National Centers and return of National Center comment	September 1978	December 1978
Methods of data collection settled		December 1978
Printing of answer forms (if necessary) and dispatch to countries	January 1979	March 1979
Pretest		August 1979
Posttest		June 1980
Completion of manuals		
Pretest		October 1979
Posttest		October 1979

	Start	Completion
Data Processing and Analysis		
Preliminary Planning:		
Outline of instruments with approximate number of items	December 1978	February 1979
Outline of codebooks (dummy)	July 1978	February 1979
Preliminary consideration of file building	July 1978	February 1979
Detailed Planning:		
Settle coding of final instruments	February 1979	October 1979
Standardize punching and coding forms for Dry Run	February 1979	March 1979
Settle analyses required by IMC	October 1978	February 1979
Settle file building and weighting procedures	February 1979	July 1979
Update of Codebooks	February 1979	July 1979
International trial of countries' capacity to produce files and undertake standard analyses (run as part of the Dry Run)	February 1979	July 1979
Write programs for basic item analyses, univariates, correlations, school reports, and for special multivariate analyses	February 1979	March 1980
Analyses for IMC and countries requiring assistance	April 1980	December 1981
Construction of data bank	December 1981	December 1982

16. References

These documents have appeared during 1979 and are of particular relevance to the Second International Mathematics Study.

Bulcock, J.W., W.F. Lee and W.S. Luk. "A Model of School Learning Using the Variance Normalization Method of Ridge Regression Analysis." Presented at IEA Methodology Symposium, Michigan State University, February 1-2, 1979.

Burstein, Leigh and Robert L. Linn. "The IEA Longtitudinal Analysis from a Multilevel Perspective: Disentangling Between- and Within-Class Relationships." Prepared for Winter 1979 Symposium on Previews of the Second IEA Mathematics Data Analysis, Michigan State University, East Lansing, Michigan, February 1, 1979.

Kifer, Edward. "The Teaching and Learning of Mathematics in Four Countries: A Cross-National Comparison." Presented at an IEA Methodology Symposium, Michigan State University, February 1-2, 1979.

Munck, Ingrid M. E. Model Building in Comparative Education. IEA Monograph Number 10. Stockholm: Almquist and Wiksell International, 1979.

McKnight, C. "Conceptualizing the complemented curriculum." Working Paper 1 of the U.S. National Coordinating Center.

McKnight, C. "Teacher practice variables in the topic-specific instruments." Working Paper 2 of the U.S. National Coordinating Center.

McKnight, C. "Conceptualizing topic-specific teacher practices." Working Paper 3 of the U.S. National Coordinating Center.

Nuthall, Graham. "Relating Classroom Process Variables to Measures of Student Achievement and Learning in the Second IEA Mathematics Study." Presented at IEA Methodology Symposium, Michigan State University, February 1-2, 1979.

Postlethwaite, T. Neville, and Arieh Lewy. Annotated Bibliography of IEA Publications (1962-1978). Stockholm: University of Stockholm, 1979.

Schmidt, William H. "Multivariate Least Squares and Structural Equations: Fitting Different Methods to Different Questions." Presented at IEA Methodology Symposium, Michigan State University, February 1-2, 1979.

Travers, Kenneth J. "The Second International Mathematics Study: an Overview." Paper prepared for IEA Methodology Symposium, Michigan State University, February 1-2, 1979.

Travers, Kenneth J. "The Curriculum Analysis: Some hypotheses and implications for the Second International Mathematics Study.

Travers, Kenneth J. "The Second IEA Mathematics Study." Journal of Curriculum Studies, 1979, II, 3, 203-210.

Tyler, Ralph W. "American Educaion in the Perspective of Education of Other Nations." (mimeo, undated).

The Second International Mathematics Study: An International Portrait of Mathematics Education

ACER Papers

Rosier, Malcolm. "Planned analyses for the Second IEA Mathematics Study-Cross-sectional Study in Australia." February 1979.

"Report of the IEA Second International Mathematics Study Data Analysis Seminar." Melbourne, ACER, February 1979.

Keeves, J. P. and R. Lewis. "Teachers, classrooms and student outcomes."

Rosier, M. "Hypotheses for the Australian National Study of Mathematics Achievement."

