



Public Examinations Examined

Thomas Kellaghan and
Vincent Greaney



WORLD BANK GROUP

Public Examinations Examined

Public Examinations Examined

Thomas Kellaghan and Vincent Greaney



WORLD BANK GROUP

© 2020 International Bank for Reconstruction and Development / The World Bank
1818 H Street NW, Washington, DC 20433
Telephone: 202-473-1000; Internet: www.worldbank.org

Some rights reserved.

1 2 3 4 22 21 20 19

This work is a product of the staff of The World Bank with external contributions. The findings, interpretations, and conclusions expressed in this work do not necessarily reflect the views of The World Bank, its Board of Executive Directors, or the governments they represent. The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

Nothing herein shall constitute or be considered to be a limitation upon or waiver of the privileges and immunities of The World Bank, all of which are specifically reserved.

Rights and Permissions



This work is available under the Creative Commons Attribution 3.0 IGO license (CC BY 3.0 IGO) <http://creativecommons.org/licenses/by/3.0/igo>. Under the Creative Commons Attribution license, you are free to copy, distribute, transmit, and adapt this work, including for commercial purposes, under the following conditions:

Attribution—Please cite the work as follows: Kellaghan, Thomas, and Vincent Greaney. 2020. *Public Examinations Examined*. Washington, DC: World Bank. doi:10.1596/978-1-4648-1418-1. License: Creative Commons Attribution CC BY 3.0 IGO.

Translations—If you create a translation of this work, please add the following disclaimer along with the attribution: *This translation was not created by The World Bank and should not be considered an official World Bank translation. The World Bank shall not be liable for any content or error in this translation.*

Adaptations—If you create an adaptation of this work, please add the following disclaimer along with the attribution: *This is an adaptation of an original work by The World Bank. Views and opinions expressed in the adaptation are the sole responsibility of the author or authors of the adaptation and are not endorsed by The World Bank.*

Third-party content—The World Bank does not necessarily own each component of the content contained within the work. The World Bank therefore does not warrant that the use of any third-party-owned individual component or part contained in the work will not infringe on the rights of those third parties. The risk of claims resulting from such infringement rests solely with you. If you wish to reuse a component of the work, it is your responsibility to determine whether permission is needed for that reuse and to obtain permission from the copyright owner. Examples of components can include, but are not limited to, tables, figures, or images.

All queries on rights and licenses should be addressed to World Bank Publications, The World Bank Group, 1818 H Street NW, Washington, DC 20433, USA; e-mail: pubrights@worldbank.org.

ISBN (paper): 978-1-4648-1418-1

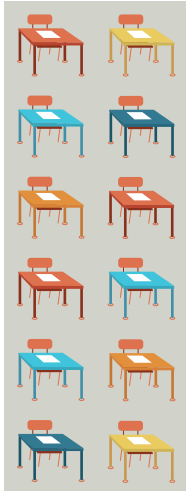
ISBN (electronic): 978-1-4648-1419-8

DOI: 10.1596/978-1-4648-1418-1

Cover design: Debra Naylor, Naylor Design.

Cover illustration: Danielle Willis.

Library of Congress Cataloging-in-Publication Data has been requested.



CONTENTS

<i>Foreword</i>	<i>xi</i>
<i>Preface</i>	<i>xiii</i>
<i>About the Authors</i>	<i>xvii</i>
1 INTRODUCTION.....	1
Overview	1
Public Examinations, Classroom Assessment, and System Assessment	4
Variety in Examination Systems.....	6
Major Public Examination Traditions	8
The Contents of This Book	11
Notes	12
References	13
2 CHARACTERISTICS AND FUNCTIONS OF PUBLIC EXAMINATIONS.....	17
Introduction.....	17
Characteristics of Public Examinations	17
Functions of Examinations.....	19
Differences among Examination Systems.....	28
Conclusion.....	35
Notes	37
References	38
3 A BRIEF HISTORY OF WRITTEN EXAMINATIONS	43
Introduction.....	43
Examinations in Imperial China.....	44
Examinations in Europe, 1400–	49
Public Examinations in Developing Economies	54
Examinations in Former Socialist Countries.....	56
Examinations in Latin America.....	58

Examinations in the United States	59
International Examinations	63
Conclusion.....	65
Notes	67
References	68
 4 THE CONSTRUCTION AND SCORING OF PUBLIC EXAMINATIONS	75
Introduction.....	75
The Examination Syllabus	75
Constructing the Examination	77
The Form of Examinee Response.....	81
Question Paper Production and Printing	86
Marking of Answer Scripts.....	88
Use of Information Technologies.....	93
Conclusion.....	95
References	96
 5 ADVANTAGES AND DISADVANTAGES OF PUBLIC EXAMINATIONS	99
Introduction.....	99
Advantages of Public Examinations	100
Disadvantages of Public Examinations	104
Conclusion.....	115
Notes	116
References	117
 6 THE PROVISION OF FEEDBACK FROM EXAMINATIONS	123
Introduction.....	123
Guidance Feedback.....	124
Incentive Feedback	131
Conclusion.....	138
Notes	139
References	139
 7 THE VALIDITY OF PUBLIC EXAMINATIONS	143
Introduction.....	143
Content Validity	145
Other Aspects of Validity.....	146
Threats to Validity	149
Conclusion.....	158
Notes	158
References	159
 8 THE RELIABILITY OF PUBLIC EXAMINATIONS	163
Introduction.....	163
Factors Associated with Examinees	164
Factors Associated with Examiners	165
Factors Associated with the Subject and How It Is Scored	165
Improving Examination Marker Reliability	168
Conclusion.....	168
Note	169
References	170

9	STANDARDS IN PUBLIC EXAMINATIONS	173
	Introduction.....	173
	Comparability of Grades in Different Subjects or Examinations.....	174
	Change in Standards over Time	175
	Dealing with Comparability Problems	178
	Defining Performance Standards	180
	Public Confidence.....	183
	Conclusion.....	184
	Notes	185
	References	185
10	EQUITY ISSUES IN PUBLIC EXAMINATIONS	191
	Introduction.....	191
	Gender.....	193
	Socioeconomic Status	196
	Place of Residence	199
	Type of School.....	203
	Ethnic or Language Group Membership	204
	Conclusion.....	207
	Notes	209
	References	209
11	THE INTEGRITY OF PUBLIC EXAMINATIONS AND MALPRACTICE.....	217
	Introduction.....	217
	Forms of Malpractice	218
	Frequency of Malpractice.....	222
	Reasons for Malpractice	223
	Strategies to Discourage and Detect Malpractice.....	224
	Ways to Control and Prevent Malpractice.....	227
	Conclusion.....	231
	Notes	233
	References	234
12	PUBLIC EXAMINATIONS PROVISION FOR CANDIDATES WITH DISABILITIES OR DIVERSE EDUCATIONAL NEEDS	239
	Introduction.....	239
	Identification of Students with Disabilities or Special Needs and Implications of Accommodation	241
	Exemptions.....	242
	Types of Accommodation	243
	Issues Arising from the Provision of Accommodations for Disabilities or Diverse Educational Needs	246
	Conclusion.....	248
	References	249
13	THE ROLE OF SCHOOL-BASED ASSESSMENT	253
	Introduction.....	253
	Strategies to Reduce the Burden of Examinations.....	254
	Dimensions of School-Based Assessment	258
	Arguments in Favor of School-Based Assessment.....	261
	Arguments against School-Based Assessment.....	263
	Addressing Problems in School-Based Assessment	266

Conclusion.....	271
Notes	272
References	273

14 CONSIDERATIONS FOR EXAMINATION REFORM 279

Introduction.....	279
Examination Reform Considerations.....	280
Options for Improving Examination Items and Questions	282
Options for Improving Administration and Reporting	285
Characteristics of a Good Examination System.....	291
Conclusion.....	294
Notes	296
References	297

GLOSSARY OF ACRONYMS AND TECHNICAL TERMS..... 301

INDEX 311

BOXES

1.1	Differences between Public Examinations and System Assessments.....	5
1.2	France: The Baccalauréat	10
1.3	United Kingdom: The General Certificate of Education A Level.....	11
3.1	Excerpts from Rules for Written Examinations, 1599.....	50
4.1	Levels in Bloom's Taxonomy of Educational Objectives for Knowledge-Based Goals (Revised)	79
5.1	Republic of Korea: Public Initiatives to Reduce Student Stress	108
5.2	Arab Republic of Egypt: Shadow Education.....	112
6.1	General Recommendations, Chief Examiner, English	125
6.2	General Recommendations, Chief Examiner, Health Education.....	126
6.3	Kenya Certificate of Primary Education: Sample Essay and Marker's Comments	128
6.4	Value-Added Modeling	137
7.1	Opening Paragraphs of Student Essays	154
7.2	Recommendations to Improve the Validity of Examinations	157
8.1	Human Factors That Affect Marking.....	166
8.2	Recommendations to Improve Marker Reliability	168
9.1	Factors That Might Affect Student Examination Grades over Time	177
9.2	Selected Countries: Academic Grades in Examinations.....	181
10.1	Examination Town Where Results Really Matter	198
10.2	China: Location and Likelihood of Performing Well on the <i>Gaokao</i>	199
10.3	Practices Associated with Examinations That May Create Inequities	208
11.1	Reasons for Examination Malpractice.....	225
12.1	Testing Accommodations	240
12.2	Document Requirements for Candidates with Special Needs	242
12.3	Special Provisions for Special Needs Candidates	244

12.4	Accommodations for Special Needs Candidates.....	245
13.1	External and School-Based Components for Science.....	258
13.2	External and School-Based Components for Foreign Language.....	259
13.3	Advantages of School-Based Assessment Compared to External Examinations.....	262
13.4	Steps for Carrying Out School-Based Assessment as Part of a Public Examination System.....	272
14.1	Contextual Factors to Be Considered When Planning Examination Reform	282
14.2	Options for Scoring Examinations.....	287

FIGURES

4.1	Hierarchy of Levels of Cognitive Processes: Bloom's Taxonomy and Revised Edition	80
5.1	Examinations' Impact on Subjects Studied	105
6.1	England and Wales: School League Table Headlines	134
13.1	Queensland: External Moderation Process	257

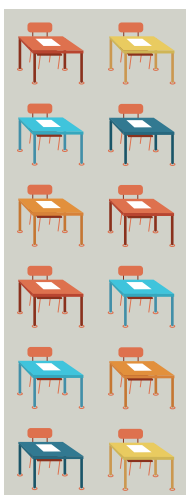
PHOTOS

B1.2.1	France: Diploma Awarded to Baccalauréat Graduates at the End of Their Second-Level Education	10
2.1	Sri Lanka: General Certificate of Education Diploma Featuring Individual Subject Grades	20
2.2	India: Secondary School Examination Diploma Certifying Subjects in Which Pass Standards Have Been Achieved	21
3.1	Ancient Chinese Public Examination	45
3.2	Examination Hall with 7,500 Cells, Canton, China, 1873.....	46
3.3	Examination Cells at School in Nanjing, China.....	47
4.1	Candidate Responding to Four-Option Multiple-Choice Question on a Machine-Scannable Answer Sheet	83
4.2	Ireland: Sorting Packets of Examination Papers Prior to Issuing Them to Markers.....	91
5.1	China: Candidates' Mothers Burning Incense and Offering Prayers for Examination Success.....	109
5.2	China: Teenagers Emerging from High School Devoted to Cramming Just before 11 p.m. after a Routine Study Day	114
6.1	School Principal Explaining How to Use Public Examination Results and Other Assessment Data to Help Improve Teaching and Learning at a Forum Organized by the Hong Kong Examinations and Assessment Authority	128
7.1	Ireland: Examination Paper Setter Checking Content Levels and Cognitive Skills Being Considered for an Economics Examination	145
10.1	Oman: Female Examination Candidates.....	194
B10.1.1	China: Part of Maotanchang Cram School's 165-Acre Campus.....	198
10.2	South Sudan: Primary School Students Taking Their Final Examination at Napata School in Ajoung Thok Refugee Camp	202

11.1	India: Parents and Other Helpers Climbing a School Wall in the East Indian State of Bihar to Help Children Cheat on the High-Stakes Grade 10 Public Examination	220
11.2	A Police Officer Displaying a Pair of Confiscated Glasses with a Hidden Camera and a Tiny Receiver Attached to a Coin, Both Pieces of Equipment Used for Cheating on Examinations	221

TABLES

4.1	Course Structure for Biology	76
4.2	Question Paper Design for Biology Examination, Class XI, 2017–18	78
4.3	Alternative Strategies for Printing	87
4.4	Measures to Improve Security of Examination Papers	88
4.5	Level Description of Objectives, Knowledge, and Understanding	90
12.1	Issues Arising from Administering Examinations to Students with Disabilities or Diverse Needs	247



FOREWORD

Curriculum-based public examinations are the dominant form of external high-stakes educational assessment in many countries of the world. These examinations are used both to certify that students have reached a prescribed level of learning and to select students for the next level of the education system or for employment. We know that the contents of these examinations exert a great influence on what teachers actually teach and what students learn.

Public Examinations Examined describes public examination practices in an extensive range of Organisation for Economic Co-operation and Development (OECD) and non-OECD countries. It draws on a very considerable body of international sources, including World Bank projects and studies on various aspects of current examination practices, and on efforts to enhance education through examination reform. The evidence is marshaled to highlight important issues related to the design of examinations, including the testing of higher-order cognitive skills, administration procedures, and use of results. There is no list here of “best practices” drawn from a limited number of countries. Instead, the reader is presented with a range of examination approaches and issues to be considered by policy makers, including evidence of what works and does not work in specific situations. Reform efforts are always context specific, but there is also something to learn from the positive and negative experiences of others. The book cautions policy

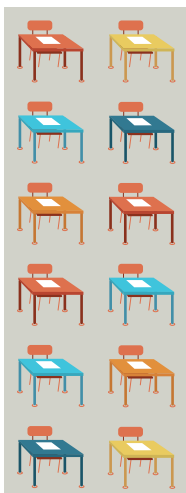
makers to appreciate that changes can give rise to both positive and unanticipated negative outcomes for teaching and student learning.

While the authors recognize that there are no easy answers to many current examination issues, they nevertheless give concrete suggestions that should apply to all systems. These include ensuring that the content of the examination is consistent with curriculum requirements; aligning curriculum, assessment, and teacher training; seeking consistency in the marking of examination papers; and using results to help improve teaching and learning. The book highlights the need for addressing examination-related inequities in the education system and tackling the ever-present and universal threat posed by examination malpractice or cheating. It also discusses the risks and benefits associated with using information technology and school-based assessment, as well as issues related to accommodating candidates with disabilities or diverse educational needs.

In the long run, improved public examinations can contribute to the personal development of individual students as well as to enhancing overall levels of knowledge and personal attributes in a system, which can in turn contribute to raising national levels of human capital. Ill-designed or poorly implemented public examinations, on the other side, can produce more harm than good. And many countries also have chosen not to use public examinations at all or to transform them radically, making them part of a continuous integrated evaluation process throughout the years of secondary education. What really matters is that countries have an assessment system that fosters a rich teaching-learning process and steers schools' and students' efforts toward the knowledge, skills, and competencies that are useful for life. This mirrors the World Bank's commitment to ensuring that all of the world's students have the opportunity to learn, giving them the chance to compete in tomorrow's economy, improve their communities, contribute to their countries' development, and move us closer to a world that is finally free of poverty and with opportunities for all. *Public Examinations Examined* can provide policy makers and other educational stakeholders with a rich source of information, reflection points, and options as they ponder how to create examination systems, and more broadly assessment systems, that promote high-quality student learning.

Jaime Saavedra

Global Director, Education
World Bank Group



PREFACE

Public examinations loom large in most educational systems across the world. They play a significant role in many countries, providing the basis for certifying a student as having completed a formal course of studies in an educational system or for employment, a particularly important consideration in countries with scarce employment opportunities and nonexistent unemployment supports. They also play a selection role, as they are widely used to choose students for the next-highest level of an educational system, sometimes at the end of primary, more often during second-level, and most commonly at the end of second-level schooling, when they are generally used for selection to tertiary-level education. They can play an important equity role in improving access as well; they help limit the effects of patronage and open up access to tertiary education to students from relatively disadvantaged backgrounds through various forms of support, including elimination of fees and provision of scholarships. They tend to enjoy a high measure of public confidence, though they have also been criticized for their overemphasis on the acquisition of low-level cognitive skills and encouragement of rote learning.

Beyond these basic roles, public examinations have considerable effects on teachers, students, and parents as well. They help focus teachers' curriculum and pedagogical priorities and provide incentives

for students to study. Because of their consequences, teachers tend to pay particular attention to the “tradition of past examinations” and gear their teaching to the content and skills featured in previous examinations rather than to broader national curriculum objectives. Because of the high stakes involved—for students, parents, teachers, and schools—these examinations can also have important unintended consequences, including their contribution to levels of student stress, early dropping out of school, and retention in grade. Furthermore, in an effort to boost the likelihood of examination success, many parents make strenuous efforts to get their children into “good schools” and engage the services of providers of shadow education in the form of tutors and grind schools. The high stakes have in addition prompted various forms of cheating and malpractice.

Despite these numerous commonalities across countries and education systems, public examination systems vary considerably throughout the world in terms of examination format, frequency of administration, extent of curriculum coverage, assessment capacity, sources of funding, susceptibility to malpractice, and uses of results. Thus modifications of current examination practices in one country or system will not necessarily be appropriate in another.

Not surprisingly given the high stakes and profound potential effects, calls to reform public examination systems are frequently heard and remarkably persistent. While any number of issues associated with examinations can be targeted by those calling for reform, serious consideration should be given to eliminating malpractice; enhancing validity, reliability, and equity; using technology to improve administrative practices; abolishing examinations that are deemed unnecessary for selection; and using examination results to enhance student learning and teaching. Efforts to bring about reforms of public examinations are likely to prove challenging, because of these examinations’ well-established traditions, imbedded administrative practices, low levels of assessment expertise, and existing teacher and student practices and expectations. Such reforms will require a series of compromises and trade-offs that will have an impact on the extent, nature, and timing of changes, and the outcome of any reform effort is likely to depend on the strength and the consistency of the support it garners from political leaders. Reform priorities will vary depending

on country contexts, including financial and human resources and the willingness of an education system to embrace change.

Against that backdrop, *Public Examinations Examined* focuses on issues related to the design, implementation, and use of examinations throughout the world. It covers practical aspects such as examination construction, candidate registration, marking, and provision of feedback on examination performance. Characteristics and functions of different types of examination systems are outlined. Technical issues reviewed include examination validity, marker reliability, and standard setting. Other topics covered include accommodating candidates with disabilities or diverse educational needs and the extent to which the chances of examination success are loaded in favor or against students of a particular population subgroup.

For historical context, the book summarizes the use of examinations dating from the early Chinese civil service selection examinations up to current times. Various forms of examination malpractice or cheating, including threats posed by modern technology, are also described. The volume concludes with suggestions that examination authorities might consider in evaluating reform choices.

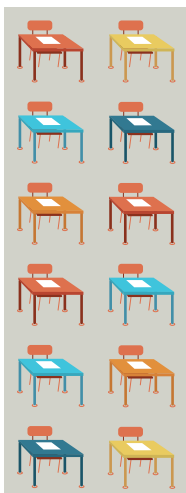
We have tried in this book to synthesize some of the key issues related to current public examination systems in both Organisation for Economic Co-operation and Development (OECD) member countries and non-OECD countries. We do not offer a cookbook approach that attempts to present the ideal solution to the complex issues surrounding examinations. Rather, we have portrayed a number of options and considerations based on evidence gathered from an extensive range of countries that proponents of examination reform might reflect on before embarking on major changes.

The book represents the final publication of Thomas Kellaghan, who died in March 2017. Thomas developed the concept for this publication and did most of the early drafting during our first two years working on the project. Initially the book was designed to build on concepts and practices covered in separate publications Thomas had authored or coauthored with various other people, as well as publications we had produced on educational assessment, including public examinations. We also drew on our joint and separate experiences with examination systems in individual countries in Europe,

Sub-Saharan Africa, the Middle East and North Africa, and South and East Asia. During the two years since Thomas's death, I've expanded the manuscript to include recent developments in the field of public examinations, advantages and disadvantages of examinations, equity issues that affect some candidates, malpractice, and specific topics pertinent to low- and middle-income countries.

Thanks are due to Marguerite Clarke, team leader of this project, and to peer reviewers Hanna Katriina Alasuutari, Sajitha Bashir, Mark Bray, Michael Crawford, Stephen P. Heyneman, Thanh Thi Mai, and Ezequiel Molina. Valuable timely inputs were received from Harsha Aturupane, Betty Jane Greaney, Hugh McManus, David Millar, Juan Manuel Moreno, Daria Ochorova, Gerry Shiel, and Algirdas Zabulionis. Additional support was provided by May Chan, Tony Dudek, Justin Frable, Kevin Gippert, James Lamb, Cathal McDonagh, Eoin McVey, Tara O'Connor, Michael Shulman, and Vineet Tripathi. The Educational Research Centre, Dublin, supported Thomas Kellaghan's work on the project; the assistance of Peter Archer, Anne Comey, Eileen Corbett, Jude Cosgrove, John Doyle, Mary Rohan, and Hilary Walshe is gratefully acknowledged. Jaime Saavedra, Omar Arias, and Luis Benveniste provided assistance and guidance at key points in the preparation of the manuscript. Publication of the book is sponsored by the Russia Education Aid for Development (READ) Trust Fund, administered at the World Bank by Julia Liberman and Victoria Levin, with the assistance of Restituto Jr. Mijares Cardenas, Dariga Chukmaitova, and Lorelei Lacdao.

Patricia Katayama, of the World Bank's Development Economics unit, was the acquisitions editor for the volume; its production was managed by Michael Harrup of the Editorial Production team. Rick Ludwick edited the book, and Sherrie Brown served as its proof-reader. They greatly assisted, each in their own way, in the process of transforming the manuscript into the final published product.



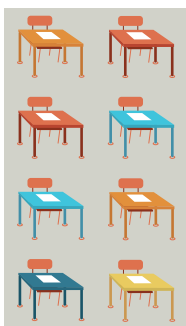
ABOUT THE AUTHORS

Thomas Kellaghan was director of the Educational Research Centre at St. Patrick's College, Dublin, from 1966 to 2009. He also worked at the University of Ibadan in Nigeria and at the Queen's University in Belfast, from which he obtained his PhD. He was author, coauthor, or editor of more than two hundred professional publications, founding editor of the *Irish Journal of Education*, and an elected fellow of the International Academy of Education. His areas of research interest covered educational assessment, including national and international assessments and public examinations; educational disadvantage; home-school relationships; teacher education; and educational policy. He served as president of the International Association for Educational Assessment from 1997 to 2001. Over the course of his career, he worked on educational assessment issues and educational reform programs in a number of African countries, in South Asia, and in Latin America. In 2004 he was awarded an honorary doctorate by the National University of Ireland for his contribution to education. Thomas died in March 2017.

Vincent Greaney has been a primary school teacher, textbook author, research fellow at the Educational Research Centre in Dublin, and visiting Fulbright professor at Western Michigan University. He obtained a PhD in educational evaluation, measurement, and

research from Boston College. A former board member of the International Reading Association, in 1996 he was elected to its Reading Hall of Fame. After joining the World Bank staff in 1990, he worked on educational projects in Sub-Saharan Africa, South and East Asia, the Middle East and North Africa, and Eastern and Central Europe and implemented a training program on national assessments of educational achievement. His extensive list of publications covers topics related to educational assessment, teacher education, reading, and promotion of social cohesion through textbook reform. He is a member of the Board of Directors of the Canadian Organization for Development through Education and of the Expert Council of the Russia Education Aid for Development program.

CHAPTER 1



INTRODUCTION

OVERVIEW

Public examinations play a crucial role in most second-level schools in Africa (Bashir et al. 2018; Kellaghan and Greaney 2004), South and East Asia and the Pacific (Dundar et al. 2014; Hill 2013), Europe (Bethell 2010; Madaus and Kellaghan 1991; West, Edge, and Stokes 1999), and the Middle East and North Africa region (World Bank 2018) because of their functions in certifying student achievement levels and in selection for the next level of the education system and for employment. In some instances, they are regarded as appropriate measures of teacher and school accountability. Their importance is underscored by their tendency to have a strong impact on the nature of teaching and learning in schools, stronger in fact than other forms of external assessment. Activities and issues related to the administration of public examinations and the release of results frequently receive extensive media coverage. As can be seen in the following chapters, high-stakes public examinations can exert strong pressure on students, their parents, teachers, and schools but also have serious consequences for users of results and for governments or examination agencies that implement them.

Examinations' effects on students can be both positive and negative. Positive effects stem from the ability of examinations to direct and

focus students in their studies and to motivate them to work even when no major decision is based on their performance, such as mid-way through second-level education. When important consequences are attached to performance, such as selection for tertiary-level education, the examination is likely to have a stronger motivational influence. On a less positive note, public examinations often ignore a wide range of knowledge and skills that cannot be measured by paper-and-pencil tests, and also restrict student opportunities to study some curriculum subjects that are not offered as examination subjects. Examinations can also contribute to stigmatizing some students as failures, increase rates of grade repetition (Madaus and Greaney 1985), represent an obstacle to promotion to second-level education (Bashir et al. 2018), lead to dropping out of school early (Kreitzer, Haney, and Madaus 1989), and result in unhealthy pressure, including self-harm (Busby 2018; Kale 2018).

Parents can affect examination performance by providing a positive home learning environment, and by recognizing their child's learning achievements. Highly motivated parents, to advance their child's educational and employment prospects, often have to make considerable personal sacrifices to increase the likelihood of examination success. These efforts can take the form of engaging in costly after-school supports ("shadow education") in the form of private tuition or "grind" schools (see chapter 5), and in some instances resorting to more extreme measures (see chapter 11). In some low- and middle-income countries, low-income parents may also have to consider the potential loss of income as a result of their child attending examination-grade classes.

Teachers can use examinations to clarify which aspects of a curriculum subject might be considered most important and identify the cognitive skills to be stressed in their classrooms. In some instances, because of a lack of curriculum documentation, the examination curriculum as reflected in past examination papers helps define the "effective" as opposed to the official curriculum, and also clarifies what is to be taught and how it should be taught. On a less positive note, teachers may devote substantial classroom time to developing examination answering techniques, encouraging rote memorization, focusing on topics within subjects that are likely to be examined

rather than on a broader curriculum, ignoring important oral and practical skills as well as higher-level cognitive skills, and directing attention to students most likely to do well.

For governments, high-stakes examinations promote a form of educational cohesion by helping ensure that teachers of individual subjects cover a common curriculum and emphasize similar skill components and knowledge in the classroom. Governments might argue that formal public examinations support a fair, impartial system of student certification and a legitimate method of allocating scarce educational benefits, such as places in higher levels of an education system. On the other hand, examination budgets in many countries represent a considerable drain on state financial resources arising from the cost of full-time and part-time staff; printing, distributing, administering, and correcting papers; issuing results; and processing appeals. At another level, they can pose a severe political risk for governments when examinations are not carried out on time; when there is evidence of malpractice, including leakage of examination questions; when students publicly protest over perceived changes in examination content and the difficulty level of papers; and when teachers' representatives use examinations as tools in labor disputes.

Public (external, national, or "exit") curriculum-based examinations play an important role in many education systems. They have a long history in their country of origin, China (see chapter 3). Examination use developed rapidly in the second half of the nineteenth century in many countries, including France, Germany, and Great Britain. One does not have to look very far to find the reasons for their growth in popularity. They were perceived to allocate scarce educational benefits in an objective and unbiased way by selecting the most talented and by removing opportunities for nepotism and favoritism. They were efficient in providing a relatively inexpensive form of assessment at a time when student numbers in education systems were growing rapidly, while their emphasis on competition was in tune with the ideas and beliefs of the time. Today, only a few industrial countries (such as Belgium, Spain, Sweden, and the United States) do not have curriculum-based external examinations at the end of secondary school. The examination models that were developed in Europe subsequently became an important component of the education system of many

countries, occupying a central role in the assessment of individual students and leading to a situation in which the great majority of third-level-bound students had to pass an advanced subject-specific examination (Britton, Hawkins, and Gandal 1996).

PUBLIC EXAMINATIONS, CLASSROOM ASSESSMENT, AND SYSTEM ASSESSMENT

It is difficult to define the term “public examination,” because of the wide variation in the format and uses of these examinations. For example, some examinations, though generally regarded as external, include an element of school-based assessment and so are not entirely external to the schools from which examinees come. The reintroduction of mass testing in the lower grades of schooling in several countries has increased the difficulty of categorization. In some cases (Australia, England, and the United States), tests are viewed primarily as (census-based) national assessments, though their use may suggest a greater affinity with public examinations. This book is concerned for the most part with the traditional type of public examination administered at the end of compulsory schooling (usually the end of lower-secondary education) and on the completion of secondary schooling (usually two or three years later).

