Secondary Analysis of the TIMSS Data

David F. Robitaille and Albert E. Beaton (Eds.)

Kluwer Academic Publishers

SECONDARY ANALYSIS OF THE TIMSS DATA

This page intentionally left blank

Secondary Analysis of the TIMSS Data

Edited by

David F. Robitaille
University of British Columbia,
Vancouver, BC, Canada

and

Albert E. Beaton
Boston College,
Chestnut Hill, MA, U.S.A

KLUWER ACADEMIC PUBLISHERS
NEW YORK, BOSTON, DORDRECHT, LONDON, MOSCOW

eBook ISBN: 0-306-47642-8
Print ISBN: 1-4020-0859-7

Created in the United States of America

Visit Kluwer Online at: http://kluweronline.com
and Kluwer's eBookstore at: http://ebooks.kluweronline.com

TABLE OF CONTENTS

Chapter		Page
	PREFACE	ix
	PART 1: INTRODUCTION	
1	TIMSS in Context: Assessment, Monitoring, and Moving Targets / *Hans Wagemaker*	3
2	TIMSS: A Brief Overview of the Study / *David F. Robitaille and Albert E. Beaton*	11
	PART 2: FOCUS ON MATHEMATICS	
3	Analyzing Student Responses in Mathematics Using Two-Digit Rubrics / *John A. Dossey, Chancey O. Jones, and Tami S. Martin*	21
4	From SIMS to TIMSS: Trends in Students' Achievement in Mathematics / *David F. Robitaille and Alan R. Taylor*	47
5	The Impact of TIMSS on the Mathematics Standards Movement in the United States / *John A. Dossey and Mary M. Lindquist*	63
6	TIMSS Mathematics Results: A Japanese Perspective / *Eizo Nagasaki and Hanako Senuma*	81
7	TIMSS, Common Sense, and the Curriculum / *Geoffrey Howson*	95
8	Advanced Mathematics: Curricula and Student Performance / *Geoffrey Howson*	113
	PART 3: FOCUS ON SCIENCE	
9	Exploring Population 2 Students' Ideas about Science / *Marit Kjærnsli, Carl Angell, and Svein Lie*	127

10	Science Achievement: A Czech Perspective / *Jana Paleckova and Jana Strakova*	145
11	TIMSS Science Results For Hong Kong: an Ailing Dragon with a British Legacy / *Nancy Law*	157
12	Science Achievement: A Russian Perspective / *Galina Kovalyova and Natalia Naidenova*	177
13	TIMSS Science Results Seen from a Nordic Perspective / *Marit Kjærnsli and Svein Lie*	193
	PART 4: FOCUS ON CROSS-CURRICULAR ISSUES	
14	Separating School, Classroom and Student Variances and their Relationship to Socio-Economic Status / *Albert E. Beaton and Laura M. O'Dwyer*	211
15	On the Relationship Between Mathematics and Science Achievement in the United States / *Min Li, Richard J. Shavelson, Haggai Kuperintz and Maria Araceli Ruiz-Primo*	233
16	Students' Attitudes and Perceptions / *Edward W. Kifer*	251
17	Analyzing Gender Differences for High Achieving Students on TIMSS / *Ina V. S. Mullis and Steven E. Stemler*	277
18	Investigating Correlates of Mathematics and Science Literacy in the Final Year of Secondary School / *Jesse L. M. Wilkins, Michalinos Zembylas, and Kenneth J. Travers*	291
19	Indicators of ICT in Mathematics: Status and Covariation with Achievement Measures / *Willem J. Pelgrum and Tjeerd Plomp*	317
20	Extra-School Instruction in Mathematics and Science / *Richard M. Wolf*	331
21	Teachers' Sources and Uses of Assessment Information / *Thomas Kellaghan and George F. Madaus*	343
	PART 5: FOCUS ON METHODOLOGY	
22	Extending the Application of Multilevel Modeling to Data from TIMSS / *Laura M. O'Dwyer*	359
23	Application of the Scale Anchoring Method to Interpret the TIMSS Achievement Scales / *Dana L. Kelly*	375
24	Effects of Adaptations on Comparability of Test Items and Test Scores / *Kadriye Ercikan and Tanya McCreith*	391

PART 6: CONCLUSION

25	A Look Back at TIMSS: What Have We Learned about International Studies / *Albert E. Beaton and David F. Robitaille*	409
	APPENDIX	419
	INDEX	423

This page intentionally left blank

PREFACE

Researchers who participate in IEA studies have a unique opportunity to work collaboratively with their counterparts from many different countries and disciplinary backgrounds over a period of several years on questions of shared academic interest. Once the data for a given study have been collected and the first round of international reports published, however, opportunities for that kind of collaboration tend to be much less frequent.