APPENDIX A

DISTRIBUTION OF COGNITIVE ITEMS

POPULATION A

176 ITEMS

CONTENT	BEHAVIOUR LEVELS				TOTAL
	I	II	III	IV	
000 ARITHMETIC	24 Items	19 Items	15 Items	2 Items	60 Items
001 Whole Numbers	A/1 A/2 A/3*	A/4* A/5*	A/6* A/7*	A/8	8
002 Common Fractions	Core/1* Core/2 A/9 A/10 A/11 D/2	Core/3 Core/4 A/12* A/13 D/1 D/3	A/14 D/5 D/4 D/6		16
003 Decimal Fractions	Core/5* B/1 B/2 B/3* C/6 ~~C/7~~ C/8	Core/6 B/4 C/9 C/10 C/11	Core/7 Core/8 B/5 C/12*		16
004 Ratio, Proportion, Percent	Core/9 Core/10 C/1 C/2	Core/11 Core/12 Core/13 C/3	Core/14 Core/15 Core/16 C/4 C/5		13
005 Number Theory	B/7	B/6* B/8		B/9	4
006 Powers	C/13* C/14 C/15				3

CONTENT	BEHAVIOUR LEVELS				TOTAL
	I	II	III	IV	
100 ALGEBRA	24 Items	15 Items	8 Items	1 Item	48 Items
101 Integers	D/7 D/8 D/9 D/10	D/11	D/12		8
102 Rationals	D/13	S1			2
103 Integer Exponents	D/14				1
104 Formulas	Core/17 A/15 A/16 A/17* D/15* D/17 D/18	Core/19 A/19 A/21 D/20 D/21 D/22*	Core/20* A/20 A/22 D/23 D/24 D/25*		19
105 Polynomials Expressions	A/18 D/16 Core/18*	D/19			4
106 Equations and Inequations	Core/21 Core/22* B/10 B/11 B/12 B/13* C/16	B/14 B/15 C/17 C/18 C/19*	Core/23	Core/24	14
107 Relations	C/21	C/20			2

CONTENT	BEHAVIOUR LEVELS				TOTAL
	I	II	III	IV	
200 GEOMETRY	8 Items	10 Items	16 Items	2 Items	36 Items
201 Classification		A/25			1
202 Properties	A/23 D/27	Core/27 D/26	A/26* A/27* D/28 D/29	Core/32* A/31*	10
203 Congruence	Core/25		Core/29 D/30		3
204 Similarity		Core/28	Core/31 A/28*		3
205 Geometric Constructions			S2		1
206 Pythagorean	Core/26 D/31		A/30		3
207 Coordinates	A/24	S3 S4	A/29		4
~~208 Deductions~~					
209 Transformation (Informal)	C/22	C/23	Core/30 C/24 C/25		5
212 Spatial Visualization		D/32	A/32		2
215 Transformational Geometry	S5	S6 S7	S8		4

CONTENT	BEHAVIOUR LEVELS				TOTAL
	I	II	III	IV	
300 PROBABILITY AND STATISTICS	2 Items	2 Items	4 Items	1 Item	9 Items
302 Organization		B/22			1
303 Representation			B/19 B/20 B/21	B/24*	4
304 Mean, Median, Mode	B/16*	B/23	B/17*		3
306 Probability	B/18				1
400 MEASUREMENT	7 Items	7 Items	8 Items	1 Item	23 Items
401 Units	C/26	C/29	Core/33		3
402 Estimation	B/25	B/28 C/30	Core/35		4
403 Approximation		Core/34* B/29			2
404 Determining Measures	Core/36 Core/37 B/27 C/27 C/28	Core/39 B/26	Core/38 Core/40 B/30 B/31 B/32* C/31	C/32	14

*Anchor Items

Item references are given by Form and number e.g., A/3 is item number 3 in Rotated Form A.
"S" items are supplementary, added to reflect I_s topics

DISTRIBUTION OF COGNITIVE ITEMS

POPULATION B

136 items (120 items plus 16 supplementary items)