Public examinations are defined primarily in terms of their purposes: to certify and select students on the basis of an assessment of their achievements in curriculum areas. Their characteristics and functions are described in greater detail in chapter 2. The examinations can be distinguished from two other forms of educational assessment: classroom assessment and system assessment (Clarke 2012; Kellaghan and Greaney 2004). Classroom assessment, which can take a multiplicity of forms, is an integral component of the teaching-learning process, much of it subjective, informal, immediate, ongoing, and intuitive. It interacts with learning as it occurs, monitoring student behavior, scholastic performance, and responsiveness to instruction. It occurs during learning, rather than when learning is presumed to be complete, and is designed to facilitate the acquisition of knowledge and skills.

System assessment, on the other hand, is a formal procedure organized by an agency outside the school, which may be carried out at either the national or international level, and is designed to describe the level of achievement, not of individual students (as is the case with public examinations) but of a whole education system or a clearly defined part of one (for example, grade 4 students) (Greaney and Kellaghan 2008; Kellaghan and Greaney 2004). The National Assessment of Educational Progress in the United States is an example of a system assessment.

Public examinations and system assessments may be similar in many respects. They may cover the same curriculum areas, use comparable methods of assessment, and require similar administrative systems. However, they also differ in a number of ways (see box 1.1).

Despite differences between public examinations, classroom assessment, and system assessments, they can share certain functions. Data from classroom assessments may be used to determine grades in a public examination (see chapter 13). Data from a national assessment (if census- rather than sample-based) may be used to make

BOX 1.1

Differences between Public Examinations and System Assessments

- *Aggregation of data.* Public examinations provide data on individuals; system assessments are typically concerned with the performance of the system.
- *Information provided.* Decisions are made about individuals on the basis of their performance on a public examination, while system assessment findings are used to make judgments about aspects of the system of education.
- *Grade level.* System assessments are usually carried out at lower grade levels of the educational system, while public examinations are carried out at higher levels.
- *Impact.* System assessments tend to have much less impact on students and teachers than public examinations.

judgments about individual students similar to those made in public examinations or to perform functions traditionally associated with classroom-based assessment (for example, relating to grade promotion; Kellaghan and Greaney 2001).

VARIETY IN EXAMINATION SYSTEMS

Variety in the nature, extent, and functions of examination systems can be related to a number of aspects of education systems. First, one might expect a public examination system to be related to the degree to which an education system is centralized (that is, one that is administered by a central authority and in which schools implement a national curriculum). However, the relationship is not straightforward. On one hand, it is true that examinations and curricula are both features of centralization and that decentralized systems (such as in the United States) do not have a national examination system. On the other hand, a centralized examination system may be considered necessary to monitor standards in a system in which control is predominantly local (for example, Norway).

Second, while all education systems are characterized by some degree of internal differentiation and specialization, they vary in the age at which differentiation occurs and in the number of specializations provided. We would expect public examinations to be associated with differentiation at an early stage in the education system and in the number of specialization routes provided. It should be noted that differentiation has often been associated with an elitist ideology: those who appear destined for higher social positions and who are considered to be most likely to contribute to economic productivity are separated at an early stage from those who appear bound for lower positions (Hopper 1977).

Third, public examinations are more likely to be a feature of an education system when participation rates at a particular level of the system are low. For example, when places for secondary education are limited, some selection is required to control entry. As participation rates increase, the need for an examination decreases. At one time, an examination at the end of primary schooling was common, both to

provide certification for students terminating their education at this point and to select students for secondary education. Because all or most students transferred to the next phase of education, examinations at this stage were abolished in many countries. However, there are exceptions. Performance on a public examination at the end of primary schooling is still used for secondary school placement in a number of countries (for example, in Kenya, Singapore, and most countries in the Caribbean).¹ Poland also requires an examination at this stage to provide information on student achievement, but not for admission to lower-secondary school (Polish Eurydice Unit 2014). Interaction between examinations and school participation rates may be reflected in the fact that the gradual elimination of public examinations, and the associated practice of grade retention, has been associated with increasing participation.

This book, it may be noted, is being published at a time when the use of examinations and of assessments in general is being proposed as a means of changing teacher behavior and classroom instruction, which in turn is expected to improve educational quality, evidenced by student learning. In this view, examinations are not just a means of obtaining information about students that can be used to make decisions about certification or selection; they are also considered to be a lever of reform (Kellaghan and Greaney 2001; Madaus, Russell, and Higgins 2009). This view is evident in the United States where, in response to the poor performance of students in international assessments of student achievement, national examinations of the type common in Europe have been proposed. This seems to be founded on the assumption that curriculum-based external “exit” examination systems based on explicit content standards are an essential ingredient of a world-class education system (Bishop 2005; Madaus and Kellaghan 1992).

Belief in the potential of examinations to contribute to improved learning is not confined to the industrial world. Two major international events—the World Conference on Education for All, held in Jomtien, Thailand, in 1990, and the World Education Forum, which met in Dakar, Senegal, in 2000—prompted a shift in focus in educational policies in low- and middle-income countries (World Bank 1995). Rather than focusing on school participation rates, ministries of education and

international aid agencies began to prioritize the acquisition of knowledge and skills. This shift in approach was supported by increased emphasis on the role of examinations and other forms of large-scale assessments. According to the 2015 World Education Forum, which took place in Incheon, Republic of Korea, the importance of “measuring learning outcomes is an essential component of the management and strengthening of education systems” (UNESCO 2015, 18).

In Central and Eastern Europe, the collapse of socialist economies opened the door to western ideas and practices, providing “solutions” to problems in education systems, in particular ones related to selection for third-level education. In Poland, for example, it was envisaged that reforms to the education system between 2002 and 2005 would be driven to a great extent by external examinations administered at the end of sixth grade in primary school, the end of third grade in lower-secondary school, and the end of the final grade in upper-secondary school (Jakubowski et al. 2010).

MAJOR PUBLIC EXAMINATION TRADITIONS

Three separate assessment traditions or paradigms have dominated educational assessment (Baird and Opposs 2018):

- The *psychometric paradigm*, which has its origin in the field of intelligence testing and is concerned with measuring a single factor or trait, tends to have items that are not related to the curriculum; places a high emphasis on item discrimination, ranking students based on achievement scores, and test reliability; and produces scores that are normally distributed.
- The *outcomes-based paradigm* is concerned with competency or mastery of a specific set of skills, does not assume a normal distribution of scores, tends to be associated with pass/fail decisions, and is often linked to vocational- or occupational-type activities.
- The *curriculum-based assessment*, as the name implies, is concerned primarily with assessing the extent to which a student has acquired the knowledge and skills listed in the official curriculum and has the potential to advance to the next level of the education system.

In contrast with the psychometric paradigm, it places a strong emphasis on validity, is not concerned with measuring a single trait, tends to give results in the form of grades, and does not assume that scores or results are normally distributed. (For example, it does not assume that very small percentages of students get either very low or very high scores.) While aspects of each approach can apply across paradigms, this book is primarily focused on curriculum-based assessment carried out through public examinations, the dominant high-stakes assessment approach in most education systems.

The large number of curriculum-based examination systems covered in this volume can be traced to a variety of examinations established in Europe in the eighteenth and nineteenth centuries. The first of these was the Abitur, which was established in Prussia in 1788. It was influenced by examination procedures in universities, evident in the extent to which it emphasized school-based assessment and the role it accorded to oral examining.² This situation may be contrasted with that in France, where the examination that was developed (the Baccalauréat) reflected a political tradition of commitment to centralized decision making and uniformity in the treatment of all citizens.

Although introduced at a later date, the Baccalauréat (1808) and a range of examinations in the United Kingdom that were developed in the middle of the nineteenth century, and have since gone through a series of transformations, have had a much wider influence than the Abitur (partly, no doubt, a function of colonial history). Examinations similar to those in the United Kingdom are currently administered in many former colonies (such as Bangladesh, Pakistan, Singapore, and Sri Lanka), while the Baccalauréat is administered in former French colonies (such as Algeria, Mali, Mauritania, and Morocco). Unlike the Abitur, these examinations are carried out by an agency outside the school and place a strong emphasis on essay writing. Boxes 1.2 and 1.3 contain vignettes of examinations in France and in the United Kingdom.

A number of general points may be made about the examinations that emerged in the eighteenth and nineteenth centuries. First, while they were initially developed by individual institutions or professions, over time governments wrested control of the examinations, resulting in

BOX 1.2**France: The Baccalauréat**

The Ministry of Education has authority over curricula and examinations throughout the country. The Baccalauréat, introduced by Napoleon in 1808, has three tracks (academic general education, technology, professional or vocational) with a variety of streams (*series*) within each. Arrangements for construction and administration of examinations are left to 30 regional academies grouped into clusters. Candidates normally take examinations in five or more subjects. The overall mean scores determine if a student passes or fails the Baccalauréat. A pass mark certifies satisfactory completion of the three-year course of further secondary education and is generally required for admission to university and some professional and training programs. Students' school-based grades are not taken into consideration in calculating a "bac" score.^a The word Baccalauréat appears to be derived from the Latin *bacca* (a berry) and *laureus* (of laurel) and is linked with the tradition of awarding a wreath of berries and laurel to recognize achievement (see photo B1.2.1).

PHOTO B1.2.1**France: Diploma Awarded to Baccalauréat Graduates at the End of Their Second-Level Education**

Sources: Cros 2009; Hawkins, Gandal, and Britton 1996.

a. Lycee (high school) and university subject specialists and inspectors from the ministry develop examination questions following ministry guidelines and grading criteria. Examinations normally consist of essay-type questions, but a number of multiple-choice items are sometimes included in a minority of subject areas. Examinations are graded by teams of teachers overseen by instructors and university subject specialists. Teachers may not grade their own students' work.

BOX 1.3**United Kingdom: The General Certificate of Education
A Level**

The General Certificate of Education (GCE) examinations are administered at the end of secondary school (Year 13). Examinations at Advanced (A) Level are based on two years' study. Examinations at Advanced Subsidiary (AS) Level are based on one year's study. The examinations are administered by four separate examination boards or awarding bodies in England and Wales and one in Northern Ireland. (Scotland has its own independent examination system.) Board syllabi and examinations follow general directions of the government Office of Qualifications and Examinations Regulation (Ofqual). Candidates usually take A-level examinations in three or four subjects. Schools and colleges may take examinations with different boards and at different times.^a

Sources: Examination boards; Ofqual.

a. The General Certificate of Secondary Education examination system replaced the earlier GCE Ordinary (O) Level and Certificate of Secondary Education systems. Subject panels composed of secondary-school teachers, subject specialists, and university staff work on a part-time basis to construct examinations and scoring guides. Early drafts are reviewed to ensure that the examinations meet syllabus objectives and that questions are clear and fair. The process is overseen by a Chief Examiner.

diminished university and professional involvement. Second, although sharing a common name, the examinations tolerated considerable variation. The Abitur varied from state to state in Germanic countries, while the Baccalauréat differed in types and streams designed to accommodate the varied characteristics of an increasing number of students sitting the examinations. Third, despite differences in examinations, equivalences have been established for the purpose of deciding on university entrance. Performance on the General Certificate of Education in the United Kingdom is generally considered equivalent to performance on the Abitur, the Baccalauréat, the Advanced Placement examinations (in the United States) and the International Baccalaureate. The Lisbon Recognition Convention (Council of Europe 1997) served to reinforce a trend toward convergence in European school leaving qualifications.

THE CONTENTS OF THIS BOOK

This book describes characteristics of public examinations in chapter 2, after which a brief description of their history is presented

in chapter 3. A description of how public examinations are constructed follows (chapter 4). Chapter 5 describes some of the advantages and limitations of examinations, including the impact of private tutoring. The following chapters discuss a number of topics associated with examinations: validity, reliability, administration, and standards and issues relating to equity and malpractice. The book also describes how examination authorities that are dedicated to the maintenance of universal standards take account of the circumstances of candidates with special needs. The book is concerned primarily with external examinations used to certify and select students, traditionally found in Europe, Africa, and Asia, but that have found their way in recent years into the United States in the form of state-mandated tests and into the post-Soviet socialist states of Central and Eastern Europe in examinations generally known as state *maturas*. The discussion draws, in particular, on the experiences of educational systems of low-income countries in developing and administering public examinations.

Readers should bear in mind that examinations and assessment systems in general are a sensitive political issue in many countries, involving discussion about many of the topics addressed in this book (such as equity, motivational effects, the use of examinations for accountability purposes, the reliability of examination results, the role of school-based assessment, and the washback effect of examinations on teaching and learning) (see, for example, Briseid and Caillods 2004). Examinations are also the objects of reform responding to forces for change, which can be political, economic, social, technological, or pedagogical (Bray 1998). Changes in policy and practice usually require negotiation and compromise (Carless 2011).

The topics discussed in this book are drawn from the experiences of examination bodies in a wide range of countries and accepted educational measurement practices, and are presented to highlight key policy issues that might be considered in efforts to improve public examination systems.

NOTES

1. Some countries (for example, Botswana and Brunei Darussalam) require out-of-school children to pass primary school leaving examinations to enter

the next level of the education system. Bashir et al. (2018) cite Kanjee (2012) and Sayed and Kanjee (2013) as sources for their list of African countries with selection examinations, by education level: (a) primary school leaving examinations: Botswana, Burkina Faso, Chad, Comoros, the Democratic Republic of Congo, Eritrea, Eswatini (formerly known as Swaziland), Ethiopia, Kenya, Mali, Mauritius, Mozambique, Niger, Nigeria, Rwanda, Senegal, Sierra Leone, Tanzania, Uganda, Zambia, and Zimbabwe; (b) lower-secondary examinations: Botswana, Burkina Faso, Chad, Comoros, Eswatini, Ethiopia, The Gambia, Ghana, Nigeria, Rwanda, Senegal, Sierra Leone, Tanzania, Uganda, Zambia, and Zimbabwe; (c) upper-secondary university entrance examinations: Angola, Benin, Botswana, Burkina Faso, Chad, Comoros, the Democratic Republic of Congo, Eritrea, Ethiopia, The Gambia, Ghana, Kenya, Lesotho, Mali, Malawi, Mauritius, Mozambique, Niger, Nigeria, Rwanda, Senegal, Sierra Leone, South Africa, Tanzania, Uganda, Zambia, and Zimbabwe; and (d) postsecondary university examinations (Mauritius, South Africa, Tanzania, and Zimbabwe).

2. The Abitur includes written and oral assessment and covers four or five examination subjects, which must include three subject areas (languages, literature and the arts; social sciences; mathematics, natural sciences and technology). It is awarded on the basis of student grades over the final years of course work and performance on written and oral examinations. The examination component accounts for a maximum of about 22 percent of the total mark, depending on the state.

REFERENCES

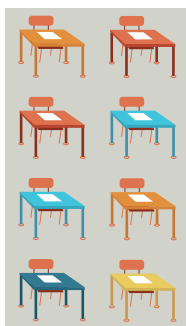
- Baird, J.-A., and D. Opposs. 2018. "The Standard Setting Project: Assessment Paradigms." In *Examination Standards: How Measures and Meanings Differ around the World*, edited by J.-A. Baird, T. Isaacs, D. Opposs, and L. Gray, 2–25. London: University College London Institute of Education Press.
- Bashir, S., M. Lockheed, E. Ninan, and J.-P. Tan. 2018. *Facing Forward: Schooling for Learning in Africa*. Washington, DC: World Bank.
- Bethell, G. 2010. "Student Assessment: Policy and Practice in Eastern Europe." In *International Encyclopedia of Education*, 3d ed., edited by P. Peterson, E. Barker, and B. McGaw, 472–78. Oxford: Elsevier.
- Bishop, J. H. 2005. "High School Exit Examinations: When Do Learning Effects Generalize?" In *Uses and Misuses of Data for Educational Accountability and Improvement: 104th Yearbook of the National Society for the Study of Education*, part 2, edited by J. L. Herman and E. H. Haertel, 260–88. Chicago: National Society for the Study of Education.

- Bray, M. 1998. "Lessons for Conceptual Understanding." In *Examination Systems in Small States: Comparative Perspectives in Policies, Models and Operations*, edited by M. Bray and L. Steward, 208–35. London: Commonwealth Secretariat.
- Briseid, O., and F. Caillods. 2004. *Trends in Secondary Education in Industrialized Countries: Are They Relevant for African Countries?* Paris: International Institute for Educational Planning, United Nations Educational, Scientific, and Cultural Organization.
- Britton, E., S. Hawkins, and M. Gandal. 1996. "Comparing Examination Systems." In *Examining the Examinations*, edited by E. D. Britton and S. A. Raizen, 201–17. Dordrecht, Netherlands: Kluwer Academic.
- Busby, E. 2018. "Pupils Self-Harm and Express Suicidal Feelings Due to Exam Stress and School Pressure, Warn Teachers." Independent, April 10. <https://www.independent.co.uk/news/education/education-news/school-pupil-mental-health-exams-school-pressure-national-education-union-neu-a8297366.html>.
- Carless, D. 2011. *From Testing to Productive Student Learning: Implementing Formative Assessment in Confucian-Heritage Settings*. London: Taylor and Francis.
- Clarke, M. 2012. "What Matters Most for Student Assessment Systems: A Framework Paper." SABER—Systems Approach for Better Education Results Student Assessment 1, World Bank, Washington, DC.
- Council of Europe. 1997. "Convention on the Recognition of Qualifications Concerning Higher Education in the European Region Lisbon." Council of Europe, Strasbourg, France.
- Cros, F. 2009. "Country Case Study: France." In *Secondary School External Examination Systems: Reliability, Robustness and Resilience*, edited by B. Vlaardingerbroek and N. Taylor, 15–28. Amherst, NY: Cambria.
- Dundar, H., T. Béteille, M. Riboud, and A. Deolalikar. 2014. *Student Learning in South Asia: Challenges, Opportunities, and Policy Priorities*. Washington, DC: World Bank.
- Greaney, V., and T. Kellaghan. 2008. *Assessing National Achievement Levels in Education*. Washington, DC: World Bank.
- Hawkins, S., M. Gandal, and E. Britton. 1996. "Examination Systems in Seven Countries." In *Examining the Examinations*, edited by E. D. Britton and S. A. Raizen, 219–49. Dordrecht, Netherlands: Kluwer Academic.
- Hill, P. 2013. *Examination Systems*. Paris: United Nations Educational, Scientific and Cultural Organization.
- Hopper, E. I. 1977. "A Typology for the Classification of Educational Systems." In *Power and Ideology in Education*, edited by J. Karabel and A. H. Halsey, 153–66. New York: Oxford University Press.

- Jakubowski, M., H. A. Patrinos, E. E. Porta, and J. Wisniewski. 2010. "The Impact of the 1999 Education Reform in Poland." Policy Research Working Paper 5263, World Bank, Washington, DC.
- Kale, S. 2018. "In India, High-Pressure Exams Are Creating a Student Suicide Crisis." *Wired*. <https://www.wired.co.uk/article/india-kota-student-suicide-exams-institutes-of-technology>.
- Kanjee, A. 2012. "Assessment and Education Quality in Sub-Saharan Africa: Prospects and Pitfalls." Paper presented at the Public Seminar on Assessment, Oxford University, Oxford, UK, October 29. <http://oxes.org.uk/wordpress/wp-content/uploads/2013/03/Assessment-and-education-quality-in-sub-Saharan-Africa-presentation-with-timings.pdf>.
- Kellaghan, T., and V. Greaney. 2001. *Using Assessment to Improve the Quality of Education*. Paris: International Institute for Educational Planning, United Nations Educational, Scientific and Cultural Organization.
- . 2004. *Assessing Student Learning in Africa*. Washington, DC: World Bank.
- Kreitzer, A. E., G. F. Madaus, and W. M. Haney. 1989. "Competency Testing and Dropouts." In *Dropouts from School: Issues, Dilemmas, and Solutions*, edited by L. Weis, E. Farrar, and H. G. Petrie, 129–52. Albany: State University of New York Press.
- Madaus, G. F., and V. Greaney. 1985. "The Irish Experience in Competency Testing: Implications for American Education." *American Journal of Education* 93: 268–94.
- Madaus, G. F., and T. Kellaghan. 1991. "Student Examination Systems in the European Community: Lessons for the United States." In *Science Assessment in the Service of Reform*, edited by G. Kulm and S. M. Malcolm, 189–232. Washington, DC: American Association for the Advancement of Science.
- . 1992. "Curriculum Evaluation and Assessment." In *Handbook of Research on Curriculum*, edited by P. W. Jackson, 119–54. New York: Macmillan.
- Madaus, G. F., M. Russell, and J. Higgins. 2009. *The Paradoxes of High Stakes Testing*. Charlotte, NC: Information Age.
- Polish Eurydice Unit. 2014. *The System of Education in Poland*. Warsaw: Foundation for the Development of the Education System.
- Sayed, Y., and A. Kanjee. 2013. "Assessment in Sub-Saharan Africa: Challenges and Prospects." *Assessment in Education: Principles, Policy and Practice* 20: 373–84.
- UNESCO (United Nations Educational, Scientific and Cultural Organization). 2015. *World Education Forum: Final Report*. Paris: UNESCO.

- West, A., A. Edge, and E. Stokes. 1999. *Secondary Education across Europe: Curricula and School Examination Systems*. London: Centre for Educational Research, London School of Economics and Political Science.
- World Bank. 1995. *Priorities and Strategies in Education*. Washington, DC: World Bank.
- . 2018. *Expectations and Aspirations: A New Framework for Education in the Middle East and North Africa*. Washington, DC: World Bank.

CHAPTER 2



CHARACTERISTICS AND FUNCTIONS OF PUBLIC EXAMINATIONS

INTRODUCTION

Chapter 1 describes the main characteristics and several functions of public examinations. In considering the contents of the chapter, two reservations should be kept in mind. First, all the functions do not apply in the case of all examination systems. Second, the fact that examination systems share certain characteristics and functions should not lead us to ignore the enormous variation that exists between systems: the range of examination procedures that can be labelled “public” or “state-wide examinations” is enormous (Klein 2010).

CHARACTERISTICS OF PUBLIC EXAMINATIONS

Examinations to which the terms “public” or “external” are normally applied possess a number of characteristics (Eckstein and Noah 1993; Keeves 1994; Kellaghan and Madaus 2003). First, they are set or controlled by an agency external to the schools from which students come. Second, the administering authority is usually a national or state government or agency or, if it does not actually administer the examinations, it will have an oversight function. Third, the examinations are based on prescribed syllabi in curriculum (or subject)

areas, such as national languages and literature, other languages, mathematics, natural sciences, social studies, arts, information and communication technologies, technology, and religion. (In this way, they differ from some university entrance exams and other admissions tests such as the SAT and the ACT, used in the United States). Syllabi are sometimes prescribed by the agency that administers the examination. In line with a tradition in which the study of classical texts was the main feature of curricula, the emphasis in examination syllabi has tended to be on content rather than on skills, though some reforms (such as in Hong Kong SAR, China, and Norway) emphasize students' acquisition of generic skills rather than subject matter. Many systems aspire to this approach. Fourth, examinations involve the application of a common test administered under controlled conditions, in which students do not have access to books or other material¹ and are separated from the classroom situation. The examinations are usually administered on fixed days to many students at the same time. There is a heavy emphasis on written tasks involving essays, though other forms of assessment, such as multiple-choice items or oral and practical tasks, may be included.

Fifth, public examinations have generally been voluntary in the sense that it was up to individual students to decide whether to take an examination. However, that situation is changing and many examinations at primary and lower secondary educational institutions are now considered compulsory (for example, in Denmark, Germany, and Portugal). Sixth, as a result of performance on an examination, the examinee is awarded a grade or mark in each subject examined. Finally, the tradition of external examinations has been to make public the content of examinations and their results. This practice existed before examinations were presented to candidates in written form. For example, after the oral examination of students at Dublin University during the first half of the nineteenth century, individuals who had been present wrote up and circulated the questions that had been asked. These questions were then used by students who were preparing for similar examinations in the future, contributing to the tradition of having examinations determine what is taught and learned (Foden 1989). Some examination systems (such as examination boards in India, Malta, and the United Kingdom) publish marking or scoring keys. In a number of countries (for example, Germany, Ireland, Slovenia, New Zealand), in addition to publishing

examination papers, answer scripts are made available to examinees following the examination to allow them to evaluate the way their responses were scored (Gabršček 1999; Steer 1999).

FUNCTIONS OF EXAMINATIONS

Most examinations serve a number of functions: certification; selection; motivating students; controlling the activities of schools; providing information that can be used in managing the educational system; holding schools, teachers, or students accountable for student achievement levels; and legitimizing membership in global society (Kellaghan 1992; Kellaghan and Greaney 1992, 2004; Kellaghan and Madaus 2003; Somerset 1996).


Certification

The most obvious function of a public examination is to assess the competence of students in relation to some agreed-upon standards, on the basis of which certification of students' achievements at the end of a period of study (lower secondary, upper secondary) is provided (Evans 2009).² Diplomas or certificates awarded describe the performance of students on each subject in the examination in letter grades (for example, A, B, C, D, E), numbers (for example, 1, 2, 3, 4, 5), percentages, or in a proficiency statement (for example, pass/fail; see chapter 9). Photo 2.1 provides an example of a candidate's grades on the Sri Lankan General Certificate of Education. Grades may be determined by simply summing marks allocated to sections of questions and across questions and components (or papers) if the examination has more than one component (or paper), a procedure that may have negative implications for validity and transparency (see chapter 7). Or different weights may be assigned to different components of an examination. In the German Abitur, for example, the student's overall grade is calculated using a complex formula to weight and aggregate the raw scores obtained in examined subjects and grades from school records. In the French Baccalauréat, mathematics and the sciences receive relatively high weightings in the science stream, while philosophy, French language and literature, history, and geography receive high weightings in the literary stream.

PHOTO 2.1**Sri Lanka: General Certificate of Education Diploma Featuring Individual Subject Grades**

FORM CT 7

Serial No. : 209472


DEMOCRATIC SOCIALIST REPUBLIC OF SRI LANKA
 (CEYLON)
Department of Examinations
National Evaluation & Testing Service
 P.O. BOX 1503, COLOMBO, SRI LANKA
OFFICIAL STATEMENT

My Ref. 11 / 3 / [] 07-March-2007

General Certificate of Education (Ordinary Level) Examination, Sri Lanka

Name in Full : []

Year of Examination : 1980

Index Number : []

Subject	Grade
FIRST LANGUAGE - SINHALA	C CREDIT
SECOND LANGUAGE - ENGLISH	C CREDIT
MATHEMATICS	S ORDINARY PASS
SCIENCE	S ORDINARY PASS
HEALTH	C CREDIT
BUDDHISM	S ORDINARY PASS
SOCIAL STUDIES	C CREDIT
PRE. VOC. STUDIES	S ORDINARY PASS

This statement is issued for use outside Sri Lanka. It may be accepted as evidence of candidate's successful performance only if it carries the embossed seal of the Department of Examinations, Sri Lanka.

RANGES OF MARKS

G.C.E. (O/L)			G.C.E. (A/L)		
Before 2001	2001 and after		Before 2001	2001 and after	
Range	Grade	Symbol	Range	Grade	Symbol
100 - 75	Distinction	D	100 - 75	Distinction	A
74 - 50	Credit	C	74 - 65	Very Good Pass	B
49 - 35	Ordinary Pass	S	64 - 50	Credit	C
34 - 00	Failure	F	49 - 35	Ordinary Pass	S
			34 - 00	Failure	F

+ Absent, × Not Decided

Prepared By Mancel

Checked By *[Signature]*

K. M. KARUNATHILAKA
 Assistant Commissioner of Examinations
 Department of Examinations,
 SRI LANKA

[Signature]
 For Commissioner General of Examinations

The certification function often tends to be overlooked because of the emphasis placed on other functions, but formal certification of academic achievement can help students gain access to employment, training, or other levels of the educational system. The diploma issued by the Indian Central Board of Secondary Education,

for example, shows that the candidate had passed the secondary school examination in eight subject areas (see photo 2.2). It should be recognized, however, that lower-level certificates lose their currency in the labor market as the number of students possessing them increases and employers use higher-level certificates to select personnel even though the levels of achievement represented by the higher-level certificates may not be required.