A major strength of IEA studies compared to other large-scale, international studies is that they are classroom based, thereby making it possible for researchers and policy makers to investigate linkages between students' achievement and a wide range of variables. Those variables could be related to instructional practices, to students' and teachers' background and attitudes, to school organizational patterns, or to opportunity to learn, to name a few. The research questions that TIMSS was designed to address make it clear that these kinds of relational, multi-variate analyses were among the major goals of the project.

The international reports of the TIMSS–95 results that were published by the International Study Center at Boston College between 1996 and 1999 were intended to provide comprehensive coverage of the basic findings of the study. They were not intended to provide in-depth analyses of research and policy issues; instead, their main purpose was to make the basic findings of the study widely available in a timely manner. This they certainly did.

The goal of the present volume is to make available the findings from a number of secondary analyses that researchers in many of the TIMSS countries have carried out since the data were collected in 1995. Thanks to the financial support provided by the U. S. National Science Foundation under Grant #REC-9815180, it has been possible to carry out some secondary analyses, and the results of those analyses are the focus of this volume. The grant made it possible to bring together 37 scholars from 10 countries for two meetings to discuss the structure of the volume and to provide feedback to them regarding their planned analyses. The grant also provided funds to provide technical support for authors in carrying out their analyses and for editing the papers they produced. Any opinions, findings, conclusions, or recommendations expressed in this book are those of the authors and do not necessarily reflect the views of the National Science Foundation.

The topics covered in this set of papers are almost as varied as the researchers who wrote them, and they illustrate the range of investigations that this kind of data makes possible. For the sake of convenience, the papers have been partitioned into

several sections on the assumption that some readers would be more interested in some topics than in others. The first, or introductory section of the book includes 2 chapters and is designed to provide a brief introduction to TIMSS as a whole as well as to this volume. The second section (Chapters 3 to 8) focuses on papers related to mathematics; the third section (Chapters 9 to 12), on science; and the fourth (Chapter 13 to 19), on topics that are more cross-curricular in nature. The fifth section (Chapters 20 to 24) contains a set of papers related to measurement and methodological topics. The sixth and last section consists of closing comments from the editors regarding a number of lessons learned from TIMSS and some suggestions for further research.

The two papers in Part I provide an introduction to the volume. In Chapter 1, Hans Wagemaker, the executive director of IEA, highlights the importance of international comparisons in education and the role of IEA studies in that effort over the past 40 years. Chapter 2, written by David Robitaille from the University of British Columbia and A1 Beaton of Boston College, both of whom were heavily involved in all phases of TIMSS, is a brief introduction to the study for readers who are not familiar with its scope and extent.

Part 2 consists of six chapters focusing on aspects of the mathematics component of TIMSS. John Dossey (Illinois State University), Chancey Jones (Educational Testing Service), and Tami Martin (Illinois State University) present an analysis of students' responses to constructed-response items, using the two-digit scoring codes developed for use in the study. The next paper, from David and Alan Taylor of the University of British Columbia, summarizes changes in students' achievement results over a period of about 20 years between SIMS, the second mathematics study, and TIMSS. John Dossey (Illinois State University) and Mary Lindquist (Columbus State University) discuss the influence of TIMSS on the development and dissemination of the curriculum and evaluation standards developed by the National Council of Teachers of Mathematics. Eizo Nagasaki and Hanako Senuma from the National Institute for Educational Research in Japan present an analysis of the TIMSS mathematics results from their perspective in Japan. In the next two papers, Geoffrey Howson of the University of Southampton, shares his insights about the curricular and instructional implications of the Population 2 mathematics results in Chapter 7 and of the Population 3 results in Chapter 8.