CONTENT	BEHAVIOUR LEVELS I	II	III	IV	TOTAL
1 SETS, RELATIONS AND FUNCTIONS	2 Items	3 Items	1 Item	1 Item	7 Items
1.1 Set Notation		1/1			
1.2 Set Operations		3/1*			
1.4 Functions	4/1 5/1	6/1	7/1	2/1	
2 NUMBER SYSTEMS	4 Items	4 Items	5 Items	3 Items	16 Items
2.1 Common Laws	1/2	5/3*	1/3		
2.2 Natural Numbers		6/3	3/3	5/4*	
2.3 Decimals	3/2		4/3	7/4	
2.4 Real Numbers		4/2	6/2*		
2.5 Complex Numbers	5/2 7/2*	7/3	2/3	3/4	
3 ALGEBRA	9 Items	6 Items	7 Items	3 Items	25 Items
3.1 Polynomials	6/4 7/5*	1/5 2/4 5/6		1/7	
3.2 Quotients of Polynomials	3/6 5/5				
3.3 Roots and Radicals	4/4 6/5	7/6	5/7 6/6*		
3.4 Equations and Inequations	3/5* 6/7*	4/5	7/7* 1/6*	2/6	
3.5 Systems of Equations and Inequations	2/2	3/7	3/8 1/4	4/6	
3.6 Matrices			2/5		

CONTENT	BEHAVIOUR LEVELS				TOTAL
	I	II	III	IV	
4 GEOMETRY	5 Items	4 Items	5 Items	4 Items	18 Items
4.1 Euclidean	7/9				
4.3 Analytic	6/8	1/8*	4/8* 5/9	2/9 1/10	
4.5 Vector Methods	4/7*	7/8 2/7	3/10	7/10	
4.6 Trigonometry	3/9 5/8*	6/9	2/8* 1/9	6/10	
(Supplementary items)		S1 - S14			(14 Items)
5 ANALYSIS	15 Items	13 Items	12 Items	4 Items	44 Items
5.1 Elementary Functions	1/11* 2/10 3/11	6/12 5/11 4/9* 4/10	4/11* 3/13* 2/11* 2/12		
5.2 Properties of Functions		3/12 5/10*	1/12	7/12	
5.3 Limits and Continuity	6/11*	7/11* 8/3*	8/5		
5.4 Differentiation	8/1 1/13	5/12*	8/13		
5.5 Applications of the Derivative	2/13 4/12 6/13 7/14 8/7		8/12	8/6 8/14	
5.6 Integration	8/2 8/8	8/9 8/11	7/13 8/4*		
5.7 Techniques of Integration	1/14*		5/13*		
5.8 Applications of Integration	2/14	4/13 8/10*	3/14	8/15	
6 PROBABILITY AND STATISTICS	4 Items	2 Items	2 Items		8 Items
6.1 Probability	2/15* 3/15*	6/14			

CONTENT	BEHAVIOUR LEVELS				TOTAL
	I	II	III	IV	
6.2 Statistics	4/14 5/14		4/15		
6.3 Distributions		1/15	5/15		
7 FINITE MATHEMATICS		1 Item	1 Item		4 Items
7.1 Combinatorics	S15	6/15 S16	7/15		

Appendix B

STRUCTURE OF POPULATION A COGNITIVE TEST
(Longitudinal)

CORE TEST

40 items

Content	Behavior Levels				Total
	I	II	III	IV	
Common and Decimal Fractions	B/6* H/5 A/30	K/8 G/5 A/6	L/22 L/24		8
Ratio, Proportion and Percent	L/3 L/6	L/23 I/8 F/1	H/9 E/8 F/25		8
Formulas, Expressions, Equations	C/14 D/25* A/17 F/17*	G/12	J/16* L/14	D/28	8
Geometry	L/27 K/20	J/27 K/28	H/24 B/32 C/27	I/26*	8
Measurement	K/30 K/35	I/39* K/36	C/36 E/39 K/27 J/40		8
Totals	13	11	14	2	40

*anchor item

Items identified by numbers used in most recent set of trial instruments (1978)

This Appendix has not been updated to show the supplementary items.

POPULATION A COGNITIVE INSTRUMENTS

176 items

Content	Test Forms				
	Core N items	Form A N items	Form B N items	Form C N items	Form D N items
Whole Numbers		8			
Common Fractions	8	6			6
Decimal Fractions			5	7	
Ratio, Prop., Percent	8			5	
Number Theory			4		
Power/Exponents				3	
Square Roots					
Dim. Analysis					
Integers, Rat., Exp.					8
Formulas		8			11
Expressions	8				
Equations			6	6	
Plane Figures	8	10			7
Transformation				4	
Probability/Stat.			9		
Measurement	8	8	7		
Supplementary Items	0	2	2	2	2
Totals	40	34	34	34	34