PHOTO 2.2

India: Secondary School Examination Diploma Certifying Subjects in Which Pass Standards Have Been Achieved

क्रम संख्या/
S. No. SSE/ 3 18 / 3 / 14

DUPLICATE

केन्द्रीय माध्यमिक शिक्षा बोर्ड
Central Board of Secondary Education

सेकण्डरी स्कूल परीक्षा, १९९८
ALL INDIA SECONDARY SCHOOL EXAMINATION 1998

यह प्रमाणित किया जाता है कि
This is to certify that

अनुक्रमिक
Roll No.

आत्मज/आत्मजा श्री
Son/Daughter of Shri

जिनकी जन्म तिथि
Born on

है, ने बोर्ड द्वारा मार्च १९९८ में आयोजित सेकण्डरी स्कूल परीक्षा
passed the Secondary School Examination of the Board held in March, 1998

विषय से
from

निम्न विषयों में उत्तीर्ण की :-
In the following subjects :

ENGLISH COURSE A	GUJARATI	MATHEMATICS
SCIENCE WITH PRAC	SOCIAL SCIENCE	WORK EXPERIENCE
PHY & HEALTH EDUCA	ART EDUCATION	

दिल्ली | AJMER
Delhi
दिनांक
Dated 27-1-2014

उप सचिव/Deputy Secretary
क्षेत्रीय कार्यालय, अजमेर/Regional Office, Ajmer
केन्द्रीय माध्यमिक शिक्षा बोर्ड / C. B. S. E.
टोडरमल मार्ग, अजमेर-३६०००१
Todaramal Marg, Ajmer-360001
Controller of Examinations

* का बिन्दु जिस विषय के आगे लगा है, वह विषय परीक्षाओं में पूरक परीक्षा में उत्तीर्ण किया है।
* against a subject indicates that the candidate passed in the subject at the Compartmental Examination

Selection

The certificate or diploma awarded on the basis of performance on an examination, in addition to certifying that an individual has reached a certain level of achievement, may also confer rights, such as the right to be considered for (if not actually admitted to) some sector of the social, professional, or educational world. In most educational systems, students' performance on an examination and the availability of places are taken into account in selection for further education. The tradition in some systems, however, is that possession of a school-leaving examination certificate (the German Abitur or the French Baccalauréat) grants automatic access to a university. Such a system is only possible when selection has taken place earlier in students' careers. A disadvantage of the system is that it creates overcrowding in universities, requiring selection during students' first two years of study when the majority are likely to be rejected because of unsatisfactory progress.

Exceptions to students' automatic right of entry to university occur in the case of programs with relatively small numbers of places for which there is stiff competition (such as medicine or veterinary science). In these cases, admissions are limited to the number of places available in a university (*numerus clausus*). In Germany, for instance, additional criteria are in place for selection to these programs: grades in specific subjects; waiting time; and results of an interview; some schools also offer a nonmandatory national aptitude test, which can increase an applicant's chances of success (Chenot 2009). Even when these additional criteria are taken into consideration, performance on the Abitur remains an important consideration in the selection process.

In many countries, the selection of students for further education from primary to lower secondary, from lower secondary to upper secondary, and from upper secondary to higher education has become the main function of public examinations. In the East Asia and Pacific region, most educational systems still continue to use exams for entrance decisions at the secondary cycle (World Bank 2018). In this context, public examinations are generally perceived to allocate scarce educational benefits in an objective and unbiased way. If there has to be selection, it has been argued in both European and emerging

market economies that examinations are a more equitable way of doing this than other procedures that have been tried (Heyneman 1985).

Examinations are used less frequently for selection as provision is made to accommodate entire age cohorts at increasingly higher grades in educational systems. In the Middle East, for example, secondary school entrance examinations are disappearing or are being used primarily for counselling and orientation rather than selection (World Bank 2008). In East Asia, the Republic of Korea; Taiwan, China; and Hong Kong SAR, China, for instance, do not use second-level admission examinations (World Bank 2018). However, examinations continue to be used in some educational systems to allocate students to different types of schools and courses (for example, Germany, Malaysia, the Netherlands, and the Russian Federation [Bolotov et al. 2013; Clark 2014; EACEA 2009]). When examinations are used in this way, major selection based on socioeconomic (SES) background will occur at the point of differentiation. Students from high SES backgrounds will tend to enter an academic type of program that will lead to third-level education, while students from low SES backgrounds will tend to enter technical, vocational, or short-cycle programs. The effects of SES are stronger at earlier than at later transition stages (Kellaghan 2015).

Examinations can contribute to the sorting of students even when secondary school curricula are not differentiated and there are no formal procedures for allocating students to schools. This is the situation in countries in the anglophone Caribbean, which continue to administer examinations at the end of primary school even though promotion to secondary school is (in theory) automatic. The examination is maintained partly in response to the pressure from parents who see it as a means of getting their children admitted to the “best” schools in an educational system in which differences between schools in the quality of education on offer can be very large.

Some countries, (for example, Azerbaijan, Greece, Hungary, Saudi Arabia) have operated separate selection and certification examinations, although having separate examinations for certification and selection is obviously more expensive than having one examination serve both purposes. There is the further problem that the selection

examination will be likely to attract greater attention and effort from students and schools than the certification examination because of the high stakes attached to it.

The practice of individual universities holding their own examinations, or other procedures, to select students, once a regular feature in the socialist republics of eastern and central Europe and elsewhere, has become less common. Some universities, however, are often reluctant to give up their control of the selection of students for a variety of reasons. In Russia these include a loss of revenue generated by preparatory courses and entrance examinations (Bolotov et al. 2013). Similarly, in Brazil, some universities have concerns over technical aspects of national tertiary selection tests (Guimarães de Castro 2012). In Vietnam, although a common examination for school leaving and university entrance has recently been established, two universities in Hanoi have indicated that they intended to hold additional entrance examinations (Pham 2015).

In England and Wales, universities differ in their control of entry. “Selective” universities specify entrance requirements, usually in terms of “points,” state what subjects applicants must have studied, and may interview prospective students. “Recruiting” universities advertise prerequisites for courses, which are lower than those required by selective universities (Baird 2009).

Monitoring Educational Standards over Time

Politicians and the media tend to regard examination performance as a form of evidence about the stability of, or the change over time in, standards of student achievement in the educational system. Thus, judgments may be made about “standards” on the basis of the change in the proportion of examination candidates who are deemed to have passed or been awarded high grades. This is problematic because examination questions and examination cohorts differ from year to year, with the result that it usually is not possible to say if, for example, an increase in the proportion of candidates awarded higher grades represented improved levels of achievement, more lenient awarding of grades, or the results of changes in the difficulty of questions. In an effort to address this question in the United Kingdom a comparison of

changes in student performance on public examinations with performance on an independent test of ability suggested that standards had declined over time (see chapter 9; Tymms, Coe, and Merrell 2005).

Motivation

Examinations may serve a motivational function. An important distinction in the case of high-stakes examination is whether motivation is intrinsic or extrinsic. Intrinsic motivation has been defined as the energizing force of self-initiated behavior and as an outgrowth of interests or internal needs (such as needs for competence or for self-determination). Thus, motivation is said to be intrinsic when the origin of action resides within the individual. Extrinsically motivated behaviors, on the other hand, are performed for, and regulated by, rewards outside the person, such as a gold star, promotion, or certification. By contrast with intrinsically motivated behaviors, behaviors that are extrinsically motivated occur, not out of interest in a task, but because they are seen as instrumental in obtaining some tangible outcome or reward (Kellaghan, Madaus, and Raczek 1996.)

The use of extrinsic motivation in the case of public examinations will be most obvious when important consequences are attached to performance. When this is the case, teachers and students will focus their efforts on examination requirements (see chapter 5). Even when certification or selection decisions are not based on performance (for example, midway through secondary schooling), the examination may be considered useful to direct students in their study or to motivate them to work. In theory, this is achieved by providing clear goals to strive for, a sense of purpose, and tangible incentives and rewards.

An obvious motivational use of examinations would appear to be to award additional marks (“bonus marks”) to candidates choosing a particular subject or a subject at a higher or more advanced level. There are a number of examples of the use of this strategy to encourage students to study mathematics at an advanced level. In Ireland (for the Leaving Certificate Examination) and in Israel (for the Bagrut), the strategy was accompanied by a considerable increase over a four-year period in student uptake (70 percent in Ireland; 100 percent in Israel), following which uptake leveled off. The proportion of low scorers

("failures") also increased, suggesting that a small number of students were sitting the advanced level examination when a less advanced level would have been appropriate. It may be hoped that students who opt for the more advanced study of mathematics to improve their examination results would in time reach a point at which they study because of an intrinsic interest in the subject.

Controlling Activities in Schools

By specifying clear goals and standards for teachers and students, examinations may be considered a way of controlling what goes on in schools by signalling what aspects of the official curriculum are to be covered and what might receive relatively little attention or be ignored. Time to be devoted to curriculum subjects such as art, music, environmental studies, and physical education may be sacrificed in the interests of devoting more hours to examination subjects. In addition, aspects of subject areas to be tested tend to receive more classroom time than untested subject areas. For instance, in the area of language many examinations place emphasis on reading comprehension and essay writing, but ignore oral fluency as the latter tends not to be examined. Examination questions tend to focus on low cognitive-level achievement (see chapter 5). At a broader level, public examinations help ensure that similar content is taught across the education system, and thus they contribute to the promotion of national homogeneity in educational standards and practice. These can be important considerations in educational systems with relatively unspecific curriculum goals, where private schools represent a considerable proportion of schools, or where there is strong local control of schools. They were also considered to be important in the Czech Republic when a higher-level secondary examination was introduced to replace internal school-based certification. It was claimed that the new centralized system would allow for a comparison of individual student performance and, in effect, meant that the state would be able to guarantee the content of the final examination and the educational level achieved by secondary-level students (Skutil and Maněnová 2009). Pilot programs in 18 provinces designed to lessen

the emphasis on rote learning in the *gaokao* examination, which drives K-12 education in China, have been described as deepening the Communist Party's control over the education system (China Policy 2019). It should be noted that national control of what goes on in schools will be diminished to the extent that students sit for regional or international examinations.

Management

Examinations can be important “management tools,” providing information that can be used to monitor the quality of the education system (UNESCO 2013). Evidence of unsatisfactory performance at a national or a regional level can provide a basis for questioning existing policies and for considering the distribution of resources (Briseid and Caillods 2004). This, in turn, can lead to action to improve student learning and to reduce educational disparities.

Accountability

An important and controversial use of examination performance is holding schools or teachers accountable for their students' achievements. This use becomes more obvious when results for individual schools are made public. A basic problem with this approach is that it fails to give due recognition to the many factors that contribute to students' scholastic progress (see chapter 6).

Income Generation

In many countries, examination-related activities generate considerable income for teachers, through offering private tuition (see chapter 5), supervising the administration of examinations, and marking examination scripts. This additional income can represent a sizable portion of some teachers' income. Fear of loss of examination-related income can generate labor relationship problems. Publishing printed or online versions of old examination papers, answer keys, or sample essays has become a thriving commercial activity in many countries (OECD 2012).

Legitimizing Membership in the Global Society

A final use of examinations is legitimate membership in the global society, facilitating the movement of students internationally to study abroad; in 2010 many of the 4.1 million students who had emigrated to pursue tertiary education abroad would have used their examination results to help gain admission (OECD 2012). We may assume that the need to facilitate students in former colonies to study in universities in Europe and elsewhere was a major consideration in examination bodies maintaining some kind of relationship with examination authorities in metropolitan countries following independence (Stanley-Marcano and Alexander 1998).³

DIFFERENCES AMONG EXAMINATION SYSTEMS

Although a range of general characteristics of examinations can be identified, there is enormous variation between countries in their public examination systems (Briseid and Caillods 2004; EACEA 2009; Eckstein and Noah 1993; Hawkins, Gandal, and Britton 1996; Hill 2013; Kanjee 2012; Madaus and Kellaghan 1991; UNESCO 2013).

The Stage at Which Examinations Are Held

There are differences at the stage at which examinations are held. Examinations are most common at the end of compulsory basic schooling and at the end of secondary schooling. However, some countries also have examinations at the end of primary schooling. Countries in which these examinations continue to be administered include some in Africa (Angola, Botswana, Burundi, Cameroon, Ethiopia, Kenya, Lesotho, Liberia, Malawi, Mauritius, Sierra Leone, Uganda, Tanzania); in Asia (Brunei Darussalam, Indonesia, Malaysia, Singapore), in Europe (Poland), and in the Middle East and North Africa region (Algeria, the Arab

Republic of Egypt, Saudi Arabia, and the Syrian Arab Republic). On the other hand, examinations at the basic education level have been abolished in Jordan, Kuwait, and Tunisia, as well as in Djibouti, Morocco, and the Republic of Yemen. However, it is of interest that, in the case of the latter three countries, implementing an automatic promotion policy did not increase the transitional rate to secondary education, indicating that expanding enrollment at this level requires more than abolishing exit or selection examinations (World Bank 2008).

Variation in Number and Nature of Bodies Involved in Administration

There is variation among countries in the number and nature of bodies involved in the administration of examinations. The most common arrangement is to have a single examination authority for a country, most usually a government ministry of education or a government agency. In a federal system, each state may have its own authority (such as in Germany and Switzerland) although a national ministry (or a committee of regional ministries) may have ultimate authority for curricula and examinations. Examinations may be administered by public or private examination boards (for example, in Bangladesh, India, Pakistan, the United Kingdom [Dundar et al. 2014]);⁴ or administration may be localized (for example, in clusters of regional *academies* in France or prefectures or municipalities in Japan). The fact that more than one agency is responsible for the administration of examinations means that several versions may exist of examinations with the same name (for example, Abitur, Baccalauréat, General Certificate of Education, Higher Secondary School Certificate). This has obvious implications for the comparability of standards. Regional examination boards offer examinations in the Caribbean, the South Pacific, and West Africa (see chapter 3). The International Baccalaureate is offered in more than 140 countries, and a number of UK boards offer examinations in many countries in Africa, Asia, and the Middle East.

Fees

Policies regarding examination fees vary greatly from country to country and even between states in some countries. In some instances governments pay for all examination-related activities (for example, Kazakhstan, Kuwait, Lebanon, Serbia); in others, funding is provided by the governments (Armenia, Uganda) or solely from student fees (Ethiopia; Nepal; Punjab, Pakistan).⁵ Fees are adjusted regularly in light of changing economic and budgetary circumstances; in 2017 one government indicated that it was not in a financial position to pay candidate fees for the following year (Teh 2017).⁶ In some instances, fees are based on number of subjects examined and in others students pay the same overall fee. Additional charges that have been imposed include fees for examinations in practical subjects, in extra subjects, or for external students, overseas students, review of marks, and repeating examinations. By supporting examination fees for less-well-off students, governments promote social equity and thus help reduce the number of students who drop out because of inability to pay (Paton 2012). In Mauritius, for instance, the state social security system refunds examination fees to families of needy students. Private examination boards almost always depend on student fees or grants to cover costs (Punch 2017).

Costs

Costs associated with public examinations tend to vary considerably by educational system, depending on the type of examination offered. Essay-type examinations are generally inexpensive to construct but expensive to score, while the reverse is true in the case of multiple-choice-type exams. Adaptive testing, which requires relatively large item pools, high-level technical information technology capabilities, and psychometric competence, can be particularly expensive. Administration costs include wages, distribution, collection of examination papers, travel and subsistence, printing, supervision, reporting results, and overhead. Other variable cost factors depend on the extent to which assessments are school based, use technology, combat malpractice, provide training for examination board staff and

invigilators, use quality control measures (such as double scoring), promote the use of results, conduct examination-related research and evaluation, and accept appeals (Connolly and Cullen 2017).

In many countries, families of examination candidates pay for private supplementary tutoring, which can greatly exceed examination registration fees (see chapter 5), in addition to examination fees (where they apply) and other school-related costs. In countries that use tests in addition to secondary school exit examinations to choose students for admission to such highly selective faculties as medicine and law, families will have to carry the additional costs for fees and frequently for supplementary private tuition.² On the other hand, public examinations can help reduce overall examination costs. The introduction of the Unified State Examination in Russia helped limit or eliminate payments parents had been making to university personnel in the form of direct bribes or for private tuition to assist their children in gaining admission under the earlier admission system (Bolotov et al. 2013).

Balance between External and Internal Components

School exit examinations differ in the extent to which they are totally external to the students' schools (in which case the student will be unknown to the examiner) and the extent to which they incorporate an element of assessment by the candidates' own teachers (see chapter 13). Examinations are externally set or verified in a majority of African and South Asian countries, in some East Asia and Pacific countries, and in European countries. However, school exit examinations are totally or almost totally external in only a few European countries, such as Finland, France, and Ireland (West, Edge, and Stokes 1999). At the other extreme, an examination may be set by or marked by teachers within a school, with some form of external moderation, as they are in Belgium, Greece, Iceland, Spain, and Sweden.

Between these extremes, a student's final grade will be a combination of marks on an external examination and marks for schoolwork over a period of time, such as grades awarded in the final year of schooling. Internal components may include oral language, fieldwork in geography, laboratory work in science, and artistic performances.

External components are usually based on essays, short-answer, and multiple-choice questions. Systems vary in the weight they apportion to the internal and external elements of an examination. For example, in the Netherlands a student's final examination mark in a subject gives somewhat more weight to the national external examination than to the internal school-based components (Scheerens et al. 2012). In Germany (Bavaria), most of the final grade in the Abitur has traditionally been based on marks earned during the last two years of secondary school.

The emphasis that education systems place on external examinations or school-based assessment seems to be heading in two directions. One group of countries (for example, Bangladesh and New Zealand) has increased the amount of external formal assessment in support of the implementation of standards-based reform. Another group (such as Bhutan, India, and Norway) has moved in the opposite direction, reducing the amount of external formal assessment in an effort to maintain a continuous flow of the whole age cohort through lower- and upper-secondary education (Briseid and Caillods 2004; Dundar et al. 2014). Traditions of public examinations, whether largely external or largely internal, have proved extremely resistant to change.

Number of Subjects a Candidate Takes

Examination systems differ in the number of subjects in which they offer examinations. There is also variation in the number of subjects a student takes in an external examination. At the end of secondary school, this ranges from a minimum of two in Russia, to three to four (A Level) in the United Kingdom, generally four in China and Germany, and five or more in Armenia, Egypt, France, India, Ireland, Israel, Kazakhstan, and South Africa.

Student Response

Examinations vary in the types of responses they require of students (see chapter 4). Essay-type questions in which candidates are required to write extended answers to questions or prompts have

predominated in many systems. Short-answer constructed responses in which candidates respond to a question by producing a word, number, or phrase are also used. In other cases, rather than requiring the candidate to construct an answer, he or she selects the correct answer from competing alternatives presented in so-called objective (multiple-choice) questions. Many examinations also include “performance” tasks, in which aural (such as listening to a recorded conversation in a foreign language) and practical competencies are assessed. A number of countries have oral components in their examinations. These not only assess language proficiency but in some instances represent the continuation of a tradition in which examinations were oral and school based, for example, in Austria, the Czech Republic, Denmark, Germany, Italy, and Switzerland.

Examination Levels

Examinations vary in whether they are offered at one level or are tiered, offering two or three levels of difficulty to provide candidates with a suitably challenging assessment experience. When offered at two levels, the examinations may be administered at different times, as in Ordinary and Advanced General Certificate of Education levels in England, Wales, and Northern Ireland. They also may be offered at the same time, requiring students to decide which level (such as Higher or Ordinary) better represents their preparatory study. This is the case in Croatia, the Czech Republic, Ireland, Poland, and Slovenia. In New Zealand, there are three levels of the National Certificate of Education. Students may take different subjects at different levels.

Concentration versus Spread of Examinations

Examinations may all be taken at the end of schooling or they may be spread over students’ final year or years of study. For example, New Zealand’s examination is spread over years 11 through 13 of secondary school. In Oman, as in some other countries in the Middle East, the final secondary school examination is administered in two sessions: one at the end of the first semester and one at the end of the school year.

Examination Options

There is variation among education systems in the range of options available under the umbrella term of a school leaving examination. In most systems, there is a single examination, which may contain a variety of subject examinations from which a candidate chooses. In other systems, a single title masks a range of very different options. Most noteworthy in this respect is the French Baccalauréat, which was once described as “an extraordinarily differentiated and complex examination system, with a host of *series*, *lignes*, and options (thirty-eight in 1988, compared to just four before 1950)” (Noah and Eckstein 1990, 88). Candidates take widely different assortments of subjects, different papers in nominally the same subject, with different weights given to the results depending on the particular selected option (sciences, economics and social sciences, and literature). As a consequence, it is not credible to speak of a single nationally comparable examination administered to all students. Rather, the Baccalauréat represents a strongly demarcated hierarchy of prestige, with mathematical options tending to be at the top and vocational options at the bottom. At the same time, it should be acknowledged that the diversification of the Baccalauréat represents a serious effort to accommodate the characteristics and needs of the increasing number of students who remain in the education system until the end of secondary school.

Choice

There is variation in the extent to which examinees are allowed a choice in the examination they are taking. In one system, all candidates are assessed in the same subjects. This is more common with younger than older students. In an alternative (and most common) system, free choice is permitted; students can select the subjects in which they wish to be examined. Some subjects, however, may be compulsory. When a subject is specified as compulsory, it is usually a national language and its literature. Several systems have at least one compulsory subject. In other systems, candidates can select subjects within a restricted framework of options.

There may also be choice in the examination components or questions in any particular subject to which a candidate responds. Most systems allow some choice, although they might also include some common tasks, but the range of options from which a choice may be made, for example, in a language test, is very large in some systems (for example, Ireland, Israel, and the United Kingdom) and nonexistent in others (for example, Germany). Choice may be provided in some subjects and not in others (for example, France). Allowing student choice has implications for the reliability and validity of an examination (see chapters 7 and 8).

The provision of choice in an examination can mean that there is great variation in the extent to which examinees sitting what is nominally the same examination have a similar experience. For example, it has been calculated that in the Irish Leaving Certificate Examination in one year there were 3,834 unique combinations of six “best” subjects (the number on which “points” for selection to third-level education were calculated). The number would be greater if differences in question selection within subjects were taken into account.⁸ Rather than being a single examination, examinations such as the leaving certificate should be considered to be a family of examinations (Mac Aogáin, Millar, and Kellaghan 2011).

CONCLUSION

All education systems are faced with two major tasks. The first is to certify the achievements of students, both to maintain standards and to provide evidence that individual students may need to pursue further education or use to seek employment. The second is the selection of students for further education in a situation that prevails everywhere: the number of applicants, at least for certain courses, exceeds the number of available places. A variety of procedures developed to meet these purposes attempt to satisfy standards of fairness and comparability while ensuring that the objectives represented in school curricula are adequately reflected. Increasing diversity in the forms of examinations to accommodate increases in the numbers and

heterogeneity of students indicates a primary concern with the certification function (in evidence in the French system). Efforts to maintain uniformity, on the other hand, with few options available to candidates (as in China, Japan, and Russia) reflect a primary concern with selection (see, for example, Eckstein and Noah 1993).

Public examinations may or may not be assigned a range of functions other than certification and selection (monitoring of educational quality, school and teacher accountability, student motivation). These additional functions are found in systems with a long tradition of public examinations as well as in ones that only recently introduced them. If anything, multiple uses seem more common in the latter than in the former.

The fact that systems of examination are used for a variety of purposes should not be taken to imply that a single system can serve all purposes equally well. For example, an examination system that efficiently selects the students most likely to benefit from further education might contribute to the identification of a technical elite. It could, however, have serious and damaging effects on the educational experiences of many students if it ignores the fact that—for many students, in fact the majority—examinations should have utility beyond that of identifying individuals for the next level of education. Such an observation points to the need to give careful consideration to the appropriateness of using public examination performance for a variety of functions. Indeed, it can be argued that purposes other than selection (for example, motivation of teachers and students, direction of teaching and learning) do not require a nationwide examination and certification system (Mathews 1985). Assigning a central role to examinations in an accountability system can be particularly problematic (see chapter 6).

Many years ago, Brerton pointed out that whatever purpose an examination is designed to serve, it should carry a reward for success and its regulations should be neither too lenient nor too stringent in relation to the circumstances in which students are working. Furthermore, the examination should be conducted fairly and honestly; and the standard of the examination should be related to students' abilities (that is, the examination should be neither too hard nor too easy for the students who take it; Brereton 1944, 2014).

These considerations still seem relevant, no matter how diverse examination systems may have become in the intervening years in their structures or functions.

NOTES

1. There are some examples of “open-book” examinations in which candidates have access to notes, reference materials, or textbooks. In a national e-assessment in Norway, students may use material they have stored on their hard drives or USBs, but may not access the internet. Rather than testing memory, these examinations are designed to assess a candidate’s ability to locate, organize, analyze, and apply relevant information.
2. As far back as 1835, the Edinburgh School of Arts in Scotland awarded certificates to students who had completed a course of study.
3. The first-ever overseas examinations of the University of London were administered in Mauritius in 1865 in response to a request from C. A. Rede, Rector of the Royal College in Mauritius, that the college be treated on the same basis as a United Kingdom provincial examination center for matriculation and BA examinations. Eleven students sat the examination in July 1865; six passed. Today, about 10,000 students in Mauritius sit the University of Cambridge International Examinations every year. Trinidad was the location of the first overseas center of the University of Cambridge Local Examinations Syndicate (UCLES). UCLES examinations were taken in Trinidad in 1863, five years after the establishment of the syndicate.
4. The number of examination boards varies considerably. There are four in England and Wales following a series of mergers. In India, the number has grown to more than 40. Over the past 30 years, federal governments have gradually increased their influence over educational matters, including assessment procedures, using financial leverage in countries in which education has traditionally been a state matter (for example, in Australia and the United States).
5. All Saber Country Reports: Kazakhstan 2011; Kuwait 2013; Lebanon 2012; Serbia 2012; Armenia 2012; Uganda 2009; Ethiopia 2012; Nepal 2012; Punjab Province, Pakistan 2012.
6. One year later some of the fees had been paid to the West African Examinations Council.
7. Medical schools in a number of countries (Australia, Ireland, and the United Kingdom), for instance, require the candidate to take the Graduate

Medical School Admissions Test (GMSAT), which assesses (a) Reasoning in Humanities and Social Sciences, (b) Written Communication, and (c) Reasoning in Biological and Physical Sciences.

8. Allowing candidates to choose 4 out of 10 topics or questions could mean that 210 choices were available; a choice of 3 out of 24 would give rise to 2,024 possible combinations.

REFERENCES

- Baird, J.-A. 2009. "Country Case Study: England." In *Secondary School External Examination Systems: Reliability, Robustness and Resilience*, edited by B. Vlaardingerbroek and N. Taylor, 29–54. Amherst, NY: Cambria.
- Bolotov, V., G. Kovaleva, M. Pinskaya, and I. Valdman. 2013. *Developing the Enabling Context for Student Assessment in Russia*. Washington, DC: World Bank.
- Brereton, J. L. (1944) 2014. *The Case for Examinations. An Account of Their Place in Education with Some Proposals for their Reform*. Cambridge, UK: Cambridge University Press.
- Briseid, O., and F. Caillods. 2004. *Trends in Secondary Education in Industrialized Countries: Are They Relevant for African Countries?* Paris: International Institute for Educational Planning, United Nations Educational, Scientific and Cultural Organization.
- Chenot, J. F. 2009. "Undergraduate Medical Education in Germany." *GMS German Medical Science* 7:Doc02. <http://www.egms.de/en/gms/2009-7/000061.shtml>.
- Clark, N. 2014. "Asia Pacific: Education in Malaysia." *World Education News and Reviews*, December 2. <http://wenr.wes.org/2014/12/education-in-malaysia>.
- Connolly, B., and D. Cullen. 2017. "European Court Rules That Exam Scripts and Comments Constitute Personal Data." William Fry, Dublin, December 21. <https://williamfry.com/newsandinsights/news-article/2017/12/21/european-court-rules-that-exam-scripts-and-comments-constitute-personal-data>.
- China Policy. 2019. "Gaokao: Reform and Reality Clash." *CP Signal*, January 25. <https://policyn.com>.
- Dundar, H., T. Bêteille, M. Riboud, and A. Deolalikar. 2014. *Student Learning in South Asia: Challenges, Opportunities, and Policy Priorities*. Washington, DC: World Bank.