Part 3 consists of 5 chapters related to the TIMSS science results. In Chapter 9, Svein Lie and his colleagues from the University of Oslo explore students' understanding of a number of fundamental concepts in science. Chapters 10 through 13 provide reflections on the science achievement results from a range of international perspectives. These include the Czech Republic Jana Paleckova and Jana Strakova), Hong Kong (Nancy Law), Russia (Galena Kovalyova), and Scandinavia (Marit Kjærnsli and Svein Lie). In each case, the authors identify and discuss the implications of the science achievement results for informing the debate about how to improve the teaching and learning of science.

The seven chapters, Chapters 14 through 20, included in Part 4 discuss a range of issues that relate to teaching and learning, but not necessarily to mathematics or science specifically. For lack of a better term, the section is described as focusing on cross-curricular issues.

In Chapter 14, Al Beaton and Laura O'Dwyer of Boston College address separating school and classroom variance using the TIMSS data. Four scholars from UCLA discuss the correlation between students' achievement in mathematics and in science in Chapter 15. Their analysis focuses on results from the United States only. Skip Kifer from the University of Kentucky provides an analysis of the student attitude data in Chapter 16. In Chapter 17 Ina Mullis and Steve Stemler from the International Study Center at Boston College focus on an analysis of gender differences in achievement in TIMSS. Chapter 18, written by Jay Wilkins, Michalinos Zembylas, and Ken Travers, provides insight into the design of and into senior secondary students' performance on the TIMSS mathematics and science literacy study. In Chapter 19, Hans Pelgrum and Tjeerd Plomp from the University of Twente summarize finding from TIMSS having to do with the impact of technology on the teaching and learning of mathematics and science. Chapter 20 was written by Dick Wolf of Teachers' College. His paper focuses on the importance of out-of-school tutoring or coaching in various countries. Chapter 21, by Tom Kellaghan of the Educational Research Centre in Dublin and George Madaus of Boston College, use data from the TIMSS teacher questionnaires to examine the sources of teachers' information about issues related to assessment and evaluation.

Part 5 of the volume focuses on issues related research methodology. In Chapter 22, Laura O'Dwyer from Boston College discusses a new technique based for estimating between-classroom variance using what she describes as a "pseudo-classroom" approach. In Chapter 23 Dana Kelly from the American Institutes for Research describes her work on the development of international benchmarks of student achievement through scale anchoring analysis. In Chapter 24, Kadriye Ercikan and Tanya McCreith from the University of British Columbia use differential-item-functioning technology to explore the impact of translation effects on item difficulty in selected countries.

Part 6 consists of a brief concluding chapter by the editors. The goal of this chapter is not to serve as a summary of what has gone before, but rather to provide an opportunity for the editors to reflect on some of the lessons learned from TIMSS, to speculate about the kinds of research that remain to be done, and to put forward a few suggestions for the consideration of researchers who will be doing these kinds of studies in the future.

As editors of this volume, we are grateful to many individuals who helped us bring this task to a successful conclusion. We are deeply indebted to Larry Suter and the U. S. National Science Foundation for the moral and financial support provided to the authors and us throughout the process of bringing this book to publication. We are, of course, very grateful to our many authors for their patience in dealing with editorial demands and for their prompt responses to our many questions and editorial suggestions. We also grateful for the support extended to us by IEA, and particularly by its Executive Director, Hans Wagemaker, throughout the process.

On the east coast of the United States, at Boston College, the project had the support of the Center for the Study of Testing, Evaluation, and Educational Policy. Their staff members were of great assistance to us in a number of ways. We also had

great technical support from Stacy Raczek, a doctoral student at Boston College, whose mastery of the software was phenomenal.

On the west coast, at the University of British Columbia, Katy Ellsworth and Bonnie Davidson took on major responsibility for getting the manuscript ready for submission to the publisher. Katy's involvement tapered off somewhat toward the end of the project as she focused on the forthcoming birth of her second child. Bonnie filled the gap admirably, and we are grateful to both of them for their contributions.