ROTATED FORM A

32 items

Content	Behavior Levels				Total
	I	II	III	IV	
Whole Numbers	E/2 D/1 I/2*	F/2* J/2	C/4* I/1*	C/10	8
Common Fractions	D/4 I/4 J/3	G/4* K/13	K/31		6
Formulas, Expressions	K/32 H/16 B/15* I/16	J/15 L/16	C/17 B/19		8
Geometry	I/31 G/28	I/23	B/28* I/24* G/23* G/24 D/34 A/31	F/33*	10
Totals	12	7	11	2	32

*anc. r item

Items identified by numbers used in most recent set of trial instruments (1978)

ROTATED FORM B

32 items

Content	Behavior Levels I	II	III	IV	Total
Decimal Fractions	B/7 G/2 G/7*	I/7	G/6		5
Number Theory	A/9	B/2* I/9		D/2	4
Algebra Equations/ Inequ. Relations	B/13 H/18 G/19 E/21*	E/22 F/19			6
Probability/Stat.	H/33* H/35	A/36 G/33	F/36* F/39 G/34 G/35	B/34*	9
Measurement	E/37 L/9	A/39 B/31 B/39	J/39 I/37 B/40*		8
Totals	12	10	8	2	32

*anchor item
Items identified by numbers used in most recent set of trial instruments (1978)

ROTATED FORM C

32 items

Content	Behavior Levels I	II	III	IV	Total
Ratio, Proportion, Percent	K/10 D/13	B/8	G/9 E/9		5
Decimal Fractions	K/1 F/6 D/9	G/40 K/2 F/7*	J/11*		7
Powers/Exponents	C/15* I/10 E/11				3
Algebra Equations Inequ. Relations	K/7 I/28	I/17 C/19 K/26 Added*			6
Trans. Geo.	J/32	C/25	E/35 B/24		4
Measurement	F/37 K/39 L/26	K/11 F/38	L/7	K/40	7
Totals	14	11	6	1	32

*anchor item

Items identified by numbers used in most recent set of trial instruments (1978)

ROTATED FORM D

32 items

Content	Behavior Levels				Total
	I	II	III	IV	
Common Fractions	A/3	A/4 L/1	D/7 L/8 B/5		6
Integers/Rational Integers	H/14 I/12 E/13 A/11	D/18	H/13		8
Rationals	H/15				
Integer Exp.	K/4				
Algebra Formulas	Added* L/15 G/15	D/27 A/15 A/14*	E/18 L/33		11
Expressions Equations	D/22	F/15	Added*		
Geometry	J/31 L/36	A/21 K/19	E/24 G/30 H/40		7
Totals	13	9	10	0	32

*anchor item

Items identified by numbers used in most recent set of trial instruments (1978)

Appendix B (continued)

STRUCTURE OF THE COGNITIVE TESTS--POP. B
(Longitudinal)
MARCH 1979

	TEST FORM 1				TEST FORM 2			
	I	II	III	IV	I	II	III	IV
Sets and Functions		1/3						14/8
Number Systems	6/1		1/4		12/2		10/10	
Algebra	10/1	8/8	5/7@	14/13		10/3	2/6	3/4
Geometry		2/7@	11/11	10/14		14/5	4/15@	14/4
Analysis	8/14@	2/14			2/10		12/12 12/14@	
Calculus	11/14 1/14@				5/19 10/18			
Probability, Statistics & Finite Math.		11/20			1/1@			
Item Count	5	4	4	2	5	2	5	3

Items marked @ are anchor items from the First Study.

	Test Form 3				Test Form 4			
	I	II	III	IV	I	II	III	IV
Sets and Functions		2/2@			3/1			
Number Systems	6/2		7/3	10/9		6/3	2/4	
Algebra	New @ 14/3	9/7	8/9		6/6	1/7		10/4
Geometry	11/10		12/9		4/13@		11/12@	
Analysis	10/7	8/1	4/16@		5/5@ 2/13		7/15@	
Calculus			9/20	1/11	2/15			
Probability, Statistics & Finite Math.	5/9@			14/1		2/1		
Item Count	6	3	5	1	5	5	4	1

Items marked @ are anchor items from the First Study.
Test Form 3 includes one anchor item recently added to the pool, without prior piloting in the Second Study. (Form C, Item 18, Husen p. 329)