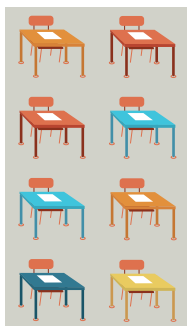
- EACEA (Education, Audiovisual and Culture Agency). 2009. *National Testing of Pupils in Europe: Objectives, Organisation and Use of Results*. Brussels: EACEA.
- Eckstein, M. A., and H. J. Noah. 1993. *Secondary School Examinations: International Perspectives on Policies and Practice*. New Haven, CT: Yale University Press.
- Evans, R. 2009. "The Beginnings of Examinations in the 19th Century." Chapter 2 in *The History of Technical and Commercial Examinations*. Technical Education Matters (website). <https://technicaleducationmatters.org/2009/08/20/chapter-2-the-beginnings-of-examinations-in-the-19th-century/>.
- Foden, F. 1989. *The Examiner: James Booth and the Origins of Common Examinations*. Leeds, UK: School of Continuing Education, University of Leeds.
- Gabršček, S. 1999. "The Right to See and Know: Slovenian Experience." Paper presented at the 25th Annual Conference of the International Association for Educational Assessment, Bled, Slovenia, May 23–28.
- Guimarães de Castro, M. H. 2012. *Developing the Enabling Context for Student Assessment in Brazil*. Washington, DC: World Bank.
- Hawkins, S., M. Gandal, and E. Britton. 1996. "Examination Systems in Seven Countries." In *Examining the Examinations*, edited by E. D. Britton and S. A. Raizen, 219–49. Dordrecht, Netherlands: Kluwer Academic.
- Heyneman, S. P. 1985. *Investing in Education: A Quarter Century of World Bank Experience*. Washington, DC: World Bank.
- Hill, P. 2013. *Examination Systems*. Paris: United Nations Educational, Scientific and Cultural Organization.
- Kanjee, A. 2012. "Assessment and Education Quality in Sub-Saharan Africa: Prospects and Pitfalls." Paper presented at the Public Seminar on Assessment, Oxford University, Oxford, UK, October 29. <http://oxes.org.uk/wordpress/wp-content/uploads/2013/03/Assessment-and-education-quality-in-sub-Saharan-Africa-presentation-with-timings.pdf>.
- Keeves, J. P. 1994. *National Examinations: Design, Procedures and Reporting*. Paris: International Institute for Educational Planning, United Nations Educational, Scientific, and Cultural Organization.
- Kellaghan, T. 1992. "Examination Systems in Africa: Between Internationalization and Indigenization." In *Examinations: Comparative and International Studies*, edited by M. A. Eckstein and H. J. Noah, 95–104. Oxford, UK: Pergamon.

- . 2015. "Family and Schooling." In *International Encyclopedia of the Social and Behavioral Sciences*, 2d ed., edited by J. D. Wright, vol. 3 751–57. Oxford, UK: Elsevier.
- Kellaghan, T., and V. Greaney. 1992. *Using Examinations to Improve Education: A Study in Fourteen African Countries*. Washington, DC: World Bank.
- . 2004. *Assessing Student Learning in Africa*. Washington, DC: World Bank
- Kellaghan, T., and G. F. Madaus. 2003. "External (Public) Examinations." In *International Handbook of Educational Evaluation*, edited by T. Kellaghan and D. L. Stufflebeam, 577–600. Dordrecht, Netherlands: Kluwer Academic.
- Kellaghan, T., G. F. Madaus, and A. Raczek. 1996. *The Use of External Examinations to Improve Student Motivation*. Washington, DC: American Educational Research Association.
- Klein, E. D. 2010. "The Differing Purposes and Effects of State-Wide Exit Examinations: A Comparison of Finland, Ireland and the Netherlands from a Government Perspective." Paper presented at 36th Annual Conference of the International Association for Educational Assessment, Bangkok, August 22–27. http://www.iaea.info/documents/paper_4d519c43.pdf.
- Mac Aogáin, E., D. Millar, and T. Kellaghan. 2011. "Transparency and Discriminatory Power of the Leaving Certificate Examination." *Irish Journal of Education* 39: 74–87.
- Madaus, G. F., and T. Kellaghan. 1991. "Student Examination Systems in the European Community: Lessons for the United States." In *Science Assessment in the Service of Reform*, edited by G. Kulm and S. M. Malcolm, 189–232. Washington, DC: American Association for the Advancement of Science.
- Mathews, J. C. 1985. *Examinations: A Commentary*. Boston: Allen and Unwin.
- Noah, H. J., and M. A. Eckstein. 1990. "Trade-Offs in Examination Policies: An International Perspective." In *Changing Educational Assessment: International Perspectives and Trends*, edited by P. Broadfoot, R. Murphy, and H. Torrance, 84–97. London: Routledge.
- OECD (Organisation for Economic Co-operation and Development). 2012. "Study Abroad." OECD Observer No. 293 (Q4), revised December 17. http://oecdobserver.org/news/fullstory.php/aid/3900/Study_abroad.html.
- Paton, G. 2012. "Schools Forced to Spend £328m on 'Over-inflated' Exams." *Telegraph*, May 9. <https://www.telegraph.co.uk/education/educationnews/9254585/Schools-forced-to-spend-328m-on-over-inflated-exams.html>.

- Pham, H. 2015. "Overhaul of University Entrance Exam Is Put to the Test." *University World News*, No. 371, June 12.
- Punch. 2017. "WASSCE, NECO Fees Too High—Former Education Commissioner." January 18. <https://punchng.com/wassce-neco-fees-high-former-education-commissioner>.
- SABER (Systems Approach for Better Education Results) Country Report. 2009. *Ethiopia: Student Assessment*. Washington, DC: World Bank.
- . 2011a. *Armenia: Student Assessment*. Washington, DC: World Bank.
- . 2011b. *Kuwait: Student Assessment*. Washington, DC: World Bank.
- . 2012a. *Kazakhstan: Student Assessment*. Washington, DC: World Bank.
- . 2012b. *Nepal: Student Assessment*. Washington, DC: World Bank.
- . 2012c. *Punjab Province, Pakistan: Student Assessment*. Washington, DC: World Bank.
- . 2012d. *Serbia: Student Assessment*. Washington, DC: World Bank.
- . 2012e. *Uganda: Student Assessment*. Washington, DC: World Bank.
- . 2013. *Lebanon: Student Assessment*. Washington, DC: World Bank.
- Scheerens, J., M. Ehren, P. Slegers, and R. de Leeuw. 2012. *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for The Netherlands*. Twente, Netherlands: University of Twente.
- Skutil, M., and M. Maněnová. 2009. "Changes in the School Leaving Examination as Part of the Transformation of Secondary School Education in the Czech Republic." *International Journal of Human Sciences* 6: 651–65. <http://www.insanbilimlari.com/en>.
- Somerset, A. 1996. "Examinations and Educational Quality." In *Assessment in Transition: Learning, Monitoring and Selection in International Perspective*, edited by A. Little and A. Wolf, 263–84. Oxford, UK: Pergamon.
- Stanley-Marcano, J., and M. C. Alexander. 1998. "Trinidad and Tobago." In *Examination Systems in Small States: Comparative Perspectives on Policies, Model and Operations*, edited by M. Bray and L. Steward, 113–19. London: Commonwealth Secretariat.
- Steer, M. 1999. "Transparent Assessment: Eight Years of Returning Answer Booklets to Candidates for National Senior Secondary School Qualifications in New Zealand." Paper presented at the 25th Annual International Association for Educational Assessment Conference, Bled, Slovenia, May 28–29.

- Teh, L. S. 2017. "Liberia: Govt Not Paying WASSCE Fees." *New Dawn*, November 20. <https://allafrica.com/stories/201711201239.html>.
- Tymms, P., R. Coe, and C. Merrell. 2005. "Standards in English Schools: Changes since 1997 and the Impact of Government Policies and Initiatives—A Report for The Sunday Times." Durham, UK: Centre for Evaluation and Monitoring, University of Durham.
- UNESCO (United Nations Educational, Scientific, and Cultural Organization). 2013. *The Use of Student Assessment for Policy and Learning Improvement*. Bangkok: Asia and Pacific Regional Bureau for Education, UNESCO.
- West, A., A. Edge, and E. Stokes. 1999. *Secondary Education across Europe: Curricula and School Examination Systems*. London: Centre for Educational Research, London School of Economics and Political Science.
- World Bank. 2008. *The Road Not Travelled: Education Reform in the Middle East and North Africa*. Washington, DC: World Bank.
- . 2018. *Growing Smarter: Learning and Equitable Development in East Asia and Pacific*. Washington, DC: World Bank.

CHAPTER 3



A BRIEF HISTORY OF WRITTEN EXAMINATIONS

INTRODUCTION

In eastern Asia, written examinations are probably as old as, or older than, most social institutions. It is in that part of the world that this chapter begins, with an outline of the history of the imperial examination system that was established in China and existed in a variety of forms for thousands of years until its abolition in 1906.¹ The use of examinations in European schools and universities described later in the chapter is much more recent. Later sections of the chapter provide an indication of the extent to which examinations have spread throughout the world.

An understanding of key aspects of the history of assessment can help us appreciate that many functions and practices associated with current public examinations have their origins in the distant past. It can also serve as a reminder that policy makers can learn from successful and unsuccessful examination reform initiatives that were carried out in different eras and in different parts of the world, and the wisdom of the aphorism (attributed to George Santayana) that “those who cannot remember the past are condemned to repeat it.”

EXAMINATIONS IN IMPERIAL CHINA

It is generally accepted that written competitive group examinations originated in China (Du Bois 1970; Eckstein and Noah 1993; Mathews 1985; Morris 1961; Teng 1943; Webber 1989). When, precisely, is less clear. According to some commentators, the use of “tests” to select civil servants (Keju) goes back as far as 2200 BC (Du Bois 1970). However, this claim is based on references that are found in Chinese classics (for example, *Book of Rites*) that were not written until about 400 or 300 BC and are examples of crediting later social institutions with an earlier existence, sometimes associated with a divine or miraculous creation (Yang, An, and Turner 2008; Wu 1982). While much that is written about the Western Zhou dynasty (1027–771 BC) is also mythological, rites and records from the era indicate that the origin of a system of universal recruitment into the civil service based on merit and involving examinations may be traced to this period (Wu 1982).

Commentators place the origin at various dates. The inauguration of a widespread system of competitive examinations by Emperor Wendi in 165 BC (Western Han dynasty), for example, is regarded by Elman (2000) as the forerunner of the elaborate imperial civil service examination set up during later dynasties. Sven and Yu (2006) identify 606 AD as the year of origin. Up to this time, non-systematic examinations were administered at irregular intervals and involved few candidates (often nominated by powerful national or local officials). Later developments during the T’ang (618–907 AD) and Song (960–1279 AD) dynasties were of particular significance as examinations were transformed from a numerically minor method of recruitment to a major, at times dominant, way of selecting officials, contributing to the breakdown of the early medieval aristocracy’s monopoly on political power (Ho 1962). Photo 3.1 depicts a civil service examination from the Song dynasty.

During successive dynasties, district, prefectural, provincial, and central examinations were held to select individuals for a wide range of positions in government and the military, including government

PHOTO 3.1**Ancient Chinese Public Examination**

Facsimile of original Chinese scroll (colored engraving) by Chinese school. Bridgeman Images/ Bibliothèque Nationale, Paris. Reproduced with permission from Bridgeman Images; further permission required for reuse.

ministers, provincial governors, education commissioners, and magistrates.

From the early Ming dynasty (probably circa 1475) until the Guangxu reform of 1898 (with some interruptions), candidates sitting imperial examinations were required to write essays based on an eight-part format (Baguwen) in response to selected quotations from the Four Books and Five Classics, the authoritative sources on Confucianism.²

The components of the eight-part essay were (a) opening, (b) introducing topic, (c) beginning discussion, (d) initial leg, (e) transition,

(f) middle section, (g) later section, (h) conclusion (Elman 2000, 394). The format was designed to test candidates' knowledge rather than indulging in obscure rhetoric (Sven and Yu 2006). It also provided uniformity in candidates' responses, which was helpful in marking their work, an important consideration as candidate numbers increased but the time available for scoring remained very limited (Lui 1974).

Features of examinations during the Ch'ing dynasty in the nineteenth century can be readily identified with conditions under which examinations are held today (Miyazaki 1981). The examinations were administered in a spacious hall or shed cut off from communication with the outside; photo 3.2 depicts an examination hall with 7,500 cells in Guangdong (Canton). Candidates sat alone at their desks (see photo 3.3); answers were written in a book of folded plain white paper, and only the candidate's number appeared on the paper. A number of features of Chinese examinations, however, have not survived. Candidates are no longer required to write through the day

PHOTO 3.2

Examination Hall with 7,500 Cells, Canton, China, 1873



Photograph in public domain.

until dark, nor are readers who grade papers prevented from leaving the hall until their work is complete.

The Chinese imperial examination system was remarkable in a number of respects, not least its longevity. By specifying the content of examinations and attaching important consequences to examination performance, cultural uniformity and consensus on basic values were maintained. The examinations were also used to identify the talent that was required to administer an empire, which they did fairly and impartially at relatively low cost, providing equal educational opportunity for all (with some notable exceptions, including women). The examination system contributed to social stratification,³

PHOTO 3.3

Examination Cells at School in Nanjing, China



Photograph by Dr. Stefan Meierhofer. Reproduced with permission (license: CC BY-SA 3.0 IGO); further permission required for reuse.

because it accepted that ruling class membership should be determined by individual merit and hard work, not family status. The actual assessment process reflected a belief that individuals could be assessed under artificial conditions, completely isolated from everyday life. It also implied that memorization (a candidate had to memorize up to 400,000 characters) and the ability to write essays were important and accurate indicators of “talent.”

By the nineteenth century, the classical education based on Ch’eng-Chu Tao learning, on which examinations were based, had lost its appeal (Franke 1960). Furthermore, the validity of the Keju was increasingly being questioned. For example, was the knowledge of Confucian thought assessed in the eight-part essay an adequate indication of the capacity of individuals to make the kinds of judgment a government official would be required to make (Lui 1974)? In response, reforms that required the “decanonization” of classical studies in favor of western learning were proposed (Lui 1974). Eventually, in 1906, discontinuation of the examination system was ordered by the dowager empress, to be replaced by school-based examinations. Details remained to be worked out when the dynasty fell in 1911.

The Chinese imperial examination system was throughout its long history criticized for its negative (if unintended) consequences (Sven and Yu 2006). The first focus of criticism recognized the role played by rote memorization: candidates preparing for the examination were supported by an array of test coaching books and learned to reproduce successful model performances without actually understanding the content of what they were writing. Second was the focus on test-taking skills (for example, the art of writing) rather than on knowledge. Third, the examinations gave rise to rampant cheating (for example, hiring substitutes to take the examination, bringing notes into the examination, and bribing officials). Finally, many anecdotal accounts and descriptions in literary works attest to many cases of mental disorders attributed to repeatedly failing the Keju examination. Efforts to deal with these consequences (for example, changing examination tasks, body searching, severe punishment for cheating) were of limited value.

In the system that was developed subsequently, high school graduate certificates were administered by individual schools. By 1985, however, the residual impact of the imperial examination

system was in evidence when Shanghai took the lead in establishing a high school graduate examination. By 1992, this type of examination, which was administered by provincial departments of education, was compulsory throughout the country (Han 1995; Yang 1995).

A similar move from examinations set by individual institutions to ones that were externally set occurred at the tertiary level. Up to 1952, the matriculation examinations for entry to third-level education were set by individual colleges and universities. In that year, the *gaokao*, the nationwide Higher Education Entrance Examination, was established and has been administered annually since, except for the period of the “cultural revolution” (1966–76), by the State Education Commission. Only individuals who have been awarded high school graduate certificates were eligible to take the Higher Education Entrance Examination (Han 1995; Yang 1995). Reaction to the overemphasis on rote learning and *gaokao* scores, together with the evidence of corrupt practices in college admissions, prompted the State Council in 2014 to authorize a series of separate phased pilot examination programs in cities and provinces, which were expected to be fully implemented by 2020 (China Policy 2019).

EXAMINATIONS IN EUROPE, 1400–

It is sometimes assumed that knowledge of the Chinese examination system provided by missionaries and travellers in China in the seventeenth century led to the introduction of written examinations in European schools. However, although the tradition of written examinations in Europe does not go back as far as in China, there was widespread use of examinations in schools at the time. Examinations featured in the method of education of the University of Paris, where they were introduced in Montaigne College in the late fifteenth century. This was based on the ideas of Johannes Cele (a schoolmaster in the Hanze city Zwolle at the end of the fourteenth century), who devised an educational system of classes, examinations, and grouping of students on the basis of “mastery” (Codina Mir 1968). In the liberal arts curriculum developed by Johannes Sturm (1507–89) of the Brethren of the Common Life for a graded school in Strasbourg, each

class had a single teacher and formal examinations were used to determine promotion (Farrell 1938; Whitehead 2007). In 1553, some 30 years before Matteo Ricci (the first superior of the Jesuit mission) arrived in China, Ignatius Loyola, the founder of the Jesuit order of priests, established the Collegium Romanum, which was based on the Paris program of education and was to become a laboratory for the development of the school system of the Jesuits. The Latin schools that made up this system spread rapidly through Europe, including to Protestant areas (such as Strasbourg and Geneva). In the seventeenth and eighteenth centuries, virtually every town in Europe had a Latin school for male students aged mostly between 6 and 18 years.

Evidence relating to examinations is available in descriptions of the *Ratio Studiorum*, which was published in its final form in 1599 and set out instructions for administration (sequence, gradation), curriculum, methodology, and discipline in schools run by the Society of Jesus (Jesuits) (Farrell 1938; McGucken 1932; Jesuits 1599, 2005). The *Ratio* provided for written and oral examinations. The wording of the rules for written examinations, excerpts of which are presented in box 3.1, is not too dissimilar to one that might be issued by a current examination agency.

BOX 3.1

Excerpts from Rules for Written Examinations, 1599

1. ...absentees...will receive no consideration in the examination unless their absence was owing to exceptional circumstances.
2. ... no one may speak to another...
3. ...come supplied ... necessary writing materials...
4. Ambiguous expressions will be construed unfavorably, and words omitted or hastily altered to avoid a difficulty will be counted as errors.
5. Seat-mates must be careful not to copy from one another.
6. ...any student who...is permitted to leave the room must deposit with the prefect...whatever he has written.
7. After a student has finished his writing assignment, he should remain at his desk and carefully check over his work, make corrections and revisions until he is satisfied.

Source: Farrell 1970.

A month's "strenuous" review before an examination was specified. Rules were provided for two types of examination. One seems to have been part of the normal cycle of teaching; the other involved competition for prizes among students. In the latter, papers were read by three "qualified and mature judges" who had to agree (by a majority vote) on an order of merit.

It is clear from an account written by Ricci in 1615 of the examination system in China that he was, not surprisingly, familiar with examining processes in Europe. In concluding his account, he described something that Europeans "might seem to be a rather strange and perhaps a somewhat ineffective method":

The judges and the proctors of all examinations, whether they be in military science, in mathematics, or in medicine, and particularly so with examinations in philosophy, are not always chosen from the senate of philosophers, nor is ever a military expert, or a medical doctor added to their number. The wisdom of those who excel in the profession of ethics is held in such high esteem that they would seem to be competent to express a proper judgment on any subject, though it be far afield from their own profession. (Ricci 1953, 41)

Examining in European universities, introduced to the University of Bologna at a period extending from 1219 to the eighteenth century, was mainly oral, consisting of questions and answers, disputation, defense of theses, or delivery of a public lecture (Teng 1943). There were, however, periodic references to the use of writing in examinations: the presentation of theses in writing in Padua around 1400 (Perreiah 1984) and the written element in Cambridge fellowship examinations in 1560 (Stray 2001). By the eighteenth century, increased student numbers and the problem of oral testing in some subjects (for example, mathematics) were pointing to the need for a change from oral to written examinations. Around 1725, an oral examination in mathematics was introduced in Cambridge to place students in order of merit. Later, candidates wrote their answers to dictated questions. Around 1790, the questions were printed. In 1792, William Farish, Professor of Chemistry and Natural Philosophy at Cambridge, developed the concept of grading students' work quantitatively (Stray 2001). Thus "was born the Cambridge

Mathematical Tripos,⁴ the grandparent of every university examination in the world.”⁵

In 1800, the Examination Statute of the University of Oxford marked the first step in that university’s examination reform (Guy 1963; Montgomery 1965; Teng 1943). Through the nineteenth century, the use of written examinations was gradually extended in both Oxford and Cambridge, often in the face of considerable opposition. Examinations outside the universities also grew, inspired by Benthamite principles of maximizing aptitude, minimizing expense, and controlling nepotism and patronage.⁶ James Booth, (vice-principal, the Liverpool Collegiate Institute) spelled out the benefits of external examinations. They would support a uniform system of education for the “middle classes”; they would encourage effective methods of teaching; they would influence education in the way that Oxford and Cambridge Universities had influenced teaching and learning in the public schools; they would promote the formation of habits of diligence and self-control; and they would provide a standard of excellence toward which middle-class schools could aspire (Roach 1971).

The growth of examinations was heavily influenced by the universities, partly because of the prestige of those institutions and partly due to the fact that those responsible for making decisions had themselves experienced examinations in the universities. Well-known figures such as William Gladstone and Robert Lowe were just some of the young men leaving Oxford and Cambridge to spread the competitive methods of the universities throughout public organizations (politics, the civil service, and the church) (Montgomery 1965). The examinations contributed to the influence of the universities and embellished their reputations at a time when they were perceived to be falling out of touch with a rapidly changing society (Roach 1971).⁷

The establishment in 1858 of examinations by the University of Oxford Delegacy and the Oxford and Cambridge “Locals” were of particular significance because it is from these that the present system of external examinations in the United Kingdom can be most directly traced (Mathews 1985). The course of development, however, was not straightforward. At first, schools were not involved; the

contract for the examination was between the university and the candidate. In time, the fact that the bulk of candidates in a center might have come from one school led to the establishment in 1862 of a parallel system of “school examinations” linked to inspections administered by the universities. These were based on schedules of study prepared by a school, but teachers marked the answers. Although these examinations continued for some time, they did not prosper to the same extent as the Oxford and Cambridge School Examinations Board’s (“Joint Boards”) examination system, established in 1874, in which entry was made by schools on behalf of candidates. The same examination was taken by different schools, and teachers played no part.

Other universities also established procedures for examining local secondary school students (Durham University 1858; University of London 1859), strengthening their control of the examinations during the second half of the nineteenth century (Brereton [1944] 2014; Montgomery 1965, 1978; Roach 1971; Wilmott and Nuttall 1975; Wiseman 1961). The twentieth century, however, saw increasing government control and centralization, together with a decrease in diversity.

In continental Europe, large-scale group written examinations were first used in the eighteenth century for selection to the civil service, as had been the case in China. Prussia led the way in 1748, and the practice was adopted in France in 1793 following the revolution.⁸ At the school level, as in Britain, universities played an important role in the introduction of written group examinations. Until the eighteenth century, each university in Prussia determined entrance qualifications. Following its introduction in 1788, the Abitur soon became a qualification examination for university entrance. In France, higher educational institutions administered the Baccalauréat examination, which had been established by Napoleon in 1808, and used it for making admission decisions. At first an oral examination, it became a written examination in 1830.⁹

The spread of examinations was not without its critics. By the 1870s, a variety of problems associated with examinations and the competitive environment in which they flourished had

been identified. It was argued that the most privileged (students who attended the “best” schools) had benefited most. Attention was directed to the evil of cramming, the focus on memorization and recall, the penalizing of independent thought and originality, and the belief that a high level of performance only meant that students possessed the abilities required for examination purposes (Roach 1971).

Today, external (public) examinations are a key feature of the educational system of most countries in Europe (EURYDICE 1999; Madaus and Kellaghan 1991; West, Edge, and Stokes 1999). In 2015, the OECD reported that 31 of 38 industrial economies held national examinations at the upper secondary school level and 14 at the lower secondary school level (OECD 2015, chart D6.1). These figures represent an increase since 2007 at the upper secondary level (from 21 of 30 listed economies) and a decrease at the lower secondary level from 17 to 14.

PUBLIC EXAMINATIONS IN DEVELOPING ECONOMIES

Many countries in Asia, Africa, and the Caribbean have public examination systems (Bray, Clarke, and Stephens 1986; Hill 2013; Kanjee 2012; Kellaghan and Greaney 1992; UNESCO Education Policy and Reform Unit 2013). As was true elsewhere, most countries in South Asia have a long legacy of using examinations to make high stakes decisions about who gains access to scarce opportunities at the next educational level (Dundar et al. 2014). In general, countries followed the practice of metropolitan countries in establishing systems in which examinations are formal, terminal, subject-based, and external to the school. As far back as 1857, as in Britain, examinations were held by universities in India (at Bombay, Calcutta, and Madras) (Eisemon 1990; Mukerji 1974). Irregularities led to the demise of the system, and by the turn of the century final school examinations were being conducted by state education departments. Concern about the quality, validity, and reliability of examinations was expressed over many years in a series of commissions that had been charged with reviewing the educational system. Following independence in

Bangladesh, India, and Pakistan, numerous suggestions were made relating to the need to eliminate what was considered to be an excessive element of chance and subjectivity in examinations, to assess higher levels of cognition, to deemphasize the role of memorization, and, in general, to reduce the influence of examinations on student learning. Several countries (Bhutan, Mongolia, Myanmar, and Nepal) have introduced new examinations or reformed existing ones since 2000 (UNESCO Education Policy and Reform Unit 2013, table 5).

In Africa, examinations have played important roles since the middle of the nineteenth century in controlling the disparate elements of educational systems, much of which were under private management, and in selecting students for scarce educational and vocational positions (Kellaghan 1992). Even before examinations were used to assess school students, they were used in the Arab Republic of Egypt in the early 1800s by the French and Ottoman Turks to select candidates for the army, and later in the century by the British to select civil servants (Hargreaves 2001). Up to and into the twentieth century, students sat for examinations set and marked in the metropolitan countries. In former British colonies, students sat examinations for which they were awarded school certificates (Kellaghan 1992). In former French colonies, students sat for the Baccalauréat following study of French curricula, which had been subjected to some adaptation to local conditions (Bray, Clarke, and Stephens 1986; Hawes 1979). Over time, examinations were partially localized, either by adapting overseas examination papers or increasing the role of local markers. In some countries, the examinations were completely localized. In others, ties were maintained with metropolitan countries through the provision of consultancy services, moderation, and printing of examination papers.

An intermediate position between autonomy and dependence on more developed systems of examinations can be found in local international examination boards, which are mostly—but not exclusively—composed of small countries. The West African Examinations Council (WAEC), the Caribbean Examinations Council (CXC), and the South Pacific Board for Educational Assessment (SPBEA) were established with the intention of sharing technical expertise and securing international recognition of the qualifications they confer. The SPBEA was established in 1980 and has nine member countries.¹⁰ The WAEC,

created in 1952, has five member countries.¹¹ The CXC, created in 1972, has 16 “participating territories” with headquarters in Barbados and an administrative operational center in Jamaica.¹² A number of innovative features have been attributed to examinations designed by the CXC: extensive syllabus development with an emphasis on critical thinking skills, teacher involvement in the design of examinations, teacher professional development in assessment, use of multimodal formats in examinations, and an elaborate school-based assessment system for most subjects (DeLisle 2009).

Until the foundation of these bodies, countries used metropolitan examinations. Their establishment demonstrated how even small states can collectively meet needs the majority might not be unable to achieve on their own (Augier and Irvine 1998). On the negative side, membership requires considerable financial input; the tensions arising from variation in the perspectives of members have been in evidence from time to time (Ndure 1998). Particular problems arose in the SPBEA, where, under the influence of Australia and New Zealand, the assignment of a substantial role to school-based assessment resulted in strains on teachers and problems in ensuring interschool and inter-country comparability of standards (Rees and Singh 1998).

Other developments in Africa include the introduction of multiple-choice questions to replace free-response items (which, it has been claimed, led to overemphasis on knowledge and skills that can be easily measured), broadening the scope of examinations by including school-based assessment, providing feedback to schools, and linking examination results with important sanctions directed at individuals, groups, or institutions (Kellaghan 1990; Kellaghan and Greaney 1992).

EXAMINATIONS IN FORMER SOCIALIST COUNTRIES

Growth in centralized systems of external examinations has been a feature of educational systems since the 1990s in former socialist countries in central, eastern, and southern Europe (Bethell 2010). In the decades up to this time, the assessment of students in primary and secondary schools had been largely oral and delegated to schools. To carry out their assessment tasks in some countries, schools set up examination boards

composed of teachers and school administrators. Even when question papers were centrally prepared for an examination at the end of secondary education (as in Poland and Romania), teachers usually evaluated the responses of their own students. This system was considered unsatisfactory for a number of reasons. First, no information was available to ministries of education about the quality of student learning. Second, conditions were not uniform for taking examinations in schools and marking was not standardized; student performance was not comparable from school to school and could not be used to select students for further education. This created a situation in which universities set their own selection examinations, which were considered more important than examinations based on the national curriculum.