David Robitaille
Al Beaton

1 INTRODUCTION

This page intentionally left blank

Chapter 1

TIMSS IN CONTEXT: ASESSMENT, MONITORING, AND MOVING TARGETS

Hans Wagemaker

Writing in the early 1970s one of the architects of the early IEA studies outlined the dilemma facing practitioners and policymakers in education alike.

> At all levels in an educational system, from the teacher in the classroom, through the administrator to the policymaker, decisions have continually to be made most of the time on the basis of very little factual information. (Postlethwaite, 1974).

Educational policy is formulated and implemented at all levels of the education system even where system-level constraints such as a centralized curriculum restrict what schools and teachers might do. Discretion at the school and classroom level always remains. How and on what basis policymakers, administrators, and teachers make decisions in the educational arena is at the heart of international comparative studies of education like TIMSS. In order to more fully understand the significance of studies like TIMSS it is worth considering the way in which interest in and the impact of studies like TIMSS have evolved.

Over the last 15 years most of the developed countries of the world have initiated or experienced significant reforms in education and the wider public sector (The World Bank, 1999). Similarly, in many low- to middle-income countries, educational reform as a means of enhancing social and economic well-being has received increasing amounts of attention. This is in part attributable to the almost universal recognition that the performance of a country's educational system is a key element in establishing a nation's competitive advantage in an increasingly global economy. Education is conceived of as being implicated in a country's economic, social, and personal development and is considered one of the key means whereby inequities, social and economic, can be reduced. Perhaps the most dramatic expression of this sentiment is contained in the report from the United States, *A Nation at Risk,* in which the authors point to the threat of economic decline as supplanting the past threat of aggressor nations (United States National Commission

D.F. Robitaille and A.E. Beaton (eds.), Secondary Analysis of the TIMSS Data, 3–10.

on Excellence in Education, 1983). Education and the decline in educational standards were cited as the cause of economic decline in the face of intensified global competition. The authors write:

> If an unfriendly foreign power had attempted to impose on America the mediocre educational performance that exists today, we might well have viewed it as an act of war. As it stands, we have allowed this to happen ourselves ... we have, in effect, been committing an act of unthinking, unilateral educational disarmament. (United States National Commission on Excellence in Education, 1983, p.5)

Although a model which ascribes economic decline in a simplistic way to a decline in educational standards is likely to be of limited value in addressing or understanding either educational or economic policy concerns, this debate served to draw attention to real concerns about educational performance, not only in the U. S. but also in may OECD countries. It is the concern for excellence (together with concerns of equity and efficiency) that has given rise to greater intensity of focus on education and educational policy development.

While education has been receiving an increased priority in the public policy arena in many countries, it has also been facing the reality that, like many other areas of public spending, there are real limits to the amount of funding that is available for educational development. What funding is available is accompanied by increasing demands for accountability and a better understanding of the relationship between educational expenditure and educational outcomes. The fullest and perhaps most extreme expression of these concerns is reflected in publications like *Reinventing Government* (Osborne & Gaebler, 1993) in which the authors argue for an educational marketplace that should be shaped by the twin imperatives of efficiency and effectiveness. The implicit argument is that increased provision and improved instructional quality are likely to produce greater numbers of better-prepared students, which in turn will result in a more internationally competitive and better-prepared workforce. The role that TIMSS might play in such an argument is to place the focus more narrowly on the assessment of quality in mathematics and science, and presumably, therefore, on the production of more productive and high-quality scientists, mathematicians, and engineers.

In general, however, what is reflected in the kinds of concerns expressed above is a shift in focus from managing issues related to the expansion of educational systems in terms of student numbers, to one of managing issues of quality and excellence. In the case of those countries in what might be described as a less advanced stage of educational development, this has meant not surrendering to the imperatives of educational expansion at the expense of considerations of quality. The change in emphasis is noted by Tuijnman and Postlethwaite (1994) who argue that, while the history of large scale assessment dates back to the early 1960s, there was a significant development toward a more systematic focus on national

monitoring with the release of reports such as *A Nation at Risk,* the release of the results of IEA's Second International Science Study, and later, again in the United States, the report from the conference of the governors of the 50 states in Charlottesville, Virginia, which sought to frame national goals for education with a strong emphasis on quality. In short, investment in education and the related policy development, it was argued, could no longer be carried out as an act of faith.