	Test Form 5 I	II	III	IV	Test Form 6 I	II	III	IV
Sets and Functions	13/1					2/3		
Number Systems	8/4	9/3@		6/4@		8/2	8/3@	
Algebra	11/7	4/3	3/11		11/5 14/7 4/9		3/10@	
Geometry	8/13@		5/10		12/7	3/13		6/13
Analysis		2/5@ 13/19			5/16@	10/5		
Calculus		5/17	6/15@		12/16			
Probability, Statistics & Finite Math.	12/11		13/20			3/19 3/20		
Item Count	5	5	4	1	6	6	2	1

Items marked @ are anchor items from the First Study

	Test Form 7 I	II	III	IV	Test Form 8 (Calculus) I	II	III	IV
Sets and Functions			10/8					
Number Systems	7/15@	13/2		14/2				
Algebra	8/5@	9/4	4/5@					
Geometry	6/10	6/12		13/9				
Analysis	15/9	11/13		13/11	2/12 3/17	4/19@	5/11@ 15/16	7/18
Calculus	1/10		12/18		14/19 15/5	15/4 2/7 14/17	14/15 15/3	15/13 15/2
Probability, Statistics & Finite Math.			6/20					
Item Count	5	4	4	3	4	4	4	3

Items marked @ are anchor items from the First Study

Appendix C

ANCHOR ITEMS

POPULATION A

TOPIC	Final Draft Item No.	Most Recent Trial No.	First Study	Changes
Whole Numbers	A/3	I/2	A/3	Nil
	A/4	F/2	A/17	Nil
	A/5	J/2	B/5	Nil
	A/6	C/4	B/2	Nil
	A/7	I/1	A/2	Nil
	B/6	B/2	A/23	Nil
Fractions	CPCI/1	B/6	B/1	Nil
	A/12	G/4	A/4	Nil
Decimals	CPCI/5	A/30	C/2	Diagram above
	B/3	G/7	B/3	Division above
	C/11	F/7	B/7	Nil
	C/12	J/11	C/7	Nil
Powers	C/13	C/15	A/6	Nil
Formulas and Expressions	CPCI/18	D/25	B/11	Nil
	CPCI/20	J/16	B/23	"cents" added to responses
	A/17	B/15	B/14	Nil
	D/15	-	B/15	Nil
	D/22	A/14	B/9	"and c" added to stem
	D/25	-	A/11	Units
Equations and Inequations	CPCI/22	F/17	A/14	Nil
	B/13	E/21	A/12	"is equivalent to"
	C/19	-	B/10	Nil

TOPIC	Final Draft Item No.	Most Recent Trial No.	First Study	Changes
Geometry	CPCI/32	I/26	A/10	Diagram above
	A/26	B/28	C/20	Line symbols
	A/27	I/24	B/13	Nil
	A/28	Earlier Trial	C/6	Units
	A/31	F/33	C/22	Diagram above
Statistics	B/16	H/33	B/4	Nil
	B/17	F/36	B/8	Nil
	B/24	B/34	A/5	Graph above
Measurement	CPCI/34	I/39	C/8	Scale above
	B/32	B/40	A/8	Units

Total 32 items

Appendix C (Continued)

ANCHOR ITEMS

POPULATION B

TOPIC	Final Draft Item No.	Most Recent Trial No.	First Study	Changes
1 Sets, Relations and Functions				
1.2 Set Operations	3/1	2/2	C/19	X,Y,X P,Q,R
2 Number Systems				
2.1 Common Laws	5/3	9/3	6/7	Nil
2.2 Natural Numbers	5/4	6/4	6/2	Nil
2.4 Real Numbers	6/2	8/3	6/4	Nil
2.5 Complex Numbers	7/2	7/5	5/15	Nil
3 Algebra				
3.1 Polynomials	7/5	8/5	7/2	Nil
3.3 Roots and Radicals	6/6	3/10	8/2	C "and" "or"
3.4 Equations and	1/6	5/7	6/11	Nil
Inequations	7/7	4/5	6/16	Nil
	3/5	=	C/18	Nil
4 Geometry				
4.3 Analytic	1/8	2/7	9/1	Nil
Geometry	4/8	11/12	9/7	Nil
4.5 Vector Method	4/7	4/13	9/13	Notation
4.6 Trigonometry	5/8	8/13	5/9	Figure above
	2/8	4/15	8/4	Nil
5 Analysis				
5.1 Elementary	1/11	8/14	6/9	Nil
Functions	4/9	5/5	5/11	"Range" deleted
	2/11	12/14	5/16	Nil
	3/13	4/16	5/19	Graph above
	4/11	7/15	6/10	Nil
5.2 Properties of Functions	5/10	2/5	5/6	Nil
5.3 Limits and	6/11	5/16	6/15	Nil
Continuity	8/3	4/19	9/15	Nil
5.4 Differentiation	5/12	5/17	9/3	Nil
5.6 Integration	8/4	5/11	9/9	Notation
5.7 Techniques of	1/14	1/14	9/5	Nil
Integration	5/13	6/15	9/4	Nil
5.8 Applications of Integration	8/10	2/17	9/10	Graph above
6 Probability and Statistics				
6.1 Probability	2/15	1/1	8/3	Nil