In this situation, corrupt practices were pervasive. University personnel offered private tutoring lessons (in one-to-one or small group situations) or preparatory courses (usually offered in institutions to larger groups) to students about to sit an entrance examination. Tutors, because of their privileged position, could tell students the topics of an examination; they could provide tuition in the topics; and in some cases they were members of examination committees that made decisions about student selection. In these situations, the money paid for tuition could be considered a bribe to gain admission, as it was completely disproportionate to the cost of tutoring. Teachers who tutored students did not have the same privileged access to the selection process that university staff may have had. Finally, tuition by teachers presented opportunities for corrupt practices. They could put pressure on students to take private lessons by, for example, failing to cover aspects of the curriculum in class.

The introduction of examination systems at the end of secondary education in former socialist countries was designed to achieve a number of objectives: to increase government control of what was assessed; to eliminate the practice of students having to sit examinations at several universities (a practice that had a negative impact on rural and less-well-off students in particular); and to address corrupt practices (Bolotov et al. 2013). Two approaches were adopted. In one, a certification examination was established that was separate from a centrally controlled examination in order to select students for further education (for example, Azerbaijan and Georgia). In an alternative approach,

similar to the situation in many western European countries, formal examinations were administered in compulsory subjects (usually the national language and mathematics) and a number of optional subjects and results were used both for certification of student achievement at the end of secondary schooling and selection for places in higher education (for example, Albania, the Russian Federation, Slovenia). The examination reform initiatives introduced in Russia during the first decade of the twenty-first century required students to take a federally managed examination at the end of grade 9 (the Final State Certification) to certify that they had met the requirements of basic education (Bolotov et al. 2013). The results of the Unified State Examination, also managed at the federal level and taken by almost a million candidates at the end of secondary education, are used not only for certification and selection to university, but also to identify the need for pedagogical change; to hold regions, schools, and teachers accountable; and to monitor educational quality (Tyumeneva 2013).

EXAMINATIONS IN LATIN AMERICA

Education throughout Latin America has been characterized by extensive reforms since the 1990s. Greater responsibility for the management and funding of education has been accompanied by increased central control and monitoring of outcomes. However, public curriculum-based examinations have played little part in this. The preference has been for tests similar to those used in the United States for the assessment of individual students and national, regional, and international assessments of student achievement for system assessment. For example, multiple-choice tests (for example, Spanish versions of the SATs) have been used since the 1980s in Colombia and Mexico to select students for high-demand programs in higher education (Rizo 2010).

An exception to this practice is a national examination at the end of secondary school (Exame Nacional do Ensino Médio), which was introduced in Brazil in 1998 (Guimarães de Castro 2012). The examination has features of European-type examinations, with its dual purpose of selecting students for third-level education and certifying the completion of secondary school, and of North American testing

with its almost total reliance on multiple-choice items, the use of which had become very popular in Brazil since the 1960s. In addition, students in a relatively small but increasing number of private schools and some high-performance selective public high schools are taking the International Baccalaureate in various countries in the region (IB 2019).

Despite the lack of a tradition of large-scale methods of assessment, the fact that countries in Latin America have enthusiastically embraced national and international (global and regional) assessments of educational systems may suggest openness to the type of public examinations common in other parts of the world. Indeed, the elaborate census-based national assessment system that was established in Chile as far back as 1978 and a more recent system in the Dominican Republic clearly serve some of the functions of a public examination (for example, assessment of individual students), as well as those of a system assessment.

EXAMINATIONS IN THE UNITED STATES

Horace Mann, faced with the impossibility in 1845 of “committee men” conducting oral examinations for “over 7,000 children” in the Boston Public Schools, saw how written essay examinations could be used for this purpose. It is perhaps not a coincidence that Mann had spent 1843 in Europe on a one-year leave of absence from his post as Secretary of the Massachusetts Board of Education. For Mann, the Kingdoms of Prussia and Saxony (where he spent time) were preeminent in the quantity and quality of their education. In addition to facilitating the examination of large numbers of students, Mann recognized that the examinations allowed the examiner to pose an identical set of questions to students, under similar conditions, in a limited time frame, producing “comparable” scores, although he did not seem to hit on the idea embodied in the Chinese practice of administering the same examinations to everyone simultaneously (Madaus and O’Dwyer 1999; Morris 1961).

Reliance on the essay-type examinations used by Mann was cut short by the invention, attributed to Frederick J. Kelly, of the

multiple-choice format in 1914. Use of this format meant that students were no longer required to construct or produce even a short answer, but simply had to select the correct question from competing alternatives. The use of the “new type” tests based on the multiple-choice format spread rapidly through American school systems in the 1920s, largely because of their efficiency and cost-effectiveness. Their perceived benefits together with the fact that normative data on student performance were provided with tests had a number of consequences. First, public examinations of the type available in other parts of the world did not become a feature of education in the United States, with the exceptions of Advanced Placement examinations and the Regents Examination in New York State. Since 1865, the latter has offered syllabi and a system of examinations to students wishing to obtain a Regents Diploma. In the absence of public examinations, individual schools provided certification of student achievement. Second, the availability of normative data shifted the focus in tests from curriculum content or skills assessed to student performance relative to those of other students. One might have expected the Carnegie-funded International Examinations Inquiry, which brought together educators from the United States and Europe,¹³ to have brought American and European assessment practices closer together (Lawn 2008; Baird and Opposs 2018). These conferences took place in the United Kingdom in 1931 in Eastbourne and in 1935 in Folkestone, and in France in 1938 in Dinard. However, the assessment approaches remained divergent, although in time multiple-choice items did find their way into many European public examination systems, sometimes as the sole item type, sometimes in combination with other types.

By the 1980s and 1990s, a “national” examination system similar to systems elsewhere was being proposed for the United States. It was claimed that the lack of common national (public) examinations in the United States was a key reason why students performed poorly in international comparisons of student achievement. An external examination system was perceived to have many benefits. First, it motivates students to work because performance has real consequences. Second, examinations provide teachers with clear and meaningful standards. Third, examinations define achievement relative to

an external standard, not relative to other students in a classroom or school. Fourth, examinations are keyed to the content of specific course sequences. Fifth, examinations signal multiple levels of achievement in a subject, not just a pass-fail grade that allows many students to pass without exertion. Sixth, examinations cover all secondary school students, not just students following advanced courses or in elite schools. Seventh, examinations assess a major portion of what students studying a subject are expected to know or be able to do (Bishop 1998a, 1998b, 2005).

The United States does not have a national examination system primarily because of political opposition to federal control of education. However, it does have a system of statewide assessments, based on standardized tests that must be administered in public schools if states are to receive federal funding. The results of the tests are used for a variety of decisions about students, schools, school systems, and teachers. The examinations, which are called assessments or tests, however, differ in many respects from traditional external examinations in other countries.

First, the tests are mandated by local authorities (for example, large school systems), state departments of education, or state legislatures under the No Child Left Behind (NCLB) Act of 2001 and its successors *Race to the Top* (2009) and the *Every Student Succeeds Act* (ESSA) of 2015.¹⁴ They are almost always developed, administered, and scored by testing companies under contract to the authorizing agency. Second, testing in the United States differs from practice elsewhere in the range of grades in which assessments are carried out. NCLB and ESSA provide for testing in a much wider range (yearly in grades 3 through 8, plus once in high school in math and reading, and less frequently in science) than is normal in public examination systems elsewhere. Third, while the tests are geared to state curriculum frameworks or standards, the emphasis at the elementary grades is on skills rather than content. Fourth, the tests have a large multiple-choice component in addition to a small number of supply-type short-answer or essay items and are supported by sophisticated psychometric analytical techniques.

Fifth, while the main function of testing is to support standards-based reform, and while ESSA removed the link between testing

programs and the high-stakes decisions associated with NCLB—which had required schools to raise reading and mathematics test scores every year or face penalties—decisions based on performance are wider than is generally the case for public examinations. They include measures that allow state intervention to assign additional resources to schools where subgroups of students are “struggling.” Schools are required to compare the level of achievement of their students with state-level achievements and report it to the state, while states are required to report the level of state achievement to the federal government.

Sixth, student performance is generally reported in the form of performance categories or standards (for example, *advanced*, *proficient*, *basic* or *needs improvement*, and *unsatisfactory* or *fail*), which are arrived at through the use of a number of standard-setting methods (Hambleton 2001; Horn et al. 2000; Kane 2001; Raymond and Reid 2001). Scores (and standards) are linked from one year to another through the use of item response theory technology (Hambleton, Swaminathan, and Rogers 1991; Mislevy 1992).

Seventh, about 50 percent of states require students to pass what are termed “exit tests” prior to graduation. Usually criterion referenced, these tests are designed to ensure that students have achieved basic competence in core curriculum areas. They tend to be less challenging than school leaving questions in countries with established public examination systems.

Eighth, although in many countries entrance to third-level institutions is based on performance on a public examination, in the United States the tradition of using aptitude (SAT 1) and achievement (SAT 2) tests administered by the College Board continues. Results of these, or of ACT tests, are used in conjunction with class rank, class grades, and other factors in making admission decisions.

Ninth, test authorities in the United States follow the tradition of public examinations in releasing test items each year. However, 10 to 20 percent of items may be withheld for year-to-year linking. Finally, in many public examinations at the secondary school level in other countries, students choose from a large range of subjects and also can

choose the items they wish to respond to in an examination; in most places in the United States, such choice is not available, and all students take examinations in the same subjects (such as language, mathematics, and science).

INTERNATIONAL EXAMINATIONS

Growth in the number of students taking international examinations in recent years has been prompted by a variety of factors: an increase in population mobility, schools opting for more demanding curricula, and scepticism about the credibility of national examination qualifications, especially in the context of students seeking admission to overseas third-level institutions.

The International Baccalaureate (IB), established in 1968, is perhaps the most widely known nonnational school leaving examination. The IB curriculum, designed for students aged 16 to 19, was developed independently of national governments and national curriculum agencies and emphasizes critical thinking and challenging of assumptions. It consists of six subject groups covering sixteen subjects and a core that includes theory of knowledge and an extended essay. Students are required to complete independent research and undertake a project that involves social or community service. Final marks are based on external and, in most instances, teacher assessments. In 2018, IB reported that 2,790 schools in 143 countries participated in the IB diploma program designed for students aged 16 to 19, an increase of 26 percent in the number of schools since 2014 (IB 2018).

Participation in IB diploma programs tends to be expensive. Examination candidates must pay separate registration and examination fees. The registration fee for the period September 1, 2017, to August 31, 2018, was US\$172, while the fee for each subject taken was US\$119. Participating schools also paid an annual fee of US\$11,650. In the United States, some school systems have supported participation in an IB program. However, some US and UK schools have opted out of the IB for financial and other reasons.

A number of British examination boards offer international versions of their school leaving examinations. These include Cambridge Assessment International Education (CIE, from its former name, Cambridge International Examinations) and Pearson-owned Edexcel (Education and Excellence). CIE is administered by Cambridge University, which established the University of Cambridge Local Examinations Syndicate (UCLES) to administer examinations to non-Cambridge students as far back as 1858 as a means of raising educational standards. The not-for-profit CIE administers school leaving examinations in 160 countries (Cambridge Assessment International Education 2018; UCLES 2007). Examinations administered at AS (Advanced Supplementary) and A (Advanced) levels are considered equivalent to the AS- and A-level examinations administered by examination boards in the United Kingdom and are accepted for entrance to universities in many countries. Students and schools may opt to combine Cambridge subject results with results from their national examinations.

As in the case of the International Baccalaureate, CIE examinations emphasize problem solving and analysis and the ability to read texts critically and to communicate results. They tend to have a limited number of multiple-choice items, along with essay-type questions, tasks that require analysis, practical assignments, practical assignments in science labs, and oral and listening tasks.

Schools wishing to enroll in a Cambridge program must meet criteria in relation to their mission and educational values, management and leadership, quality of teaching, and physical environment and facilities (such as having facilities for art, music, languages, and information and communications technology).

Schools must pay a registration fee and annual fees to gain access to online materials and training services. CIE also publishes hardcover materials. Candidates pay a fee, which varies by subject and location for each examination entry. In the United States, fees typically run between US\$78 and US\$86 per subject. The norm is for students to take examinations in three or four subjects. CIE has worked with such countries as Botswana, Mauritius, and Namibia to prepare examinations with a particular national focus and with Singapore to prepare examinations leading to joint Singapore-Cambridge Examination certificates.

CONCLUSION

An outstanding feature of the nineteenth and twentieth centuries has been the growth in the number of educational systems that administer public examinations. These examinations are now to be found in most parts of the world, in many cases replacing school-based assessment. As student retention rates increased, the number of levels at which examinations are held decreased. Many countries have now discontinued public examinations at the end of primary school, and several do not administer them before the end of secondary schooling. It would seem that when high stakes are attached to an assessment, however, the preference is for an external assessment.

Many of the features of the examination system of Imperial China can be found in the examination systems that spread throughout the world in the nineteenth and twentieth centuries: defining culturally valued knowledge (privileging “book knowledge” rather than practical ability); rewarding “talent”; providing equality of educational opportunity; contributing to social stratification; focusing on memorization and essay writing; and, in many systems, separating assessment from other educational activities. There were, however, differences between the Chinese systems and examinations systems that developed in Europe. First, while competition and norm-referenced decisions were features of both, the high failure rate in China meant that competition loomed larger in that country. Furthermore, the use in Europe of examinations to determine entry to a profession implied a standards-based approach. Second, an underlying assumption in the Chinese systems, testified to by the large numbers of candidates who repeated examinations, was that effort and hard work led to success. In Europe, on the other hand, the use of examinations to select and classify students at an early stage in their careers for differentiated education provision reflected a belief in the important role played by ability in determining students’ scholastic progress.

Although examination systems throughout the world share many features, it is of interest to note differences between them. Of particular interest is the contrast, noted at the conferences in the 1930s (Lawn 2008) and that continues to exist, between the use of essay-type examinations in Europe and multiple-choice tests in the

United States and Latin America, as well as in the methods used to select students for further education.

Although external public examinations have been in operation in some European countries for as long as two centuries, the United States is a newcomer to the scene. The manner in which testing programs are being implemented in American states suggests a limited awareness of the nature, variety, and history of public examinations. American programs differ from practice in most European countries in their total reliance on evidence obtained from an external test, in a heavy reliance on multiple-choice tests, in the testing of elementary school students, in the narrow range of curriculum areas assessed, and in the lack of choice (either to take a test or in the test taken). Following ESSA, there may be a move away from a sole focus on performance on standardized tests, allowing for the use of multiple measures of student learning, along with other indicators of student success in making school accountability decisions.

As one observes the development of public examinations, it is interesting to note the resilience of claims of strong links between external examinations, school achievement, and global competitiveness. In the mid-nineteenth century, Professor Liebig of Giessen in Germany, Britain's then main industrial competitor, said that "if no examination is introduced the best schemes will fail, and produce no effect: introduce the examination, and all the rest follows of itself" (Foden 1989, 74). Liebig's view from Germany echoed a view—still popular today—that students and teachers need to be motivated, that competition improves motivation and learning, and that examinations are a necessary and cost-effective means of raising educational standards and securing national competitiveness (Madaus and Kellaghan 1991).

It is often claimed that examinations were also introduced to promote meritocracy and equity. It can be argued that the reforms of Frederick the Great and Napoleon were less concerned with such issues than with a desire to build a modern unitary state, staffed on the basis of talent (Eckstein and Noah 1993). In Britain, it can also be argued that Gladstone's determination to effect change was influenced not a little by the obvious administrative shortcomings exhibited by the army and civil service during the Crimean War

(Montgomery 1965). In light of these observations, it is difficult to make the case that examinations, whatever the motivation in their introduction, played a major role in the promotion of equity in educational systems or in the distribution of life chances. Insofar as social mobility was a feature of life over the past century and a half, it would seem to owe more to the provision of schools than to the institution of examinations, which probably did little more than reflect students' access to, and earlier success in, the educational system.

NOTES

1. Korea and Vietnam had similar systems.
2. The Four Books: Analects of Confucius, Great Learning, Doctrine of the Mean, Book of Mencius. The Five Classics: Poetry/Songs, Book of Documents, Book of Rites, Book of Changes, Spring and Autumn Annals.
3. Mencius (371–289 BC) set out the Confucian theory of social stratification: “Those who labor with their minds rule others, and those who labor with physical strength are ruled by others” (Ho 1962, 17).
4. The Tripos was an examination at Cambridge University to qualify for a bachelor's degree. The name may have come from the fact that students used to sit on a three-legged stool while taking oral examinations.
5. A UK Board of Education report (Acland 1911, 6) noted, “It has been stated that the first written examination known in Europe was introduced by R. Bently in 1702 at Trinity College Cambridge. It seems doubtful, however, whether further research would not reveal the existence of earlier written examinations.” Teng's (1943, 273) conclusion is less circumspect: “The year 1702 remains the reliable date to mark the beginning of the written examination in Europe.” However, it fails to take account of the use of written examinations in secondary schools.
6. Many bodies were involved in examining during the second half of the nineteenth century. These included the College of Preceptors, the City and Guilds of London Institute, the General Medical Council, the Royal Institute of Chemistry, the Royal Institute of Civil Engineers, and the Royal Society of Arts.
7. The London University Matriculation examinations (1838) preceded the Oxford and Cambridge Locals. However, they differed in function. While the original incentive to pass “Locals” was to obtain a certificate granted by a university, the incentive to pass the London Matriculation examination was to enter a course of study for a degree. The Matriculation

examination gradually became a leaving examination for secondary schools, certifying “a good school education” (Brereton 1944).

8. The examinations lapsed after 10 years, but were reestablished in the 1840s. Civil service examinations were not established in England until 1850.
9. Halls (1965) traces the origin of the Baccalauréat to the thirteenth-century *determinance* of the Sorbonne University, which was an oral examination to decide whether students were fit to embark on university studies.
10. Fiji, Kiribati, Nauru, Tokelau, Tonga, Tuvalu, Samoa, the Solomon Islands, and Vanuatu. Australia and New Zealand are the main donors.
11. The Gambia, Ghana, Liberia (joined in 1974), Nigeria, and Sierra Leone (Ndure 1998).
12. Anguilla, Antigua and Barbuda, Barbados, Belize, the British Virgin Islands, Cayman Islands, Dominica, Grenada, Guyana, Jamaica, Montserrat, St. Kitts and Nevis, St. Lucia, St. Vincent and the Grenadines, Trinidad and Tobago, and Turks and Caicos Islands (Augier and Irvine 1998).
13. Educators from England, France, Germany, and Scotland attended the first meeting. Finland, Norway, and Sweden joined later.
14. See “Reauthorization of the Elementary and Secondary Education Act,” U.S. Department of Education, Washington, DC: (a) “No Child Left Behind Act, 2001”; (b) “Race to the Top Program, 2009”; (c) “Every Student Succeeds Act (ESSA) 2015.”

REFERENCES

- Acland, A. H. D. 1911. *Report of the Consultative Committee on Examinations in Secondary Schools*. London: Her Majesty's Stationery Office
- Augier, R., and D. Irvine. 1998. “Caribbean Examinations Council.” In *Examination Systems in Small States: Comparative Perspectives on Policies, Models and Operations*, edited by M. Bray and L. Steward, 145–61. London: Commonwealth Secretariat.
- Baird, J.-A., and D. Opposs. 2018. “The Standard Setting Project: Assessment Paradigms.” In *Examination Standards: How Measures and Meanings Differ around the World*, edited by J.-A. Baird, T. Isaacs, D. Opposs, and L. Gray, 2–25. London: University College London Institute of Education Press.
- Bethell, G. 2010. “Student Assessment: Policy and Practice in Eastern Europe.” In *International Encyclopedia of Education*, 3d ed., edited by P. Peterson, E. Barker, and B. McGaw, 472–78. Oxford: Elsevier.
- Bishop, J. H. 1998a. “Do Curriculum-Based Exit Exam Systems Enhance Student Achievement?” CPRE Research Report Series RR-40, Consortium

- for Policy Research in Education, Graduate School of Education, University of Pennsylvania, Philadelphia.
- . 1998b. “The Effect of Curriculum-Based External Exit Exam Systems on Student Achievement.” *Journal of Economic Education* 29: 171–82.
- . 2005. “High School Exit Examinations: When Do Learning Effects Generalize? In *Uses and Misuses of Data for Educational Accountability and Improvement: 104th Yearbook of the National Society for the Study of Education*, part 2, edited by J. L. Herman and E. H. Haertel, 260–88. Chicago: National Society for the Study of Education.
- Bolotov, V., G. Kovaleva, M. Pinskaya, and I. Valdman. 2013. *Developing the Enabling Context for Student Assessment in Russia*. Washington, DC: World Bank.
- Bray, M., P. B. Clarke, and D. Stephens. 1986. *Education and Society in Africa*. London: Arnold.
- Brereton, J. L. (1944) 2014. *The Case for Examinations: An Account of Their Place in Education with Some Proposals for Their Reform*. Cambridge, UK: Cambridge University Press.
- Cambridge International Assessment Education. 2018. “What We Do.” <http://www.cambridgeinternational.org/about-us/what-we-do/>.
- China Policy. 2019. “Gaokao: Reform and Reality Clash.” *CP Signal*, January 25. <https://policyn.com/?r=https%3A%2F%2Fpolicyn.com%2F%2F19-01-24-gaokao-reform-and-reality-clash%2F&login=true>.
- Codina Mir, G. 1968. *Aux sources de la pédagogie des Jésuites: le “Modus Parisiensis.”* Rome: Institutum Historicum Societatis Iesu. <https://archive.org/details/bhsi28>.
- De Lisle, J. 2009. “Entrance Examination beyond National Borders: Trinidad and the Caribbean Examinations Council.” In *Secondary School External Examination Systems: Reliability, Robustness and Resilience*, edited by B. Vlaardingerbroek and N. Taylor, 262–89. Amherst, NY: Cambria.
- Du Bois, P. H. 1970. *A History of Psychological Testing*. Boston: Allyn and Bacon.
- Dundar, H., T. Bêteille, M. Riboud, and A. Deolalikar. 2014. *Student Learning in South Asia: Challenges, Opportunities, and Policy Priorities*. Washington, DC: World Bank.
- Eckstein, M. A., and H. J. Noah. 1993. *Secondary School Examinations: International Perspectives on Policies and Practice*. New Haven, CT: Yale University Press.
- Eisemon, T. O. 1990. “Examination Policies to Strengthen Primary Schooling in African Countries.” *International Journal of Educational Development* 10: 69–82.

- Elman, B. A. 2000. *A Cultural History of Civil Examinations in Late Imperial China*. Berkeley and Los Angeles: University of California Press.
- EURYDICE. 1999. *European Glossary on Education: Vol. 1—Examinations, Qualifications and Titles*. Brussels: EURYDICE.
- Farrell, A. P. 1938. *The Jesuit Code of Liberal Education: Development and Scope of the Ratio Studiorum*. Milwaukee, WI: Bruce.
- , trans. 1970. *The Jesuit Ratio Studiorum of 1599*. Washington, DC: Conference of Major Superiors of Jesuits. <http://admin.wyk.edu.hk/documents/A01%20Jesuit%20Education%20%26%20Ignatian%20Pedagogy/Ratio%20Studiorum.htm>.
- Foden, F. 1989. *The Examiner: James Booth and the Origins of Common Examinations*. Leeds, UK: School of Continuing Education, University of Leeds.
- Franke, W. 1960. *The Reform and Abolition of the Traditional Chinese Examination System*. Cambridge, MA: Asia Center, Harvard University.
- Guimarães de Castro, M. H. 2012. *Developing the Enabling Context for Student Assessment in Brazil*. Washington, DC: World Bank.
- Guy, B. J. 1963. *The Chinese Examination System and France, 1569–1847: Studies on Voltaire and the Eighteenth Century*. Geneva: Institut et Musée Voltaire.
- Halls, W. D. 1965. *Society, Schools and Progress in France*. Oxford: Pergamon.
- Hambleton, R. K. 2001. "Setting Performance Standards on Educational Assessment and Criteria for Evaluating the Process." In *Setting Performance Standards: Concepts, Methods and Perspectives*, edited by G. J. Cizek, 89–116. Mahwah, NJ: Erlbaum.
- Hambleton, R. K., H. Swaminathan, and H. Rogers. 1991. *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.
- Han, N. 1995. "The Higher Education Entrance Examination in China." In *Admission to Higher Education: Issues and Practice*, edited by T. Kellaghan, 228–32. Dublin: Educational Research Centre and Princeton, NJ: International Association for Educational Assessment.
- Hargreaves, E. 2001. "Assessment in Egypt." *Assessment in Education: Principles, Policy and Practice* 8: 247–60.
- Hawes, H. 1979. *Curriculum and Reality in African Primary Schools*. Harlow, Essex, UK: Longman.
- Hill, P. 2013. *Examination Systems*. Paris: United Nations Educational, Scientific and Cultural Organization.
- Ho, P.-T. 1962. *The Ladder of Success in Imperial China: Aspects of Social Mobility, 1368–1911*. New York: Columbia University Press.

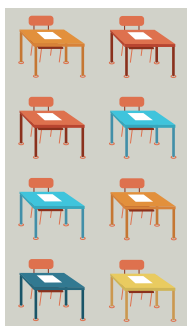
- Horn, C., M. Ramos, I. Blumer, and G. F. Madaus. 2000. "Cut Scores: Results May Vary." *NBETPP Monographs* 1 (1): 1–31.
- International Baccalaureate. 2018. "The IB Diploma Programme Statistical Bulletin, May 2018 Examination Session." International Baccalaureate Organization, Cardiff
- . 2019. "The IB by Region." <https://www.ibo.org/about-the-ib/the-ib-by-region/>.
- Jesuits and Claude Nicholas Pavur. (1599) 2005. *The Ratio Studiorum: The Official Plan for Jesuit Education*. St. Louis, MO: Institute of Jesuit Sources.
- Kane, M. T. 2001. "So Much Remains the Same: Conception and Status of Validation in Setting Standards." In *Setting Performance Standards: Concepts, Methods, and Perspectives*, edited by G. C. Cizek, 53–88. Mahwah, NJ: Erlbaum.
- Kanjee, A. 2012. "Assessment and Education Quality in Sub-Saharan Africa: Prospects and Pitfalls." Paper presented at the Public Seminar on Assessment, Oxford University, Oxford, UK, October 29. <http://oxes.org.uk/wordpress/wp-content/uploads/2013/03/Assessment-and-education-quality-in-sub-Saharan-Africa-presentation-with-timings.pdf>.
- Kellaghan, T. 1990. "Examination Systems in Africa." In *Changing Educational Assessment: International Perspectives and Trends*, edited by P. Broadfoot, R. Murphy, and H. Torrance, 98–105. London: Routledge.
- . 1992. "Examination Systems in Africa: Between Internationalization and Indigenization." In *Examinations: Comparative and International Studies*, edited by M. A. Eckstein and H. J. Noah, 95–104. Oxford, UK: Pergamon.
- Kellaghan, T., and V. Greaney. 1992. "Using Examinations to Improve Education: A Study in Fourteen African Countries." Technical Paper 165, Africa Technical Department Series, World Bank, Washington, DC.
- Lawn, M., ed. 2008. *An Atlantic Crossing? The Work of the International Examination Inquiry, Its Researchers, Methods and Influence*. Oxford, UK: Symposium.
- Lui, A. Y.-C. 1974. "Syllabus of the Provincial Examination (*hsiang-shih*) under the Early Ch'ing (1644–1795)." *Modern Asian Studies* 8 (3): 391–96.
- Madaus, G. F., and T. Kellaghan. 1991. "Student Examination Systems in the European Community: Lessons for the United States." In *Science Assessment in the Service of Reform*, edited by G. Kulm and S. M. Malcolm, 189–232. Washington, DC: American Association for the Advancement of Science.
- Madaus, G. F., and L. M. O'Dwyer. 1999. "A Short History of Performance Assessment." *Phi Delta Kappan* 80: 688–95.