As interest in global competitiveness and local accountability has increased, so too has interest in international comparisons of educational performance. What then is the significance of, and what are the benefits of participating in international comparisons of educational achievement, and how does the Third International Mathematics and Science Study meet these expectations?

TIMSS is intended to monitor the success of mathematics and science instruction and the context in which it occurs in each of the countries that participated in the project. Three major conceptual elements drive the TIMSS design. These elements include the intended curriculum (the curriculum as described at the policy level), the implemented curriculum (the curriculum as students experience it at the school and classroom level), and the attained curriculum (the curriculum as represented by student outcomes). Through the mechanisms of a curriculum analysis, a video study, achievement tests, and background questionnaires that gathered information from schools, teachers, and students, the conceptual design was realized, providing a unique opportunity to observe and examine how mathematics and science instruction is provided in some 40 countries. What is significant is that the TIMSS design provided for an examination of those policy variables related to schooling, curriculum, and instruction that are affected by policy intervention. Furthermore, it established international benchmarks for achievement and key policy variables that allow countries to monitor their performance in an increasingly global community.

While the emphasis in comparative studies of educational achievement is often seemingly focused primarily on the achievement data, the interpretation of such system level data is not straightforward. The significance of the extensive data collected by the multiple strategies employed by TIMSS lies in the fact that countries that do not take into account the differences in the respective education systems when introducing policy reform based on comparative data risk not only disappointment, but also the possibility of developing polices that are potentially counter-productive in addressing perceived educational needs.

Moreover, the data collected through the background questionnaires allows policymakers to address particular policy needs and concerns related not only to the quantity, quality, and content of mathematics and science instruction but also to identifying factors that may be linked to achievement or to sub-populations of national importance (such as gender and ethnicity). While it is not always possible in the international context to collect data on, for example sub-groups of interest that are internationally comparable (e.g., ethnicity), the TIMSS design permits the

collection of these variables as international options. For example, the TIMSS reports (Beaton et al., 1996) included information not only on such things as the characteristics of the students' home environment (e.g., books in the home), the characteristics of instructional practices (e.g., classroom organization) but also on some of the affective characteristics of the student populations (e.g., student attitudes to mathematics and science) and their relationship to achievement.

While much of what studies like TIMSS do is to describe "what is" in terms of how education is practiced in a country (the within-country perspective), the power of such studies is most fully realized when the international context they provide is considered (the between-country perspective). Given the differences in the ways in which education is organized and practiced across cultures and societies, a comparative perspective such as that provided by TIMSS not only enables an understanding of its many forms, but also serves to expand a nation's horizon as to what might be possible. As Foshay et al. (1962) noted:

> If custom and law define what is educationally allowable within a nation, the educational systems beyond one's national boundaries suggest what is educationally possible.

Identifying models or practices of education from countries around the world as a means of reflecting on one's own practice and experience is, arguably, a key function of international comparative studies like TIMSS.

TACTICAL VERSUS STRATEGIC APPROACH

While the evidence that there is an increasing interest in and awareness of the need to invest in international comparative studies of educational achievement has been presented, there is a more recent demand for the creation of international benchmarks against which a country's performance may be measured. This is associated with a growing awareness of the need to move from the process of tactical decision making and ad hoc participation in studies like TIMSS to a more strategic investment that recognizes the dynamic nature of change in the educational environment. As noted in *Education Sector Strategy* (The World Bank, 1999), market economies and a constantly changing educational, political, and economic environment prevail in countries accounting for over 80 percent of the world's population. Education is deemed vital, with those who can compete with the best having an enormous advantage in a faster paced world economy over those who are less well prepared. Globalization of markets and factors such as knowledge exacerbate these impacts. Given this dynamic it is not surprising that a demand has emerged for the regular monitoring of educational quality, particularly for those economies engaged in strategic educational reform.

Repeated monitoring over time, provided that the tests in different years can be linked, can provide evidence not only of changes in levels of performance but also