TOTAL 30 Items

APPENDIX D

Final Form, Population A Test (Cross-Sectional)

CORE FORM

Item	Code No.	Cont.	Beh.	Diff.	Item	Code No.	Cont.	Beh.	Diff.
1	014	102	II	53	21	033	303	II	68
2	027	206	I	31	22	018	106	IV	44
3	006	003	III	54	23	010	006	I	57
4	012	101	I	65	24	037	404	III	37
5	030	209	I	44	25	016	104	III	53
6	025	203	III	39	26	008	004	II	48
7	032	302	II	50	27	035	304	I	34
8	040	404	III	34	28	021	201	II	56
9	022	202	I	70	29	029	207	III	41
10	015	104	I	57	30	036	401	I	58
11	039	404	I	64	31	002	001	IV	58
12	017	106	I	63	32	023	202	IV	45
13	031	212	III	64	33	009	004	I	55
14	004	002	II	76	34	011	008	II	30
15	034	303	III	66	35	007	003	II	64
16	020	110	III	42	36	024	203	III	62
17	003	002	I	68	37	038	402	II	43
18	005	003	I	52	38	028	207	I	56
19	026	204	III	69*	39	019	107	II	64
20	001	001	I	85	40	013	101	III	51
Mean Diff.						(N = 40)			54.25

* First Survey

Final Form, Population A Test (Cross-Sectional) - Continued

FORM A					FORM B				
Item	Code No.	Cont.	Beh.	Diff.	Item	Code No.	Cont.	Beh.	Diff.
1	052	104	III	43	1	086	106	I	45
2	055	107	III	52	2	100	401	I	54
3	050	101	III	83	3	093	206	I	30
4	062	208	III	39	4	085	104	II	63
5	072	403	III	47	5	088	105	I	28
6	173	215	I	-	6	095	208	II	51
7	067	304	III	66	7	104	404	III	62
8	169	207	II	-	8	074	001	IV	36
9	057	201	III	71	9	073	001	I	87
10	044	002	III	69	10	176	205	III	-
11	043	002	II	52	11	080	005	I	60
12	069	402	II	63	12	078	003	III	43
13	058	202	II	36	13	083	102	II	53
14	064	211	I	84	14	096	204	II	38
15	051	102	IV	45	15	090	202	I	54
16	046	004	I	62	16	089	201	II	71
17	048	006	I	74	17	101	401	III	72
18	063	209	I	30	18	097	302	I	34
19	065	301	II	58	19	099	304	IV	49
20	047	004	II	70	20	171	215	II	-
21	045	003	II	40	21	103	404	III	65
22	049	101	I	70	22	081	008	II	46
23	071	404	II	49	23	077	003	II	61
24	066	303	III	74	24	079	004	III	42
25	070	404	I	49	25	091	203	I	81
26	060	204	I	57	26	075	002	II	65
27	041	001	I	88	27	076	002	III	46
28	056	110	I	68	28	087	106	III	35
29	053	105	I	54	29	084	103	I	28
30	061	207	II	59	30	102	402	II	49
31	068	401	I	80	31	098	303	II	59
32	054	106	II	58	32	092	204	III	67
33	042	001	I	79	33	094	207	III	50
34	059	202	IV	33	34	082	101	II	41
Mean Diff.		(N = 32)		59.44			(N = 32)		52.03