- Mathews, J. C. 1985. *Examinations: A Commentary*. Boston: Allen and Unwin.
- McGucken, W. J. 1932. *The Jesuits and Education*. Milwaukee, WI: Bruce.
- Mislevy, R. J. 1992. "Linking Educational Assessments: Concepts, Issues, Methods, and Prospects." ERIC ED353302. Educational Testing Service, Princeton, NJ.
- Miyazaki, I. 1981. *China's Examination Hell: The Civil Service Examinations of Imperial China*, translated by C. Schirokauer. New Haven, CT: Yale University Press.
- Montgomery, R. 1965. *Examinations: An Account of Their Evolution as Administrative Devices in England*. London: Longmans Green.
- . 1978. *A New Examination of Examinations*. London: Routledge and Kegan Paul.
- Morris, N. 1961. "An Historian's View of Examinations." In *Examinations and English Education*, edited by S. Wiseman, 1–43. Manchester, UK: Manchester University Press.
- Mukerji, S. N. 1974. *History of Education in India (Modern Period)*. Baroda, India: Acharya Book Depot.
- Ndure, M. 1998. "West African Examinations Council." In *Examination Systems in Small States: Comparative Perspectives on Policies, Models and Operations*, edited by M. Bray and L. Steward, 181–97. London: Commonwealth Secretariat.
- OECD (Organisation for Economic Co-operation and Development). 2007. *Education at a Glance 2007: OECD Indicators*. Paris: OECD.
- . 2015. *Education at a Glance 2015: OECD Indicators*. Paris: OECD.
- Perreiah, A. R. 1984. "Logic Examinations in Padua circa 1400." *History of Education* 13: 85–103.
- Raymond, M. R., and J. B. Reid. 2001. "Who Made Thee a Judge? Selecting and Training Participants for Standard Setting." In *Setting Performance Standards: Concepts, Methods, and Perspectives*, edited by G. J. Cizek, 119–57. Mahwah, NJ: Erlbaum.
- Rees, T., and G. Singh. 1998. "South Pacific Board for Educational Assessment." In *Examination Systems in Small States: Comparative Perspectives on Policies, Models and Operations*, edited by M. Bray and L. Steward, 162–80. London: Commonwealth Secretariat.
- Ricci, M. (1615) 1953. *China in the Sixteenth Century: The Journals of Matteo Ricci, 1583–1610*, translated by Louis J. Gallagher. New York: Random House.
- Rizo, F. M. 2010. "Assessment Practice in Policy Context: Latin American Countries." In *International Encyclopedia of Education*, 3d ed., edited by P. Peterson, E. Baker, and B. McGaw, 479–85. Oxford, UK: Elsevier.

- Roach, J. 1971. *Public Examinations in England, 1850–1900*. Cambridge, UK: Cambridge University Press.
- Stray, C. 2001. “The Shift from Oral to Written Examinations: Cambridge and Oxford, 1700–1900.” *Assessment in Education: Principles, Policy and Practice* 8: 33–50.
- Sven, H. K., and L. Yu. 2006. “Chronic Consequences of High-Stakes Testing? Lessons from the Chinese Civil Service Exams.” *Comparative Education Review* 50: 46–65.
- Teng, S.-Y. 1943. “Chinese Influence on the Western Examination System.” *Harvard Journal of Asiatic Studies* 7: 267–312.
- Tyumeneva, Y. 2013. *Disseminating and Using Student Assessment Information in Russia*. SABER—Systems Approach for Better Education Results 10. Washington, DC: World Bank.
- UCLES (University of Cambridge Local Examinations Syndicate). 2007. “Memorandum Submitted by University of Cambridge International Examinations.” <https://publications.parliament.uk/pa/cm200607/cmselect/cmeduski/memo/test&ass/ucm4502.htm>.
- UNESCO (United Nations Educational, Scientific, and Cultural Organization) Education Policy and Reform Unit. 2013. *The Use of Student Assessment for Policy and Learning Improvement*. Bangkok: UNESCO Asia and Pacific Regional Bureau for Education.
- Webber, C. 1989. “The Mandarin Mentality: Civil Service and University Admissions Testing in Europe and Asia.” In *Test Policy and the Politics of Opportunity Allocation: The Workplace and the Law*, edited by B. Gifford, 33–60. Dordrecht, Netherlands: Kluwer Academic.
- West, A., A. Edge, and E. Stokes. 1999. *Secondary Education across Europe: Curricula and School Examination Systems*. London: Centre for Educational Research, London School of Economics and Political Science.
- Whitehead, M. 2007. “‘To Provide for the Edifice of Learning’: Researching 450 Years of Jesuit Educational and Cultural History, with Particular Reference to the British Jesuits.” *History of Education* 36: 109–43.
- Wilmott, A. S., and D. L. Nuttall. 1975. *The Reliability of Examinations at 16+*. London: Macmillan.
- Wiseman, S., ed. 1961. *Examinations and English Education*. Manchester, UK: Manchester University Press.
- Wu, C. L. 1982. *The Chinese Heritage*. New York: Crown.
- Yang, L., D. An, and J. A. Turner. 2008. *Handbook of Chinese Mythology*. New York: Oxford University Press.

Yang, X. 1995. "The Enrolment and Examination System to Higher Education in China." In *Admission to Higher Education: Issues and Practice*, edited by T. Kellaghan, 223–27. Princeton, NJ: International Association for Educational Assessment and Dublin: Educational Research Centre.

CHAPTER 4



THE CONSTRUCTION AND SCORING OF PUBLIC EXAMINATIONS

INTRODUCTION

This chapter describes a number of major components of a public examination: the examination syllabus; constructing the examination; the form of examinee responses; the production and printing of examination papers; and the marking of examination scripts. The use of information technologies in examinations is briefly considered.

THE EXAMINATION SYLLABUS

The terms “curriculum” and “syllabus” are used in different ways. One way is to regard the “curriculum” as referring to all elements of programs taught in schools (subjects to be taught; knowledge, skills, and attitudes to be developed for each subject; topics and content to be taught; teaching strategies to be suggested). The term “syllabus,” on the other hand, is used to refer to the document that specifies the details of an examination: the knowledge, skills, topics, and so on that will be examined and how they will be examined.

The following topics are usually included in a syllabus document:

- Aims of the examination
- Assessment objectives or domains

- Weightings of domains (in a table of specifications)
- Scheme of assessment
- Description of papers
- Table of specifications (topic weightings)
- Details of school-based element (if appropriate)
- Marking details
- Grade descriptions
- Exemplar material (examples of items or sample papers)
- Notes to guide teachers

The syllabus is an important element in ensuring the transparency of the examination. It provides information to students, teachers, examination writers, examiners, and examination administrators, acting as a contract between the assessment body and its users (see table 4.1).

The contents and structure of a syllabus are determined by the assessment body. In some countries, assessment bodies make syllabi available on their websites. Excerpts from syllabus information relating to course structure, question paper design, and marking scheme are provided in tables 4.1 and 4.2.

TABLE 4.1
Course Structure for Biology

THEORY				
Time: 3 Hours		Max Marks: 70		
	Title		No. of Periods	Marks
I	Diversity of Living Organisms		23	7
II	Structural Organization in Plants and Animals		22	12
III	Cell: Structure and Function		22	15
IV	Plant Physiology		35	18
V	Human Physiology		40	18
Example: Diversity of Living Organisms: One of four chapters in Unit 1 Chapter 1: The Living World What is living? Biodiversity; Need for classification; three domains of life; taxonomy and systematics; concept of species and taxonomical hierarchy; binomial nomenclature; tools for study of taxonomy-museums, zoological parks, herbaria, botanical gardens.				

(continued)

TABLE 4.1
Course Structure for Biology (continued)

PRACTICALS		
Time Allowed: Three Hours	Max. Marks: 30	
Evaluation Scheme		Marks
One Major Experiment Part A (Expt. No. 1, 3, 7, 8)		5
One Minor Experiment Part A (Expt. No. 6, 9, 10, 11, 12, 13)		4
Slide Preparation Part A (Expt. No. 2, 4, 5)		5
Spotting Part B		7
Practical Record + Viva Voce		4
Project Record + Viva Voce		5

Source: India, Central Board of Secondary Education 2017.

CONSTRUCTING THE EXAMINATION

Syllabi and general regulations for examinations are usually specified by a government ministry. Ideally, the actual design of an examination based on a syllabus should be an iterative process, beginning with a delineation of the conceptual framework that includes a description of the construct that will be assessed. The content and the nature and level of specific knowledge, skills, and understanding to be elicited should be described in a blueprint that will guide the development of the examination specifications (Lane and Stone 2006). In most countries, committees composed of ministry officials, officials from curriculum and examination bodies, classroom teachers, and university personnel construct examinations (Britton, Hawkins, and Ganda 1996). In some systems, provision is made for internal and external review of examination questions, translation verification, and field testing. Work on the preparation of examination papers (and marking schemes) may begin up to two years before they are required.

The construction of an examination may be based on a taxonomy (or classification system) of learning outcomes. The most popular taxonomy was developed by Bloom (1956) and his colleagues in the mid-1950s. Both it and a subsequent revision (Anderson and Krathwohl 2001) present cognitive processes as a continuum of increasing complexity ranging from remembering to creating. Airasian (1994), citing Tyler’s work (Tyler 1934), points out that because of

TABLE 4.2**Question Paper Design for Biology Examination, Class XI, 2017–18**

No	Typology of Questions	Very Short Answer (VSA) (1 mark)	Short Answer-I (SA-I) (2 marks)	Short Answer-II (SA-II) (3 marks)	Value based question (VBQ) (4 marks)	Long Answer (LA) (5 marks)	Total Marks	% Weight
1	Remembering – (Knowledge based Simple recall questions, to know, specific facts, terms, concepts, principles, or theories, Identify, define, or recite, information)	2	1	1			7	10%
2	Understanding – (Comprehension – To be familiar with meaning and to understand conceptually, interpret, compare, contrast, explain, paraphrase information)		2	4		1	21	30%
3	Application – (Use abstract information to concrete situation, to apply knowledge to new situation, Use given content to interpret a situation, provide an example, or solve a problem)		2	4		1	21	30%
4	High Order Thinking Skills (Analysis and Synthesis – Classify, Compare, Contrast, or differentiate between different pieces of information, Organize and/or integrate unique pieces of information from a variety of sources)	2		1		1	10	14%
5	Evaluation – (Appraise, judge, and/or justify the value or worth of a decision or outcome, or to predict outcomes based on values)	1		2	1		11	16%
	TOTAL	5X1	5X2	12X3	1X4	3X5	70(26)	100%

Source: India, Central Board of Secondary Education 2017.

low correlations between scores on memory tests and scores on tests of reasoning and the application of principles, one could not rely on tests of information to provide a valid indication of a student’s ability to apply, analyze, or interpret.

The taxonomy had a number of advantages. First, it provided a panorama of goals (broader than might otherwise be considered). Second, it highlighted the importance of specifying objectives involving skills and abilities as distinct from memorized knowledge. Third, it provided a framework for analyzing the relative emphasis given various levels of behavior in a curriculum or test. Fourth, it showed that some goals depend on the earlier learning of other goals (Tyler 1934).

The taxonomy has been revised (see box 4.1) and translated into many languages. Despite problems in establishing its hierarchical

BOX 4.1

Levels in Bloom’s Taxonomy of Educational Objectives for Knowledge-Based Goals (Revised)

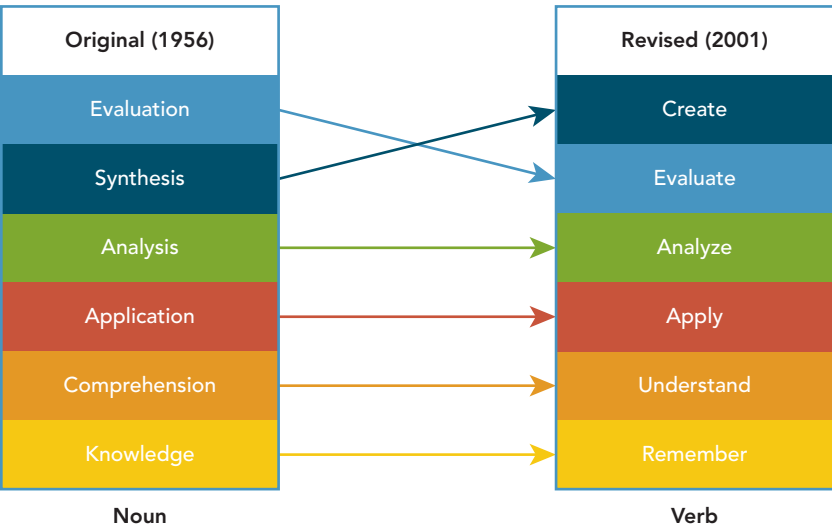
Remembering	Retrieving, recalling, or recognizing knowledge from memory. Producing definitions, facts, or lists or reciting or retrieving material.
Understanding	Constructing meaning from different types of functions, written or graphic messages, or activities such as interpreting, exemplifying, classifying, summarizing, inferring, comparing, or explaining.
Applying	Carrying out or applying previously learned facts, rules, or concepts in new situations, or activities such as translating, constructing, illustrating, building, or implementing.
Analyzing	Breaking materials or concepts into parts, determining how the parts relate to one another or to an overall structure or purpose. Mental actions include differentiating, organizing, and attributing, as well as being able to distinguish between the components.
Evaluating	Making judgments based on criteria and standards through checking and critiquing.
Creating	Putting elements together to form a coherent or functional whole; reorganizing elements into a new pattern or structure, as well as developing, designing, and generating.

Sources: Anderson and Krathwohl 2001; Wilson 2016.

nature, the taxonomy or modifications of it have been used in many countries throughout the world (Lewy and Báthory 1994). Revisions were designed to make the classification more relevant to assessment and teaching practice. It has also been widely used by testing and examination bodies to clarify curriculum objectives and test items. A major feature of revisions has been the use of outcome-oriented language and behavioral objectives. Nouns have been changed to verbs. For example, “knowledge” has become “remember” and “analysis” has become “analyze” (see figure 4.1).

Examinations in which students are required to answer using extended essays may not contain a sufficient number of questions to allow the same level of taxonomic detail as is possible in a multiple-choice test. Examination setters, however, should still try to ensure that important sections of the syllabus are sampled. The extent to which they actually do this is often open to question. In many cases, the approaches adopted fall short of what can be regarded as satisfactory, involving repetition of identical or similar questions from year to year, a focus on textbook knowledge rather than on competencies and core concepts, and overuse of multiple-choice items (Dundar et al. 2014).

FIGURE 4.1
Hierarchy of Levels of Cognitive Processes: Bloom’s Taxonomy and Revised Edition



Rather than constructing a new examination from scratch every time one is required (usually once or twice a year), some examination systems construct item banks from which items or tasks are selected for an examination. To constitute an item bank (rather than just an item pool), the items should have certain characteristics. They should be associated with a clear definition of outcomes, should be organized and catalogued (content area and expected cognitive process), and should provide calibration data (indexes of difficulty and discrimination) or slope parameters and location or threshold parameters used in item response theory analyses (Cartwright 2015) on measurement characteristics. Item banks provide a ready source of items with known statistical and content characteristics and can facilitate monitoring standards over time. A study of the reasons behind the decline in the United States' National Assessment of Educational Progress scores between 1984 and 1986 showed that small, almost innocuous changes, such as changing the order of the items or having students fill in ovals rather than circling letters, affect scores (Beaton and Zwick 1990). What is less clear is that the assumption of item response theory about a single latent (underlying) ability applying to all items is the appropriate one for student examination behavior. Furthermore, the advanced technical skills required to construct and maintain banks and select items with specific psychometric characteristics may be beyond the resources of some examination systems.

THE FORM OF EXAMINEE RESPONSE

Four major types of examination items or tasks and associated responses can be identified: extended essays, multiple-choice selection-type items, short written supply-type items, and performance tasks.

Extended Essays

There is a strong tradition, going back to the Chinese imperial examination system (see chapter 3), of requiring examinees to write extended essays in examinations. The format is used not just to assess candidates' writing skills (for example, how they organize material),

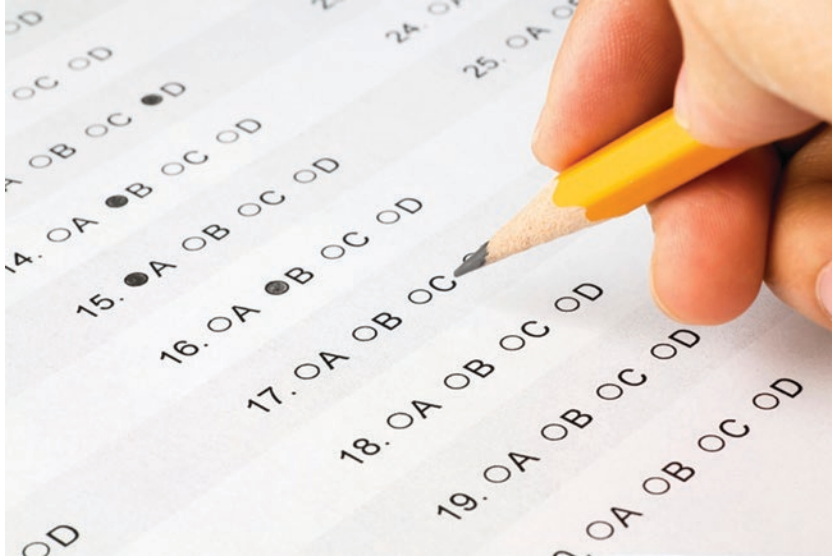
but also to assess their knowledge in a range of curriculum areas. An advantage of the essay question is that it may require a response from the examinee that has qualities of novelty and complexity. A disadvantage is that judges of the quality of a response may differ in their appraisals. To help guide such judgments, rubrics or specific marking guidelines need to be developed, typically by a group of experts, and defined by the group's knowledge of the content domain and their experience as educators. Rubrics provide guidance for scorers by identifying key elements required in a response, to which a precise number of marks is assigned. A further disadvantage of the essay question is the relatively small number of essays that can be accommodated in an examination, with the result that the domain to be assessed is not adequately represented. A final disadvantage is the high cost of scoring.

Multiple-Choice Selection-Type Items

Multiple-choice selection-type items were developed early in the twentieth century largely to address the problems of the unreliability and cost associated with essay-type items (see chapter 3). In a multiple-choice item, a candidate responds to a question or incomplete statement (in the stem), by selecting a response from among a number of alternatives (usually four or five). A major reason for the popularity of this test item is its “objectivity” (there is only one predetermined correct response), the simplicity of responding on an answer sheet that is processed on a scanner (see photo 4.1), and the ease with which items can be scored, including through electronic means.

Short Written Supply-Type Items

Rather than having a candidate select a response as in the multiple-choice item, in the short written supply-type item the candidate is required to respond by supplying a word, numbers, or other information. These items are frequently used in conjunction with multiple-choice items. They can provide greater insight into candidates' thinking processes (for example, the steps in solving a mathematical problem), but subjectivity can be a problem in scoring, as in the case of extended essays.

PHOTO 4.1**Candidate Responding to Four-Option Multiple-Choice Question on a Machine-Scannable Answer Sheet**

Photograph by Alberto G./Flickr (license: CC by 2.0 IGO). <https://www.flickr.com/photos/albertogp123/5843577306/>.

Performance Assessments

Performance assessments often “seek to emulate the context or conditions in which the intended knowledge or skills are actually applied” (Joint Committee AERA, APA, and NCME 2014, 77). Requiring candidates to apply their problem-solving skills in real-life situations provides a direct assessment of student achievement, measuring important learning outcomes that cannot be measured using other assessment formats. Some form of performance assessment is a feature of many public examination systems. It can be very focused, as in the case of the oral assessment of a candidate’s competence in a second language or the assessment of the construction of an object in woodwork. Performance assessment can be broader, as is the case in public examinations in England and Wales, which require teachers to assess what is called “coursework.” The International Baccalaureate also requires internal course work

assessment in its examination subjects. Coursework may take many forms: a project (for example, a historical investigation, a geography field trip, a mathematics task), a portfolio of work, or a performance in drama or music (Gipps and Stobart 2003).

While performance assessment has many attractive features, its use in large-scale assessments such as public examinations is problematic in a number of ways. First, it is time consuming; it has been estimated that between 10 and 23 tasks would be required to obtain an acceptable level of reliability (Shavelson, Baxter, and Gao 1993). Second, scoring is complex, requiring multidimensional information. For example, in the United Kingdom regulations for biology in the General Certificate of Secondary Education require that candidates be marked for, among other things, demonstrated competence in developing and testing hypotheses; managing risks when carrying out practical work; collecting, processing, and interpreting primary and secondary data; and drawing evidence-based conclusions (Ofqual 2015). It can be difficult to summarize in a meaningful way the information that could be incorporated into a candidate's results in a public examination. Third, it may be difficult to distinguish between the contribution of a candidate and that of adults who may have provided assistance. In England and Wales, these issues led to a retrenchment toward more traditional forms of assessment (Cresswell 1995). Fourth, schools may suffer when their teachers are assigned during the school year to other schools to administer oral and practical examinations.

Differences between Item Types

The limited research evidence available on differences between item types and the possible consequences of differences suggests the following conclusions (Martinez 1999). First, item types differ in their typical cognitive demands and in the range of cognitions they are designed to assess. Multiple-choice items, because they tap memory and convergent thinking, are frequently identified as failing to elicit the full spectrum of cognitive activity valued by educators. Constructed response items, on the other hand, can range widely in the knowledge, structures, processing strategies, and self-regulatory

functions they demand. They can require simple recall in short supply-type items, or in the case of extended essays, responses can range from ones that reproduce memorized essays to ones that reflect complex cognitive processes, including the ability to organize, synthesize, and evaluate knowledge.

Second, item types differ in their susceptibility to contamination by personal characteristics of candidates. Multiple-choice items have been considered to be particularly amenable (especially when items are poorly constructed) to candidate proficiency in using specific response strategies (for example, a strategy in which options are narrowed by eliminating the most implausible). Examinees that suffer from test anxiety are more likely to be adversely affected when faced with the task of generating a response than when required to select a response option. Females tend to perform better on essay-type items, while males tend to perform better on multiple-choice items (see chapter 10).

Third, item types differ in their potential effects on the educational processes and outcomes involved in the period leading up to the examination. In preparing for an examination, teachers and students are likely to have in mind the format of the examination. Richer outcomes tend to be associated with constructed response items than with multiple-choice items. For example, students who expected essay-type examinations, compared to ones preparing for a multiple-choice examination, paid greater attention to the structure of the curriculum and to strengthening their grasp of the subject matter at a more global level (Martinez 1999). The use of performance assessment in examinations is particularly lauded for its potential to affect learning and teaching, to shape instructional practice, and to encourage strategies that foster reasoning, problem solving, and communication (Frederiksen 1984; Gipps and Stobart 2003; Lane and Stone 2006).

The role of response formats in examinations and, in particular, how formats might affect anticipatory teaching and learning should be important considerations in the construction of examinations. Such considerations would probably lead to the conclusion that a variety of formats is desirable if the purpose of the examination is to elicit a wide spectrum of valued cognitions.

QUESTION PAPER PRODUCTION AND PRINTING

Writing examination questions or items is a professional task. Ensuring that production quality standards are observed, that security is maintained, and that deadlines are met are largely administrative tasks. These tasks involve facilitating production and monitoring (for example, arranging meetings and arranging the flow of materials among typesetters, artists, printers, and proofreaders).

The following steps are normally required to produce good-quality question papers.

1. Question papers and marking schemes should be written in conjunction with each other. They should include the answers that examiners are looking for and indicate how marks should be awarded if answers are only partially correct. A marking scheme should specify the knowledge and skills to be assessed and the values to be assigned to specific aspects of performance. Marking schemes should be revised after candidates have taken the examination. Papers may be written by one person (chief examiner) or, more usually, by a team of writers, or they may be made up of items taken from an item bank.
2. A professional panel should moderate the draft examination papers and marking schemes, a process that involves reviewing and revising papers to check for factual accuracy, clarity of expression and appropriateness of the papers and supporting drawings or artwork, and the draft marking scheme.
3. Moderated manuscripts may be prepared or set up in-house (for example, in a confidential materials unit) or by a security typesetter.
4. Artwork may be produced in house (for example, by desktop publishing) or by contract artists.
5. Proofs need to be checked by a chief examiner or specialists, amended, and checked again.
6. The question paper should be “passed for printing” by the chief examiner.
7. A parallel examination may be produced in case the original examination has to be replaced at short notice, for example, for security reasons.

Table 4.3 lists a number of options available for printing examination papers, together with their advantages and disadvantages.

Leakage of question papers threatens the credibility of an examination (see chapter 11). However, it is often difficult to identify when

TABLE 4.3

Alternative Strategies for Printing

Strategies	Advantages	Disadvantages
In-house facility	Scheduling, quality control, security, purchasing controlled by examinations body.	High capital outlay and recurrent costs for maintenance. Need for trained staff and space for printing and storage.
Government printers	No investment in equipment. No recurrent costs for maintenance and staff. Advantageous rates may be available. High security option may be available (as for confidential government papers).	May be difficult to insist on high levels of services and, in particular, to control time of printing and delivery. Quality may be lower than in commercial sector. Costs may be higher. Security may be difficult to control.
Local commercial printers	No investment in equipment. No recurrent costs for maintenance and staff. Quality and levels of service may be high, especially in competitive markets. Relatively low risk of poor service as printer will be bound to meet contractual agreements.	Security may be difficult to control. Direct costs may be higher than in-house prices.
International specialist security printers	No investment in equipment and low recurrent costs for maintenance and staff. High levels of security, quality, service. Additional services (for example, packing) may be available.	High cost of printing. High cost of freight. Payment may require hard currency. No support of local industry.

Source: World Bank 2001.

TABLE 4.4
Measures to Improve Security of Examination Papers

Measure	Comments
Elimination of proofreading of master copy	This may reduce leakage but will probably increase the number of errors, particularly in technical subjects.
"Just-in-time" printing	Reducing the time between printing and the date of the examination (to perhaps only a few days) reduces the opportunity for leakage.
Supervision of print shops	Where external printing sources are used, examination authority officers will be present in the print facility to ensure that extra copies are not run off and that all printing source materials including files and waste materials are dealt with appropriately. Mobile and iPhones and other photographic equipment are not allowed in the printing facility.
Isolation of printers	Printers may not be allowed leave the print shop (day or night) until the printing of papers is complete.
Packaging	Employing trusted employees who do not have close relatives taking the examination and who carefully monitor the quantity and content of examination papers in packages.

Source: World Bank 2001.

exactly an examination paper was leaked from the system, for instance, when papers were being prepared, being transported, or stored. On the assumption, without a great deal of evidence, that the potential for leakage is greatest during the printing phase, examination bodies often go to great lengths to improve security when papers are being printed and packaged (see table 4.4). Online examinations, discussed later, while eliminating many of the risks associated with printed examination papers, can also be leaked or hacked prior to the examination.

MARKING OF ANSWER SCRIPTS

There are two major approaches to scoring essays: holistic and analytic (Meadows and Billington 2005). In holistic scoring, raters make a single general judgment about the quality of a response as a whole

and assign one score, using a scoring rubric with criteria that identify benchmark scores associated with grades. Experienced markers may be asked to select answers that exemplify grades. In analytic scoring, raters evaluate candidates' responses according to a number of features (for example, content, organization, focus, mechanics, and ideas) considered essential to a good answer, and assign a score indicating a level of quality to each one. In some analytic scoring methods, domains are weighted; that is, ones considered to be more pertinent to the construct being measured will contribute more to the overall score (Lane and Stone 2006). Tisi et al. (2013) cite sources that claim that holistic marking is less valid and possibly less reliable than analytic marking.

After candidates have taken an essay-type examination, a sample of scripts marked by experienced examiners and distributed across key grade boundary ranges may be reviewed by senior examiners. Marking schemes will be finalized on this basis.

The setting of grade thresholds is a key step in this process (Cambridge Assessment International Education 2019). A grade threshold is the minimum number of marks that a candidate needs to be awarded a particular grade in a paper or a subject. Normally, the aim is to set each threshold every year at a level that matches the level in the previous year. A threshold may be lowered if questions seem to have been more "difficult" than last year or raised if questions seem to have been "easier." Grades are not predetermined or awarded on a quota system.

Levels-Based Marking Schemes

Levels-based marking schemes are commonly used in marking extended response items. Because such items have scope for multiple valid approaches, point-based marking or the provision of exemplars is not appropriate. The examiner has to make an initial assessment of a candidate's response, classify it into a single defined level, and assign a mark (Pinot de Moira 2013). Table 4.5 provides an example of levels (and associated marks) for a General Certificate of Education Citizenship Studies unit. It should be noted that levels-based marking schemes can vary in several ways: in the number of levels in the

TABLE 4.5

Level Description of Objectives, Knowledge, and Understanding
GCE Citizenship Studies Unit 1 Generic Mark Scheme for Items 1–5, Summer 2011

Level	Assessment Objective AO1
3	(4–5 marks) Answers demonstrate a range of citizenship understanding and an accurate understanding of relevant citizenship concepts and theories. A range of examples is used to relate knowledge and understanding to citizenship issues.
2	(2–3 marks) Answers are characterized by a good level of citizenship knowledge and an understanding of relevant citizenship concepts and theories. Examples are used to relate knowledge and understanding to citizenship issues.
1	(1 mark) Answers are characterized by limited citizenship knowledge and limited understanding of relevant concepts and theories. Candidates may make a limited attempt to use examples to relate knowledge and understanding to citizen issues, or no examples may be present.
	(0 marks) No relevant response.

Source: Pinot de Moira 2013.

Note: GCE = General Certificate of Education.

scheme, in the number of marks within a level, in the distribution of marks between levels, and in the inclusion (or omission) of indicative content within levels.

In addition to bearing responsibility for preparing and printing examination papers, examination agencies tend to devote considerable resources to recruiting and training markers, packaging and distributing papers, moderating marks, collecting and storing marked papers, checking and issuing results, and dealing with appeals. Factors that contribute to the success of these activities include the existence of carefully designed administrative procedures (South Africa, Directorate, Assessment and Examinations 2007), attention to detail, and access to adequate secure distribution and storage facilities (see photo 4.2). Marking procedures used by examination agencies include centralized and home-based marking of paper scripts and home-based on-screen or online marking.

PHOTO 4.2

Ireland: Sorting Packets of Examination Papers Prior to Issuing Them to Markers



Photograph © Ireland State Examination Commission/Alan Betson/Irish Times. Reproduced with permission from the Irish Times and the Ireland State Examination Commission; further permission required for reuse.

Centralized Marking

In centralized marking, examination scripts are returned to a central location or locations and marked over a period of weeks under the direct supervision of senior examiners or markers. The process tends to be relatively expensive, involving fees and travel, subsistence, and accommodation costs. It offers the benefits, however, of a useful form of in-service teacher education, as it highlights the essential elements of the curriculum being assessed, standards expected, and marking procedures. Furthermore, it helps in establishing networks of teachers in different subject areas.

Marking Paper Scripts at Home

In home-based marking arrangements, practicing and retired teachers and frequently graduates with relevant subject matter expertise receive packages of scripts from the examination board for scoring in their own homes. Novice markers in particular receive training in marking processes. Senior examiners monitor samples of marked scripts and provide regular feedback to markers. While home marking allows markers flexibility in their use of time and is less expensive than centralized marking, it is not subject to the same degree of supervision as other approaches. It also requires markers to carry out additional administrative work in the form of packing and mailing scripts to supervisors and to the examination agency.

On-Screen Marking

In on-screen marking, examinees' scripts are scanned into digital format and sent to examiners. A variety of advantages have been associated with this procedure. First, it facilitates item-level marking, which reduces errors due to the idiosyncrasies of individual markers. Second, it allows for flexibility in marking allocation. Markers can be sent items or scripts when they are ready, with items that are difficult to mark sent to more experienced markers. Third, on-screen marking allows for continuous monitoring of an individual marker's performance. Problems can be flagged at an early stage and examiners whose marking standards drift can be identified. Fourth, on-screen marking removes the

economic and logistical burdens of sending paper scripts to and from examiners or of bringing markers to a central location. Fifth, it increases the speed of marking. Finally, because marks are directly inputted to the system, errors associated with incorrect addition or transcription of marks are eliminated (Tisi et al. 2013).

USE OF INFORMATION TECHNOLOGIES

Recent years have witnessed a substantial increase in the use of information technologies (IT) in examination systems. IT can be used to manage one or more individual components of the examination cycle, such as candidate registration, candidate responses, marking of candidates' responses (described earlier), and issuing of results, or can offer a complete national-level e-examination system. Many examination authorities are investing heavily in technology. The Kenyan National Examination Commission, for example, is committed to carrying out an IT needs assessment, implementing an IT upgrading program, installing data backup and recovery programs, exploring the use of the cloud, and using IT to enhance security (KNEC 2015). In 2018 the Arab Republic of Egypt launched a large-scale initiative that included the development of a new computer-based examination system involving students in grades 10, 11, and 12, the results of which are to be aggregated to produce a final secondary school graduation mark.

Modern IT can support a complete national-level e-examination system. In Georgia, for example, a country that only recently introduced public examinations, school graduation examinations for about 50,000 students in eight subjects are delivered through a computer-adaptive testing (CAT) system developed by an agency of the Ministry of Education National Examinations Center (Bakker 2014). CAT uses a process of systematically selecting questions for each individual candidate, maximizing the precision of the examination by using information based on the candidate's responses to previous questions. A correct response will lead to a candidate being presented with a more difficult question, while an incorrect one leads to an easier question. The test normally starts with a question of average difficulty. This type of assessment supports an accurate measurement of student

achievement using fewer questions than one that requires all students to respond to all items. However, CAT requires considerably more items than a conventional multiple-choice test and a large amount of pilot testing (to determine item parameters). As responses are scored in the sequence in which they appear on the screen, students are not normally able to skip items and return to them later, thus denying them a feature of other forms of assessment. Issues may arise about the extent to which the selected items represent an adequate coverage of the syllabus.

Some educational systems include automated scoring of writing tests both for national assessments and for high-stakes testing. In Australia, for instance, a computerized system is trained to use the same marking criteria as used by human markers and to apply these to narrative and persuasive writing tasks submitted on the computer. Criteria include ability to engage the reader, organization of text features, use of ideas, development of character and setting, language precision, spelling accuracy, and correct use of grammar, punctuation, and paragraphing (ACARA 2016). Students' responses that cannot be scored by the computer are redirected to a human marker. The system, introduced on a phased basis, requires students to be able to use computers to complete writing tasks, and schools to have appropriate hardware and access to reliable internet service. Programs have been developed to grade essays written in different languages (Pearson Education 2010). Issues remain about the ability of automated scoring systems to assess aspects of writing such as critical thinking, creativity, rhetorical style, and the factual accuracy of claims (Zhang 2013).

Several advantages have been attributed to the use of IT in examinations: increased speed in processing, increased security, increased accuracy, increased opportunities for analysis of performance, and reduced costs (due to savings on storage, printing, and payment of examiners), especially when using cloud technologies, which are ideal when there are peaks in data generation and processing.

Disadvantages associated with IT are the dependence on high-tech machines and special stationery (often imported), carrying with it the risk that the system may fail at a crucial point; dependence in many systems on multiple-choice items (although CAT can be applied to other formats), which may have undesirable backwash effects on

teaching and learning; the distance it creates between teachers and the assessment process; and the fact that it results in loss of payment to examiners that often supplements low teacher salaries. A final disadvantage of IT in an examination is that candidates need to be familiar with a range of procedures, for example, filling in optical mark reader forms, using computer keyboards, and following instructions on the computer. If there is variation in the extent to which candidates possess these competencies, they could become a source of construct-irrelevant variance in candidates' performance; in other words, performance on the examination could be affected by factors or skills that are irrelevant to the construct the examination is intended to assess (see chapter 7).

Experience with computer-based large-scale assessment programs in the United States has been mixed. In some states (for example, California) satisfaction was expressed with computerized systems, but computerized tests were scrapped in Tennessee in 2016 and traditional tests reinstated. In Texas, answers to more than 14,000 tests were accidentally erased, while in Alaska, the cutting of a fiber optic cable caused severe disruption in testing. Additional difficulties have included limited broadband, especially in rural schools; old computers; and inconsistent technical support (Brown 2016). Other states have experienced problems with online tests, including cyberattacks, log-in difficulties, and technical errors (Herold 2016). Many of these issues are likely to be resolved with improvements in technology, but the problem of computer hacking of data examination databases and documents will remain a concern.

CONCLUSION

The type of activities described in this chapter should serve to underline the complexity of the administration of a public examination system. This, in turn, should point to the need for adequate resources and personnel with the diversity of skills required to carry out a wide range of tasks. These include preparing examination tasks; organizing the administration of those tasks; scoring examinees' responses; analyzing results; and providing feedback to examinees, schools, and the

wider society. Some of these tasks are basically technical and can be addressed by the provision of adequate funding and training. Others may be more challenging because of their philosophical and political implications. Stakeholders should have opportunities to express their views about the nature of achievement, how it should be assessed, and the uses that are made of information about examinees' performance.

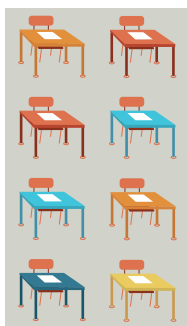
REFERENCES

- ACARA (Australian Curriculum, Assessment and Reporting Authority). 2016. *NAPLAN 2016 Narrative Writing Marking Guide*. Sydney: ACARA.
- Airasian, P. W. 1994. "The Impact of the Taxonomy on Testing and Evaluation." In *Bloom's Taxonomy: A Forty-Year Retrospective—Ninety-Third Yearbook of the National Society for the Study of Education*, part 2, edited by L. W. Anderson and L. A. Sosniak, 82–102. Chicago: National Society for the Study of Education.
- Anderson, L. W., and D. R. Krathwohl, eds. 2001. *A Taxonomy of Learning, Teaching and Assessment: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Bakker, S. 2014. *The Introduction of Large-Scale Computer Adaptive Testing in Georgia: Political Context, Capacity Building, Implementation, and Lessons Learned*. Washington, DC: World Bank.
- Beaton, A. E., and R. Zwick. 1990. "The Effects of Changes in the National Assessment: Disentangling the NAEP 1985–86 Reading Anomaly." Princeton, NJ: Educational Testing Service. Cited in G. F. Madaus, "Testing as a Social Technology" (Bosi Lecture in Education and Public Policy, Boston College, Chestnut Hill, MA, 1990).
- Bloom, B. S., ed. 1956. *Taxonomy of Educational Objectives: Handbook 1—Cognitive Domain*. London: Longmans Green.
- Britton, E., S. Hawkins, and M. Ganda. 1996. "Comparing Examination Systems." In *Examining the Examinations*, edited by E. D. Britton and S. A. Raizen, 201–17. Dordrecht, Netherlands: Kluwer Academic.
- Brown, E. 2016. "Technical Glitches Plague Computer-Based Standardized Tests Nationwide." *Washington Post*, April 14.
- Cambridge Assessment International Education. 2019. "March 2019 Grade Threshold Tables." <https://www.cambridgeinternational.org/programmes-and-qualifications/cambridge-upper-secondary/cambridge-igcse/grade-threshold-tables/march-2019/>.

- Cartwright, F. 2015. "Item and Test Analyses." In G. Shiel and F. Cartwright, *Analyzing Data from a National Assessment of Educational Achievement*, 123–266. Washington, DC: World Bank.
- Cresswell, M. J. 1995. "Technical and Educational Implications of Using Public Examinations for Selection to Higher Education." In *Admission to Higher Education: Issues and Practice*, edited by T. Kellaghan, 38–48. Princeton, NJ: International Association for Educational Assessment and Dublin: Educational Research Centre.
- Dundar, H., T. Béteille, M. Riboud, and A. Deolalikar. 2014. *Student Learning in South Asia: Challenges, Opportunities, and Policy Priorities*. Washington, DC: World Bank.
- Frederiksen, N. 1984. "The Real Test Bias: Influence of Testing on Teaching and Learning." *American Psychologist* 39: 193–202.
- Gipps, C., and G. Stobart. 2003. "Alternative Assessment." In *International Handbook of Educational Evaluation: Part 2—Practice*, edited by T. Kellaghan and D. L. Stufflebeam, 549–75. Dordrecht, Netherlands: Kluwer Academic.
- Herold, B. 2016. "Technology in Education: An Overview." *Education Week*, February 5. <https://www.edweek.org/ew/issues/technology-in-education/index.html>.
- India, CBSE (Central Board of Secondary Education). 2017. "Biology." <http://cbseacademic.in/curriculum.html>.
- Joint Committee AERA (American Educational Research Association), APA (American Psychological Association), and NCME (National Council on Measurement in Education). 2014. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- KNEC (Kenya National Examinations Council). 2015. "Strategic Plan, 2015–2020. Popular Version." KNEC, Nairobi.
- Lane, S., and C. A. Stone. 2006. "Performance Assessment." In *Educational Measurement*, 4th ed., edited by R. L. Brennan, 387–431. Westport, CT: American Council on Education and Praeger.
- Lewy, A., and Z. Báthory. 1994. "The Taxonomy of Educational Objectives in Continental Europe, the Mediterranean, and the Middle East." In *Bloom's Taxonomy: A Forty-Year Retrospective—Ninety-Third Yearbook of the National Society for the Study of Education*, part 2, edited by L. W. Anderson and L. A. Sosniak, 146–63. Chicago: National Society for the Study of Education.
- Martinez, M. E. 1999. "Cognition and the Question of Test Item Format." *Educational Psychologist* 34: 207–18.

- Meadows, M., and L. Billington. 2005. "A Review of the Literature on Marking Reliability." National Assessment Agency, Manchester, UK. https://cerp.aqa.org.uk/sites/default/files/pdf_upload/CERP_RP_MM_01052005.pdf.
- Ofqual (United Kingdom, Office of Qualifications and Examinations Regulation). 2015. "GCSE (Graded A* to G) Controlled Assessment Regulations." Ofqual, Coventry, UK.
- Pearson Education. 2010. "Intelligent Essay Assessor (IEA) Fact Sheet." <https://images.pearsonassessments.com/images/assets/kt/download/IEA-FactSheet-20100401.pdf>.
- Pinot de Moira, A. 2013. "Features of a Levels-Based Mark Scheme and Their Effect on Marker Reliability." Technical Report, Centre for Education Research and Policy, Manchester, UK.
- Shavelson, R. J., G. P. Baxter, and X. Gao. 1993. "Sampling Variability of Performance Assessments." *Journal of Educational Measurement* 30: 215–32.
- South Africa, Directorate, Assessment and Examinations. 2007. *Provincial Curriculum Guidelines: Script Control Manual*. Zwelitsha: Province of Eastern Cape Department of Education, Republic of South Africa. <http://www.ecexams.co.za/0.9PCG%209.pdf>.
- Tisi, J., G. Whitehouse, S. Maughan, and N. Burdett. 2013. "A Review of the Literature on Marking Reliability Research." Ofqual/13/5285. National Foundation for Educational Research, Slough, Berkshire, UK.
- Tyler, R. W. 1934. *Constructing Achievement Tests*. Columbus, OH: Bureau of Educational Research, Ohio State University.
- Wilson, L. O. 2016. "Anderson and Krathwohl—Bloom's Taxonomy Revised: Understanding the New Version of Bloom's Taxonomy." <http://thesecondprinciple.com/teaching-essentials/beyond-bloom-cognitive-taxonomy-revised/>.
- World Bank. 2001. "Toolkit on Public Examination Systems in Developing Countries." World Bank, Washington, DC.
- Zhang, M. 2013. "Contrasting Automated and Human Scoring of Essays." *R & D Connections* 21. Princeton, NJ: Educational Testing Service.

CHAPTER 5



ADVANTAGES AND DISADVANTAGES OF PUBLIC EXAMINATIONS

INTRODUCTION

As noted in chapter 1, the use of assessment is widely proposed as a lever of reform—a means of improving educational quality. Assessment is ascribed an active role, one in which it is consciously used in the education process to change teacher behavior and classroom instruction, which in turn is expected to raise the standard of student learning. While the argument is applied to many kinds of assessment, the role of teachers' assessment practices in the classroom is critically important, and is an integral part of the teaching learning process. However, probably largely for logistic and cost considerations, most resources for reform have been invested, not in directly changing teachers' classroom practices, but in forms of assessment controlled by an agent outside the school that are expected to affect teachers' behavior. One focus of such reform is the public examination system.

Public examinations derive their importance largely from the fact that examination performance forms the basis of important decisions about the educational and vocational futures of students. In performing their discriminatory or ranking function, examinations have

acquired a legitimacy based on the view that the qualifications they confer provide a fair indication of achievement and that the distribution of benefits on the basis of such qualifications is equitable. As a consequence, achievement comes to be defined by the examinations, since it is examination performance, and not any other kind of achievement, that is rewarded. Thus examinations become obtrusive measures of student achievement. Little (1982) concludes that since examinations represent the ultimate goal of the educational career, they define the important aspects of a school curriculum and dictate to a large degree the quality of the school experience for both teacher and student alike.

ADVANTAGES OF PUBLIC EXAMINATIONS

A number of specific advantages have been attributed to examinations (see, for example, Heyneman 1987; Hill 2013; Madaus and Kellaghan 1991; Tyumeneva 2013). They are considered to be a relatively objective and impartial means of distributing educational benefits, providing an assessment procedure that is unaffected by personal relationships between teachers and students. They provide incentives for students to study and help focus teachers' efforts on key aspects of the curriculum, an important consideration if teachers are unclear about the contents of national curricula. They provide a measure of educational quality and instill public confidence in the system. The level of anonymity associated with the examination process may have an equity-enhancing function as in the process it also contributes to the creation throughout the education system of national homogeneity in standards and practice. These factors, together with their wide acceptance in the community, help explain their longevity.

At a broader level, examinations reduce the effects of patronage and have been credited with opening higher education and a range of occupations to a wide population of students. Alternative methods of selection based on family status, political influence, parental occupation, or random selection are generally regarded as unacceptable.

By introducing “points” systems in conjunction with examinations, some countries have greatly increased the effective use of examination results. In Oman, for example, a highly computerized selection system requires students to apply to one body, the Higher Education Admissions Center, rather than to more than 50 public and private higher educational institutions (Oman 2019). Decisions are based on students’ school leaving examination marks, subject requirements, and the number of places available in higher education institutions. The system has contributed to a speedier selection process, reduced student travel to sit entrance examinations in several institutions, increased transparency, and resulted in fewer instances of malpractice. The system is modeled on one used by the Irish Central Applications Office, which oversees undergraduate applications to colleges and universities, and since 2019 has informed applicants of the outcome three days after the initial release of examination results to the students.

Performance on a high-stakes examination can help students from disadvantaged backgrounds, who might have otherwise been forced to drop out, advance to the next highest level of the education system. Scholarships have been available in many countries to provide financial support to high-scoring students from such backgrounds (Little, Aturupane, and Shojo 2013).¹ Scholarship schemes have been augmented with additional means of providing financial and other forms of support to students in need, including government and state grants, support from nonprofit groups, extended time to complete programs of study, and internships.

How examinations might contribute to raising academic standards has been a matter of debate. Supporters of their use argue that using examinations as guides for teaching leads to an alignment of instructional and learning processes with examinations. At their simplest, the examinations should provide a sense of purpose for teachers’ efforts and tangible incentives for students. They should also direct teachers’ and students’ attention to important topics and skills (Adnan and Mahmood 2014; Eisemon 1990; Heyneman and Ransom 1992; Ross and Maehlck 1990; Somerset 1996; Spratt 2005; Tyumeneva 2013).

In anticipation of problems with this position, it is argued that if the alignment is to be beneficial, the scope and quality of the

examination should be satisfactory. The examination should, for example, contain questions or prompts that require higher-order processes instead of verbatim recall, present a clear concept of desired goals, and contain a manageable number of assessment foci around which teachers can organize their instruction (Popham 1983). The impact will be strengthened if feedback is provided to schools on how their students have performed in specific subject areas and even in individual sections of an examination. Finally, as an incentive, high stakes should be associated with performance on the examinations. That is, how students perform should have important consequences for students, teachers, or schools.

If external curriculum-based examinations, in addition to serving their traditional functions of certification and selection, improve human capital (represented by the cognitive skills assessed in examinations), this would be an important consideration as such improvement would have the potential to substantially improve the long-term economic well-being of a country (Hanushek and Woessmann 2011).

The availability of data from international studies in which representative samples of students from a number of educational systems sit for the same achievement tests provides a possible source of empirical evidence regarding the contribution to student achievement of factors, including public examinations, that vary from country to country (Greaney and Kellaghan 2008). Several analyses (at the country and individual student level) of data from international studies have been carried out in search of an answer to the question, Do students in education systems that have end-of-secondary-school external examinations perform better in international studies of student achievement (reading, mathematics, science) than students in education systems that do not have such examinations? The studies that provided the data for analysis were the International Assessment of Educational Progress (IAEP), the Third International Mathematics and Science Study (TIMSS), the International Association for the Evaluation of Educational Achievement Reading Literacy Study, and the Organisation for Economic Co-operation and Development Programme for International Student Assessment (PISA).

The findings of analyses do indeed support the view that students in education systems with curriculum-based external examinations

have higher levels of achievement than students in systems without such examinations (Bishop 1995, 1998a, 1998b, 1999; Fuchs and Woessmann 2004; Hanushek and Woessmann 2011; Woessmann 2000). For example, after controlling for resources and family background in an analysis of TIMSS data, having external examinations was significantly related to student achievement in mathematics and science. The findings were not entirely consistent across studies, however. There was greater consistency in the case of mathematics than of reading or science. Not surprisingly, the findings indicated that a wide range of factors were related to student achievement: student characteristics, family background, home inputs, resources and teachers, and institutional structures (Fuchs and Woessmann 2004), and it was recognized that examinations were not the most important determinant of achievement (Bishop 1998b).

The precise nature of the contribution of examinations to student achievement has been a topic of debate. For instance, some have argued that curriculum-based external examinations contribute to the national achievement levels of students, as they provide performance information that can be used to hold schools and students accountable (Hanushek and Woessmann 2011). These examinations have also been associated with higher minimum standards of entry into teaching and a greater likelihood of hiring teachers who have majored in the subject they will teach, and additional hours of instruction in examination subjects (Bishop 1999). They also put pressure on teachers to improve their teaching, as colleagues will be aware of the examination results (Bishop 2005). It has also been asserted that because examinations signal levels of achievement in specific subjects to users (colleges and employers), and bring increased extrinsic rewards for learning, they stimulate changes in priorities and behavior, including studying harder, that result in higher achievement (Bishop 1995).

A consideration of the role of motivation in learning, in particular the implications of using extrinsic rather than intrinsic motivation to promote learning (see chapter 2), would suggest that the role proposed for external examinations in raising educational standards may underestimate the complexity of the psychological processes involved, as well as fail to anticipate the range of consequences that may ensue (Kellaghan, Madaus, and Raczek 1996).

Finally, the assumption that all external examinations bear a similar relationship to accountability is open to question. In a case study of three education systems with long-established public examinations at the end of secondary education, only one (the Netherlands) used examination results systematically to hold schools accountable. In the other two systems (Finland, Ireland), publication of examination results was rejected, and thus, the results could not be used to hold schools accountable (Klein 2010), although of course they could have had an accountability role at the level of individuals and families.²

DISADVANTAGES OF PUBLIC EXAMINATIONS

Although there can be little doubt that examinations to which high stakes are attached affect teaching and learning, the effects are not always as intended; neither are they always positive.³ The negative effects of such examinations have been a recurring theme in their history, from the days of examinations in Imperial China, through the period of their introduction in Europe in the eighteenth and nineteenth centuries, to the present day. For example, while the introduction of the Tripos in Cambridge University might have been considered a cause for celebration by some (see chapter 3), some mathematicians concluded that training for the examination stifled the creative thinking of young mathematicians for a century (Martinez 1999).

Underrepresentation of the Domain

The most obvious criticism that can be leveled at high-stakes examinations is that a pencil-and-paper test, usually carried out under artificial conditions in a limited time frame, cannot measure many of the skills that most curricula are designed to develop. Thus, performance on the examination will be unrepresentative of performance on the target domain of knowledge and skills and so cannot be accepted as evidence that the examinee has acquired a degree of competence in the objectives of a course as prescribed in a curriculum (see chapter 4).

Underrepresentation in an examination of the knowledge and skills specified in a curriculum (such as oral fluency) is associated with a

series of negative effects on teaching and learning. Some relate to written examinations; others relate to any kind of examination to which high stakes are attached.

Focus on Topics Examined

Students and teachers will focus their efforts on topics that are to be examined, often excluding both cognitive and noncognitive areas that are not relevant to the examination. They will tend to emphasize the scholastic skills tested in examinations (particularly ones involving literacy and numeracy), excluding or spending less time on worthwhile educational experiences, such as ones involved in social studies and activities that might foster citizenship, science, art, physical education, and music (see figure 5.1). Teachers will refer to past examination papers in deciding what areas of the curriculum to study (and not study) in a phenomenon known as “teaching to the test.”

FIGURE 5.1
Examinations' Impact on Subjects Studied



Source: Cartoon © 2019 David Horsey. All rights reserved. Distributed by Tribune Content Agency, LLC. Reproduced with permission from Tribune Content Agency; further permission required for reuse.

In effect, the content of previous examinations comes to define the implemented curriculum.

The focus on topics tested in examinations will also tend to inhibit the development of curriculum variety that may be required to meet local and student needs.

Focus on Achievement at a Low Taxonomic Level

The negative impact of the examination will be more pronounced if the items in an examination measure achievement at a low taxonomic level, involving the recall or recognition of factual information, rather than the ability to synthesize information or apply principles to new situations (for example, in making inferences, developing a logical sequence of steps to solve a problem, or arguing a case). In preparation for the examination, the focus will be on rote memorization and the development of algorithmic skills, while process, planning, and perseverance skills will be accorded little attention. That is the situation in most countries according to Black (1996, 20): examination questions can be answered by routine procedures often based on rote learning rather than by “thoughtful translation and application of principles and procedures.”

In an earlier study of the Irish Leaving Certificate Examination (Madaus and Macnamara 1970), the nature of the abilities students displayed in the examination were analyzed using Bloom’s (1956) *Taxonomy of Educational Objectives*. Findings indicated that examinations focused on the lower levels of the taxonomy (knowledge) rather than on the higher levels of analysis, synthesis, or evaluation. The findings of a study of examinations in eight African countries are similar (Little 1982). In that study, there was a high degree of emphasis on the achievement of cognitive skills, with little reference to concrete real-life situations. Furthermore, the majority of items measured knowledge and application. There was, however, considerable variation among subjects in the taxonomic levels examined. In the 1978 Zambian Junior Secondary School Leaving Certificate, for example, the majority of items in geography (60 percent), civics (93 percent), and religion (90 percent), tested knowledge, while in mathematics the taxonomic level was much higher (92 percent of items tested application). It was surprising, perhaps, that practical subjects such as

health science and agricultural science emphasized knowledge rather than comprehension or application.

Time does not seem to have altered the situation regarding the taxonomic level assessed in examinations. A World Bank review of education in South Asia concluded that public examinations typically focused on memorization and information recall rather than critical thinking and problem-solving skills (Dundar et al. 2014). In Bangladesh, more than 80 percent of total examination marks were awarded for recall of facts (Hossain 2009). More recently, a 2017 study noted that questions on higher-order skills were almost entirely lacking in papers from a number of examination boards in Pakistan and one in India, and less so in papers set by examination authorities in Uganda and Nigeria (Burdett 2017).

Focus on Examination-Taking Skills

Apart from effects on the scope and emphasis of the curriculum as implemented by teachers, high-stakes testing results in considerable effort and time being invested in the development of the kind of skills that help students do well in an examination. There are reports that teachers teach to past examination papers and use the same types of items in their own tests and even in their teaching. For example, when examinations use the multiple-choice format, normal teaching may be presented in the form of statements accompanied by a range of options from which students are required to choose (for example, “The Capital of South Africa is Cape Town, Johannesburg, Pretoria, Soweto”). The influence of examinations on teaching is not confined to examination grades but has been observed in grades much lower than the ones in which examinations are held (Kellaghan and Greaney 1992).

Impact on Reform

In situations where examinations play a dominant role in determining what goes on in classrooms, the influence of a high-stakes examination may limit the outcomes of educational reform efforts. In Jordan, for example, the impact of a large-scale reform program that included,

among other goals, the development of students' problem-solving and critical-thinking skills was limited by the emphasis on rote learning in the all-important secondary school leaving examination, the Tawjihi (UNESCO 2014).⁴

Focus on Students Likely to Succeed in the Examination

Schools may concentrate their efforts on students who are most likely to succeed in an examination, with the result that potential low scorers are excluded from the examination, are retained in grade, or drop out of school.⁵ Teachers may focus, in particular, on students who are close to the minimum passing standard in an attempt to raise the overall success rate.

Anxiety

Written examinations may not be suitable for all students. For some, the degree of stress they engender makes them unsuitable for assessing student achievement. Pressure caused by examinations has been associated with loss of sleep, skipping meals, tension within families, loss of confidence, use of drugs and alcohol, and self-harm. In the Republic of Korea, authorities have implemented a variety of steps to reduce the level of student stress (see box 5.1). Concern has been expressed in the United Kingdom over the number of students seeking counseling for problems arising from examination-related stress (Adams 2015). Examination-related suicides have been reported in many countries,

BOX 5.1

Republic of Korea: Public Initiatives to Reduce Student Stress

"Commonly on test day, the allocation of buses and subways are expanded to avoid traffic jams, students are also escorted by police officers with the siren on, and even aircraft take-offs or landings are forbidden during the listening test of the English section."