Final Form, Population A Test (Cross-Sectional) - Continued

FORM C					FORM D				
Item	Code No.	Cont.	Beh.	Diff.	Item	Code No.	Cont.	Beh.	Diff.
1	114	101	IV	45	1	159	212	II	52
2	130	304	I	43	2	155	202	III	58
3	175	207	II	-	3	141	003	III	67
4	108	003	I	45	4	154	201	III	42
5	133	401	III	59	5	146	009	I	77
6	136	404	II	33	6	140	003	I	54
7	125	204	II	30	7	143	004	III	79
8	132	304	I	50	8	164	402	I	63
9	105	001	II	38	9	153	110	IV	55
10	129	303	III	34	10	150	105	I	59
11	135	403	II	48	11	163	306	I	42
12	109	003	III	74	12	147	102	I	48
13	126	207	II	40	13	145	008	II	60
14	131	302	II	33	14	138	001	III	74
15	116	104	I	44	15	139	002	I	61
16	107	002	I	63	16	144	005	IV	54
17	119	107	I	60	17	168	404	IV	21
18	121	201	I	66	18	166	404	I	29
19	128	202	III	39	19	148	104	II	60
20	110	004	III	58	20	167	402	III	51
21	120	110	II	43	21	152	107	I	75
22	112	006	I	60	22	137	001	II	63
23	113	101	I	52	23	162	304	II	58
24	172	103	II	-	24	174	215	II	-
25	127	208	III	45	25	157	208	II	55
26	124	203	II	39	26	142	004	II	29
27	123	202	III	34	27	170	215	III	-
28	117	106	II	54	28	165	403	II	45
29	134	402	III	36	29	156	204	III	55
30	122	202	II	79	30	151	106	I	40
31	106	001	III	71	31	149	104	III	72*
32	115	104	I	28*	32	158	209	III	60
33	111	005	II	78	33	160	303	II	57*
34	118	106	I	50	34	161	303	II	48*
Mean Diff.		(N = 32)		49.09			(N = 32)		55.09

* First Survey

Final Form, Population A Test (Cross-Sectional) - Continued

CORE FORM

Content	Level I	II	III	IV
000	5	5	1	1
100	3	1	3	1
200	4	1	6	1
300	1	2	0	0
400	2	1	2	0
Total	15	10	12	3

FORM A / FORM B

Content	Form A Level I	II	III	IV	Form B Level I	II	III	IV
000	4	3	1	0	2	3	3	1
100	3	1	3	1	3	3	1	0
200	4	3	2	1	3	4	3	0
300	0	1	2	0	1	1	0	1
400	2	2	1	0	1	1	3	0
Total	13	10	9	2	10	12	10	2

FORM C / FORM D

Content	Form C Level I	II	III	IV	Form D Level I	II	III	IV
000	3	2	3	0	3	3	3	1
100	5	3	0	1	4	1	1	1
200	1	5	3	0	0	3	5	0
300	2	1	1	0	1	4	0	0
400	0	2	2	0	2	1	1	1
Total	11	13	9	1	10	11	10	3

APPENDIX E

FINAL FORM

POPULATION B TEST (CROSS-SECTIONAL)

Population B - Final Forms - Cross-sectional

FORM 1

Item	Code No.	Cont.	Beh.	Diff.
1	027	5.1	III	56
2	096	3.3	II	64
3	099	4.1	I	62
4	103	5.6	III	57
5	015	5.7	II	64
6	129	7.1	III	--
7	028	5.5	I	61
8	059	6.2	I	82
9	093	2.5	II	39
10	070	5.2	II	59
11	128	4.2	III	--
12	079	3.1	I	72
13	032	2.3	I	52
14	005	3.5	III	73
15	076	1.4	II	59
16	100	4.5	IV	35
17	084	4.6	II	72

FORM 2

Item	Code No.	Cont.	Beh.	Diff.
1	002	2.1	I	83
2	006	3.1	III	59
3	134	4.5	I	--
4	072	5.4	II	63
5	055	5.1	II	44
6	086	5.3	I	57
7	116	5.6	II	41
8	050	3.4	III	45
9	069	4.3	III	83
10	058	5.8	II	48
11	001	1.2	II	86
12	045	6.1	I	79
13	127	3.	III	--
14	066	3.1	I	91
15	104	5.5	I	--
16	003	2.1	III	49
17	080	3.3	I	67

FORM 3

Item	Code No.	Cont.	Beh.	Diff.
1	082	3.4	I	84
2	062	2.5	I	33
3	115	5.8	II	45
4	057	5.5	III	35
5	074	6.2	II	31
6	031	1.2	II	87
7	033	2.2	III	49
8	049	3.3	I	66
9	123	4.	III	--
10	008	3.1	II	35
11	135	3.	II	--
12	020	3.6	III	45
13	056	5.1	III	46
14	061	1.4	I	70
15	106	5.4	I	73
16	111	5.5	IV	42
17	009	4.3	III	73

FORM 4

Item	Code No.	Cont.	Beh.	Diff.
1	046	1.4	I	97
2	019	3.1	II	77
3	025	4.6	I	--
4	113	5.7	II	--
5	091	1.4	III	72
6	124	3.	I	--
7	023	4.6	III	29
8	131	4.	II	--
9	048	2.3	III	62
10	054	5.1	II	75
11	089	6.1	III	38
12	065	3.2	I	74
13	092	2.5	I	83
14	030	6.1	I	55
15	012	5.1	III	53
16	119	5.5	IV	51
17	110	5.3	III	44

Content	Form 1 I	Form 1 II	Form 1 III	Form 1 IV	Form 2 I	Form 2 II	Form 2 III	Form 2 IV	Form 3 I	Form 3 II	Form 3 III	Form 3 IV	Form 4 I	Form 4 II	Form 4 III	Form 4 IV
1		1				1			1	1			1		1	
2	1	1			1		1		1		1		1		1	
3	1	1	1		2		3		2	2	1		2	1		
4	1	1	1	1	1		1				2		1	1	1	
5	2	1	2		2	4			1	1	2	1		2	2	1
6	1				1					1			1		1	
7			1													
Total	6	5	5	1	7	5	5	-	5	5	6	1	6	4	6	1

Mean Difficulties: Form 1 (15) 61; Form 2 (14) 64; Form 3 (15) 54; Form 4 (13) 62

Item	FORM 5 Code No.	Cont.	Beh.	Diff.	FORM 6 Code No.	Cont.	Beh.	Diff.	FORM 7 Code No.	Cont.	Beh.	Diff.	FORM 8 Code No.	Cont.	Beh.	Diff.
1	078	2.2	II	30	004	3.5	I	73	081	3.3	III	65	078	5.1	II	87
2	064	2.2	IV	55	101	5.3	II	65	083	4.3	I	66	075	6.3	III	31
3	067	3.3	III	73	077	2.4	III	75	105	7.1	III	33	029	5.8	I	57
4	109	5.6	III	52	014	5.4	I	62	018	2.5	III	50	036	3.2	I	86
5	035	3.4	II	--	010	4.6	III	47	043	5.1	III	42	136	4.5	II	--
6	114	5.6	II	28	130	4	I	--	088	5.5	III	48	120	5.8	I	60
7	039	4.6	I	59	024	4.3	IV	73	038	3.5	III	61	063	2.1	II	59
8	132	4.	II	--	053	4.3	III	47	108	5.3	II	29	016	1.4	IV	72
9	090	7.1	II	--	068	4.6	I	40	126	4.2	II	--	007	3.4	IV	49
10	037	3.5	III	73	060	6.2	III	49	017	3.5	I	45	125	7.1	II	--
11	102	5.2	IV	--	094	2.3	IV	68	117	5.5	III	62	052	4.5	I	65
12	040	4.5	III	42	097	3.4	III	53	041	5.1	I	83	042	5.2	II	56
13	112	5.5	I	38	044	5.8	III	38	011	4.3	I	47	021	3.4	IV	49
14	022	4.5	II	58	121	4.3	I	--	133	4.2	I	--	085	4.6	IV	28
15	118	5.4	III	40	051	3.5	IV	53	098	4.5	II	72	026	5.1	III	68
16	095	3.1	I	46	073	5.7	III	21	047	2.4	III	57	034	2.5	IV	49
17	122	5	II	--	107	5.6	I	76	013	5.2	I	80	071	5.1	II	53

Content	FORM 5 I	II	III	IV	FORM 6 I	II	III	IV	FORM 7 I	II	III	IV	FORM 8 I	II	III	IV
1																1
2		1		1			1	1			2			1		1
3	1	1	2		1		1	1	1		2		1			2
4	1	2	1		3		2	1	3	2			1	1		1
5	1	2	2	1	2	1	2		2	1	3		2	3	1	
6							1								1	
7		1									1			1		
Total	3	7	5	2	6	1	7	3	6	3	8	--	4	6	2	5
M ... tres	(12) 50				(15) 57				(15) 56				(15) 58			

139b