Source: Kwon, Lee, and Shin 2017, 66.

including China, the Arab Republic of Egypt, India, Korea, and the Russian Federation (see, for example, BBC 2019). Examinations also tend to contribute to high anxiety levels among parents. The popular media in many countries regularly feature scenes of anxious parents outside examination sites or praying for examination success in places of worship (see photo 5.1).

Malpractice

In trying to obtain high scores on an examination, students (and sometimes teachers and others) may resort to various forms of malpractice (see chapter 11). This is a serious issue, particularly in many emerging market economies, where bribing or intimidating examiners, purchasing copies of examination papers before the examination, impersonation, copying, and collusion between examinees, examination invigilators, and examiners are common.

PHOTO 5.1

China: Candidates' Mothers Burning Incense and Offering Prayers for Examination Success



Photograph © Sim Chi Yin/Magnum Photos. Reproduced with permission from Magnum Photos; further permission required for reuse.

Narrowing of Cognitive Processes

The results of experiments and field studies suggest that examinations or tests toward which high stakes or sanctions are attached affect students' learning goals and strategies, involvement in learning tasks, and attitudes toward learning, in particular attitudes toward improving their competence. These studies distinguish between *learning* (or *mastery*) *goals*, which reflect the concern of individuals to develop new skills and increase their competence, understanding, and mastery of something new, and *performance* (or *ego*) *goals*, which reflect individuals' concern to demonstrate their ability and to gain favorable judgments or avoid negative evaluations of their ability or competence. Different patterns of behavior, cognition, and affect have been found to be associated with pursuit of these two goal categories. For example, learning-goal students are more likely to apply self-regulating effective learning and problem-solving strategies, while performance-goal students make more use of strategies that are superficial or short term, such as rote memorization, rehearsing, routine drilling, and the accumulation of factual knowledge. Furthermore, students who pursue learning goals are more positive and exhibit a preference for challenging work and risk taking, while students with performance goals tend to avoid challenging tasks and risk taking, especially if their self-concept of ability is low (Ames 1992; Grolnick and Ryan 1987; Kellaghan, Madaus, and Raczek 1996).

Most people would probably agree that it would be preferable if schools promoted the development of learning goals rather than performance goals. This view is very much in line with current standards-based reform efforts, particularly in the United States, that specify challenging definitions of the proficiencies that students will need in the information-based economies of the twenty-first century: higher-order thinking skills, problem-solving abilities, investigative and reasoning skills, improved means of communication, and a commitment to lifelong learning. The problem with high-stakes examinations is that they tend to foster performance goals rather than learning goals.

Motivational Effects

There seems little doubt that if sanctions are attached to students' examination performance, some students will work quite hard for the examinations. However, two issues arise in this case. One relates to the type of motivation that is engendered. The long-term consequences of using intrinsic or extrinsic motivation are likely to differ. It can be argued that because examinations rely primarily on extrinsic motivation, they will not provide optimal conditions for developing or sustaining interest in learning (intrinsic motivation). Intrinsic motivation is more likely to result in higher levels of sustained interest in an activity and higher levels of learning.

The second issue relates to the extent to which examinations are successful in motivating *all* students. One would expect a variety of social and psychological factors to determine whether or not individuals will select certain goals in the first place and, if they do, the degree of effort they will use to pursue the goals. The factors include (a) individuals' perceptions of their ability as high or low, (b) their belief that effort will lead to a desired goal (which will reflect their past record of achievement), and (c) individuals' perceptions of the locus and controllability of causes of success and failure, specifically if the locus is perceived to be internal (such as ability, effort) or external (such as luck, a quota system, or malpractice) (Kellaghan, Madaus, and Raczek 1996). When students reach a negative conclusion about their prospects of success, there is evidence that some respond by indifference to the examination process (Hargreaves 1989); others by disappointment and anger (Edwards and Whitty 1994), health problems (Hurrelmann and others 1988), discord and open rebellion (Kariyawasam 1993); and, in extreme cases, as noted earlier, by committing suicide (Nock and Kazdin 2002).

Increase in the Resources Devoted to Examination Preparation

Private supplementary tutoring (also called "shadow education") may be defined as instruction in academic subjects provided by tutors for

a fee to supplement mainstream schooling; the instruction may be one to one, in small groups, or in large classes (Bray 1999, 2003, 2013; Bray and Kobakhidze 2014; Bray, Kwo, and Jokic 2016; Chan and Bray 2014). Private supplementary tutoring is firmly rooted in the culture and education systems of some countries. Terms associated with schools that offer additional tuition or cram schools include *bimbel* (Indonesia), *buxiban* (China), *dershane* (Turkey), *frontistirio* (Greece), *hagwon* (Korea), *juku* (Japan), and *escuela de cursos intensivos* (Spain). Over a century ago, the head of Mauritius' Royal College complained that students were taking private lessons provided by staff of the college, but that nothing could be done about it (Foondun 2002).⁶ Private tutoring has greatly expanded in recent years and has been interpreted as part of a global shift from public to private institutions, with a variety of previously public roles being given to the private sector ("marketization") (Chan and Bray 2014). Such tutoring has been described as a "major industry" and a "major phenomenon" in Egypt; Hong Kong SAR, China; India; Japan; Kenya; Korea; Mauritius; and Taiwan, China. In Egypt, where 60 percent of secondary school students receive private tuition, the amount spent annually has been reported to be equivalent to US\$2.4 billion, which amounted to 27 percent of government spending on education in 2011 (UNESCO 2014) (see box 5.2).

BOX 5.2

Arab Republic of Egypt: Shadow Education

In many instances the same teachers that teach in the secondary schools have their own private tutoring sessions (with as many as 60 students in a "class") in which they teach topics at a greater depth and with greater attention to detail than their own teaching in the public education system classes. These private lessons are taken by a majority of students in general secondary education, who often are absent from regular school classes to work on their private tutoring. They are also seen at the primary and preparatory level, although to a lesser degree. They are expensive in relation to the average family income and consume students' evening or afternoon hours, in a dramatic preparation for the *thanawiya amma*.

Source: OECD 2015.

India has very high rates of tutoring. Government statistics indicated in 2014 that about one in five primary students received private tuition and that this figure rose to over one in three at the senior and senior secondary levels (Government of India, 2015, A-24, cited in Ghosh and Bray 2018). A small-scale qualitative study in urban Maharashtra involved science students who had just completed class 12 (the end of senior secondary school) and who had attended private-aided schools until class 10. During this time, they had received private supplementary tuition (PST) (Bhorkar and Bray 2018). The results indicated that students had availed themselves of PST throughout their school careers; PST played an increasingly important role as students approached Class 10, when they took the Secondary School Certificate examination, a finding which was consistent with the findings of another study conducted in Karnataka (Ghosh and Bray 2018). The impact of PST became more pronounced during grades 11 and 12, as students prepared for their final High School Certification and entrance examinations for tertiary education. In their final two years, these students spent more time receiving PST than in their mainstream schools, in which they had to register in order to sit for the High School Certification. Both Indian studies indicate that parents made considerable financial sacrifices to pay for private tuition; in one study they paid between two and eight months of their household incomes for additional tutoring in expectation of gains in their children's examination performance (Bhorkar and Bray 2018). In Sri Lanka, where public education is free, data suggest that PST accounts for 36 percent of the household education budget and involves participants from all socioeconomic groups (Dundar et al. 2017).

Positive effects attributed to private tutoring relate to improving student achievement, helping students understand mainstream lessons, or playing a remedial role. Private tutoring also provides income for teachers; for some, it actually serves as the main source of income, even if teachers keep their jobs to provide legitimacy and as a means of recruiting clients (UNESCO 2014).

A variety of negative effects have been attributed to private tutoring. First, although it may be considered a private investment in human capital, private tutoring can be quite expensive, especially for poor families. Second, it increases stress on students. Third, it can

PHOTO 5.2

China: Teenagers Emerging from High School Devoted to Cramming Just before 11 p.m. after a Routine Study Day



Photograph © Sim Chi Yin/Magnum Photos. Reproduced with permission from Magnum Photos; further permission required for reuse.

result in fatigue in both students and teachers (see photo 5.2). Fourth, students come to rely more on private tutors than on mainstream teachers, sometimes neglecting to attend their regular classes. Fifth, private tutoring, insofar as it is successful, maintains and exacerbates social inequalities as prosperous households can invest in more and better tutoring than poor households. Sixth, teachers may make less effort in class on the assumption that students have a “safety net” outside the school. Seventh, private tutoring is open to corrupt practices when it is provided by students’ mainstream teachers or when tutors have a role in examining students. For this reason, teachers are prohibited from tutoring their own mainstream students in some jurisdictions (for example, Hong Kong SAR, China).

Two aspects of the relationship between private tutoring and examinations merit comment. First, to the extent that tutoring focuses on the preparation of students for examinations (identifying topics likely to be examined, teaching “examination technique,”

providing sample answers, taking advantage of the marking scheme), it serves to reinforce the role of examinations in distorting the educational process as envisaged in school curricula. Second, private tutoring reinforces the obstruction of efforts to make the education system less dependent on examinations as examinations are seen as essential ingredients in the demand for tutors' services.

Government responses to private tuition reflect a wide range of perceptions of its value and the feasibility of controlling it (Bray 2003). In some jurisdictions, direct action has been taken to limit it. In Cambodia and Myanmar, for example, private tutoring has been banned at different times over various concerns, including its perceived negative impact on social equity. Korea imposed curfews on private tutoring institutions (*hagwons*) operating after 10 pm in an effort to address the problems of student fatigue and lack of sleep. Mauritius prohibited private tutoring for standards 1 to 3 and placed limitations on the number of students permitted in a group and on the number of hours that students may be tutored. In other jurisdictions, authorities feel constrained and helpless. For example, the capacity to police tutorial operations does not exist in many countries (such as Kenya and Nigeria). Other jurisdictions leave matters to the market, considering private tutoring as something to be encouraged on balance (such as Singapore and Taiwan, China).

CONCLUSION

The advantages attributed to public examinations lie primarily in their objectivity and impartiality in fulfilling the essential societal function of assessing students' knowledge and competencies and in the decisions that follow assessment. Additional perceived advantages relate to the extent that examinations contribute, at the micro level, to the development of individual student learning and, at the macro level, to the development of educational quality, realized as the cognitive output of the education system, which, in turn, contributes to a nation's global capital. Commonly perceived disadvantages associated with public examinations include distorting the curriculum, teaching to the test, grade retention and early dropout, and a variety

of forms of malpractice. It may be noted in passing that, although most of the comments in the literature on the negative effects of high-stakes assessments relate to public examinations, observers in Latin America point out that the practice of “teaching to the test” is increasingly present in several countries; this practice has been attributed to the implementation of census-based national assessments and the publication of results for individual schools (Rizo 2010).

What all this indicates is that no examination system is likely to be perfect. Trade-offs and compromises that take account of the state of development of an educational system, as well as cultural and economic realities, will be required among competing goals and values, while exceptional efforts may be required to minimize anticipated negative effects.

NOTES

1. For example, Sri Lanka’s Grade 5 Scholarship Examination was originally designed to select high-achieving disadvantaged students for scholarships for highly ranked schools. Over time its purpose has changed and it no longer is limited to children from disadvantaged families.
2. In England, Wales, and Northern Ireland, however, major newspapers publish results in the form of league tables, a process that promotes a highly public form of school accountability.
3. Berliner 2011; Clarke, Haney, and Madaus 2000; Crooks 1988; DeLuca 1993; Eisemon, Patel, and Abagi 1987; Haertel 1989; Haney 2000; Hill 2013; Kellaghan and Greaney 1992, 2004; Little 1982; Madaus 1988; Madaus and Clarke 2001; Madaus and Greaney 1985; Madaus and Kellaghan 1992; Madaus, Russell, and Higgins 2009.
4. The reform program was the Education Reform for the Knowledge Economy, 2003–09.
5. Bashir et al. (2018) report that data from some African countries (such as Burkina Faso, the Democratic Republic of Congo, Ethiopia, Ghana, Kenya, and Rwanda) recorded an increase in the repetition rate in the grade immediately prior to the national examination, while others (such as Côte d’Ivoire, Malawi, Mozambique, Nigeria, and Senegal) showed high repetition rates in the grade in which the exam was administered. Madaus and Greaney (1985), using historical Irish data, report an increase

in the grade 5 repetition rate, the grade before the national primary school leaving examination; the examination was subsequently abolished.

6. An unintended effect of the introduction of the civil service examination in Britain was the opening of special “cramming” establishments to prepare students for examinations.

REFERENCES

- Adams, R. 2015. “Surge in Young People Seeking Help for Exam Stress.” *Guardian*, May 14. <http://www.theguardian.com/education/2015/may/14/calls-to-childline-over-exam-stress-break-records>.
- Adnan, U., and M. A. Mahmood. 2014. “Impact of Public Examination on Teaching of English: A Washback Perspective.” *Journal of Education and Practice* 5: 132–39. <http://iiste.org/Journals/index.php/JEP/article/view/10649/10876>.
- Ames, C. 1992. “Classrooms: Goals, Structures, and Student Motivation.” *Journal of Educational Psychology* 84: 261–71.
- Bashir, S., M. Lockheed, E. Ninan, and J.-P. Tan. 2018. *Facing Forward: Schooling for Learning in Africa*. Washington, DC: World Bank.
- BBC News. 2019. “India School-Leaving Exam: The Controversy That Cost 23 Lives.” July 18. <https://www.bbc.com/news/world-asia-india-48050020>.
- Berliner, D. 2011. “Rational Responses to High-Stakes Testing: The Case of Curriculum Narrowing and the Harm That Follows.” *Cambridge Journal of Education* 41: 287–302.
- Bhorkar, S., and M. Bray. 2018. “The Expansion and Roles of Private Tutoring in India: From Supplementation to Supplantation.” *International Journal of Educational Development* 62: 148–56.
- Bishop, J. H. 1995. “The Impact of Curriculum-Based External Examinations on School Priorities and Student Learning.” Center for Advanced Human Resource Studies Working Paper 95-30, School of Industrial and Labor Relations, Cornell University, Ithaca, NY.
- . 1998a. “Do Curriculum-Based Exit Exam Systems Enhance Student Achievement?” Research Report Series RR-40, Consortium for Policy Research in Education, Graduate School of Education, University of Pennsylvania, Philadelphia.
- . 1998b. “The Effect of Curriculum-Based External Exit Exam Systems on Student Achievement.” *Journal of Economic Education* 29: 171–82.

- . 1999. "Are National Examinations Important for Educational Efficiency?" *Swedish Economic Policy Review* 6: 349–98.
- . 2005. "High School Exit Examinations: When Do Learning Effects Generalize?" In *Uses and Misuses of Data for Educational Accountability and Improvement: 104th Yearbook of the National Society for the Study of Education*, part 2, edited by J. L. Herman and E. H. Haertel, 260–88. Chicago: National Society for the Study of Education.
- Black, P. 1996. "Commentary." In *Examining the Examinations*, edited by E. D. Britton and S. A. Raizen, 19–21. Dordrecht, Netherlands: Kluwer.
- Bloom, B. S., ed. 1956. *Taxonomy of Educational Objectives: Handbook 1—Cognitive Domain*. London: Longmans Green.
- Bray, M. 1999. *The Shadow Education System: Private Tutoring and Its Implications for Planners*. Paris: International Institute for Educational Planning.
- . 2003. *Adverse Effects of Private Supplementary Tutoring*. Paris: International Institute for Educational Planning.
- . 2013. "Benefits and Tensions of Shadow Education: Comparative Perspectives on the Roles and Impact of Private Supplementary Tutoring in the Lives of Hong Kong Students." *Journal of International and Comparative Education* 2: 18–30.
- Bray, M., and M. N. Kobakhidze. 2014. "Management Issues in Research on Shadow Education: Challenges and Pitfalls Encountered in TIMSS and PISA." *Comparative Education Review* 58: 590–620.
- Bray, M., O. Kwo, and O. Jokic, eds. 2016. *Researching Private Supplementary Tutoring*. Dordrecht, Netherlands: Springer.
- Burdett, N. 2017. "Review of High Stakes Examination Instruments in Primary and Secondary School in Developing Countries." RISE Working Paper 17/018. https://www.riseprogramme.org/sites/www.riseprogramme.org/files/publications/RISE_WP-018_Burdett.pdf.
- Chan, C., and M. Bray. 2014. "Marketized Private Tutoring as a Supplement to Regular Schooling: Liberal Studies and the Shadow Sector in Hong Kong Secondary Education." *Journal of Curriculum Studies* 46: 351–88.
- Clarke, M., W. Haney, and G. F. Madaus. 2000. "High Stakes Testing and High School Completion." *National Board of Educational Testing and Public Policy Statements* 1: 1–11.
- Crooks, T. J. 1988. "The Impact of Classroom Evaluation Procedures on Students." *Review of Educational Research* 58: 438–81.
- De Luca, C. 1994. *The Impact of Examination Systems on Curriculum Development: An International Study*. ED.94/WS.17. Paris: United Nations Educational, Scientific, and Cultural Organization.

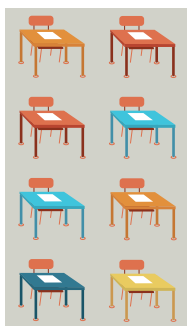
- Dundar, H., T. Béteille, M. Riboud, and A. Deolalikar. 2014. *Student Learning in South Asia: Challenges, Opportunities, and Policy Priorities*. Washington, DC: World Bank.
- Dundar, H., B. Millot, M. Riboud, M. Shojo, H. Aturupane, S. Goyal, and D. Raju. 2017. *Sri Lanka Education Sector Assessment: Achievements, Challenges, and Policy Options*. Washington, DC: World Bank.
- Edwards, T., and G. Whitty. 1994. "Education: Opportunity, Equality and Efficiency." In *Paying for Inequality*, edited by A. Glyn and D. Miliband, 44–64. London: Rivers Oram.
- Eisemon, T. O. 1990. "Examination Policies to Strengthen Primary Schooling in African Countries." *International Journal of Educational Development* 10: 69–82.
- Eisemon, T. O., V. L. Patel, and J. Abagi. 1987. "Read These Instructions Carefully: Examination Reform and Improving Health Reform in Kenya." *International Journal of Educational Development* 7: 1–12.
- Foondun, A. R. 2002. "The Issue of Private Tuition: An Analysis of the Practice in Mauritius and Selected South-East Asian Countries." *International Review of Education* 48: 485–515.
- Fuchs, T., and L. Woessmann. 2004. "What Accounts for International Differences in Student Performance? A Re-examination Using PISA Data." Institute for Economic Research, Munich.
- Ghosh, P., and M. Bray. 2018. "Credentialism and Demand for Private Supplementary Tutoring: A Comparative Study of Students Following Two Examination Boards in India." *International Journal of Comparative Education and Development* 20: 33–50.
- Government of India. 2015. "Key Indicators of Social Consumption in India: Education." National Sample Survey 71st Round, Ministry of Statistics and Programme Implementation, National Sample Survey Office, New Delhi.
- Greaney, V., and T. Kellaghan. 2008. *Assessing National Achievement Levels in Education*. Washington, DC: World Bank.
- Grolnick, W. S., and R. M. Ryan. 1987. "Autonomy in Children's Learning: An Experimental and Individual Difference Investigation." *Journal of Personality and Social Psychology* 52: 890–98.
- Haertel, E. 1989. "Student Achievement Tests as Tools of Education Policy: Practices and Consequences." In *Test Policy and Test Performance: Education, Language and Culture*, edited by B. R. Gifford, 25–60. Dordrecht, Netherlands: Kluwer Academic.
- Haney, W. 2000. "The Myth of the Texas Miracle in Education." *Education Policy and Analysis Archives* 8: 41.

- Hanushek, E. A., and L. Woessmann. 2011. "How Much Do Educational Outcomes Matter in OECD Countries?" Paper presented at the 52nd Panel Meeting on Economic Policy, Rome. (Also available in *Economic Policy* 26 (67) (July 2011): 427–91.)
- Hargreaves, A. 1989. "The Crisis of Motivation in Assessment." In *Educational Policies: Controversies and Critiques*, edited by A. Hargreaves and D. Reynolds, 41–63. New York: Falmer.
- Heyneman, S. P. 1987. "Uses of Examinations in Developing Countries: Selection, Research, and Education Section Management." *International Journal of Educational Development* 7: 251–63.
- Heyneman, S. P., and A. W. Ransom. 1992. "Using Examinations and Testing to Improve Educational Quality." In *Examinations: Comparative and International Studies*, edited by M. A. Eckstein and H. J. Noah, 105–20. Oxford, UK: Pergamon.
- Hill, P. 2013. *Examination Systems*. Paris: United Nations Educational, Scientific and Cultural Organization.
- Hossain, M. Z. 2009. "Creative Questions: Validity and Reliability of Secondary School Certificate Examination." *Bangladesh Education Journal* 8: 33–43. Cited in H. Dundar, T. Béteille, M. Riboud, and A. Deolalikar, *Student Learning in South Asia: Challenges, Opportunities, and Policy Priorities* (Washington, DC: World Bank, 2014), 301.
- Hurrelmann, K., U. Engel, B. Holler, and E. Nordlohne. 1988. "Failure in School, Family Conflicts, and Psychosomatic Disorders in Adolescence." *Journal of Adolescence* 11: 237–49.
- Kariyawasam, T. 1993. "Learning, Selection and Monitoring: Resolving the Roles of Assessment in Sri Lanka." Paper presented at the conference "Learning, Selection and Monitoring: Resolving the Roles of Assessment," Institute of Education, University of London.
- Kellaghan, T., and V. Greaney. 1992. "Using Examinations to Improve Education: A Study in Fourteen African Countries." Technical Paper 165, Africa Technical Department Series, World Bank, Washington, DC.
- . 2004. *Assessing Student Learning in Africa*. Washington, DC: World Bank.
- Kellaghan, T., G. F. Madaus, and A. Raczek. 1996. *The Use of External Examinations to Improve Student Motivation*. Washington, DC: American Educational Research Association.
- Klein, E. D. 2010. "The Differing Purposes and Effects of State-Wide Exit Examinations: A Comparison of Finland, Ireland and the Netherlands from a Government Perspective." Paper presented at the 36th Annual Conference of the International Association for Educational Assessment, Bangkok, August 22–27. http://www.iaea.info/documents/paper_4d519c43.pdf.

- Kwon, S. K., M. Lee, and D. Shin. 2017. "Educational Assessment in the Republic of Korea: Lights and Shadows of High-Stake Exam-Based Education System." *Assessment in Education: Principles, Policy & Practice* 24: 60–77.
- Little, A. 1982. "The Role of Examinations in the Promotion of 'The Paper Qualification Syndrome.'" In *Paper Qualifications Syndrome (PQS) and Unemployment of School Leavers: A Comparative Sub-regional Study*, edited by J. Oxenham, 176–95. Addis Ababa, Ethiopia: International Labour Office.
- Little, A., H. Aturupane, and M. Shojo. 2013. "Transforming Primary Education in Sri Lanka: From a 'Subject' of Education to a 'Stage' of Education." South Asia Human Development Working Paper 61, World Bank, Washington, DC.
- Madaus, G. F. 1988. "The Influence of Testing on the Curriculum." In *Critical Issues in Curriculum: Eighty-Seventh Yearbook of the National Society for the Study of Education*, part 1, edited by L. N. Tanner, 83–121. Chicago: University of Chicago Press.
- Madaus, G. F., and M. Clarke. 2001. "The Adverse Impact of High Stakes Testing on Minority Students: Evidence from One Hundred Years of Test Data." In *Raising Standards or Raising Barriers? Inequality and High-Stakes Testing in Public Education*, edited by G. Orfield and M. L. Kornhaber, 85–106. New York: Century Foundation Press.
- Madaus, G. F., and V. Greaney. 1985. "The Irish Experience in Competency Testing: Implications for American Education." *American Journal of Education* 93: 268–94.
- Madaus, G. F., and T. Kellaghan. 1991. "Student Examination Systems in the European Community: Lessons for the United States." In *Science Assessment in the Service of Reform*, edited by G. Kulm and S. M. Malcolm, 189–232. Washington, DC: American Association for the Advancement of Science.
- . 1992. "Curriculum Evaluation and Assessment." In *Handbook of Research on Curriculum*, edited by P. W. Jackson, 119–54. New York: Macmillan.
- Madaus, G. F., and J. Macnamara. 1970. *Public Examinations: A Study of the Irish Leaving Certificate*. Dublin: Educational Research Centre.
- Madaus, G. F., M. Russell, and J. Higgins. 2009. *The Paradoxes of High Stakes Testing*. Charlotte, NC: Information Age.
- Martinez, M. E. 1999. "Cognition and the Question of Test Item Format." *Educational Psychologist* 34: 207–18.
- Nock, M. K., and A. E. Kazdin. 2002. "Examination of Effective, Cognitive, and Behavioral Factors and Suicide-Related Outcomes in Children." *Journal of Clinical Child and Adolescent Psychology* 31: 48–58.

- OECD (Organisation for Economic Co-operation and Development). 2015. *Schools for Skills—A New Learning Agenda for Egypt*. Paris: OECD.
- Oman, Ministry of Higher Education, Higher Education Admission Center. 2019. “Student Guidebook.” Ministry of Higher Education, Muscat, Oman. http://www.heac.gov.om/media/doc/GED_guidebook_2019-2020.pdf.
- Popham, W. J. 1983. “Measurements as an Instructional Catalyst.” *New Directions for Testing and Measurement* 17: 19–30.
- Rizo, F. M. 2010. “Assessment Practice in Policy Context: Latin American Countries.” In *International Encyclopedia of Education*, 3d ed., edited by P. Peterson, E. Baker, and B. McGaw, 479–85. Oxford, UK: Elsevier.
- Ross, K. N., and L. Maehlck, eds. 1990. *Planning the Quality of Education: The Collection and Use of Data for Informed Decision Making*. Oxford: Pergamon and Paris: United Nations Educational, Scientific, and Cultural Organization.
- Somerset, A. 1996. “Examinations and Educational Quality.” In *Assessment in Transition. Learning, Monitoring and Selection in International Perspective*, edited by A. Little and A. Wolf, 263–84. Oxford, UK: Pergamon.
- Spratt, M. 2005. “Washback and the Classroom: The Implications for Teaching and Learning of Studies of Washback from Exams.” *Language Teaching Research* 9: 5–29.
- Tyumeneva, Y. 2013. *Disseminating and Using Student Assessment Information in Russia*. Washington, DC: World Bank.
- UNESCO (United Nations Educational, Scientific, and Cultural Organization). 2014. *Global Monitoring Report 2013/4: Teaching and Learning—Achieving Quality for All*. Paris: UNESCO.
- Woessmann, L. 2000. *Schooling Resources, Educational Performances: The International Evidence*. Kiel, Germany: Kiel Institute of World Economics.

CHAPTER 6



THE PROVISION OF FEEDBACK FROM EXAMINATIONS

INTRODUCTION

Public examinations are used primarily for two purposes: to certify student achievement, that is, to assess student competence in relation to some agreed standard, and to select students for further education or for jobs. They are usually considered “high stakes” as they have importance consequences for students and their teachers, schools, and parents, as is the case when performance on a public examination determines graduation, promotion, or selection for further education or a job. Although examination results may be subject to close scrutiny by politicians, schools, the teaching profession, and the public for what they can tell about the functioning of the education system, their potential for improving the quality of teaching and learning is, probably, rarely realized. This chapter describes how in some education systems feedback about education performance is formalized in guidance and incentives, with the objective of positively affecting educational practice.

GUIDANCE FEEDBACK

The Organisation for Economic Co-operation and Development reported feedback data on the 2006 performance of students on a public examination at the end of lower-secondary schooling (grade 9 or 10) for a number of member states (OECD 2008).¹ Feedback to schools described as “high level of influence” (OECD 2008, 471) was provided in Iceland, Ireland, and Scotland, and as “moderate” in Estonia and France. In Italy, Portugal, and Turkey, examinations results were deemed to have minimal if any impact on the feedback provided to schools and teachers. The students’ examination performance was considered to have had no financial impact in terms of school budgets or monetary awards to schools or teachers, except in Scotland and Estonia, where the impact was considered to be low. “Moderate assistance,” justified by student examination results, was provided in Estonia, France, Ireland, and Scotland to help teachers improve their teaching skills. Perhaps of greater significance is the number of countries indicating that they provided no feedback whatsoever on student examination performance.

The situation relating to feedback in the Russian Federation is of particular interest, because national-level public examinations have only comparatively recently been established in that country. An important use of examination results in Russia is the identification of deficiencies in student learning; arising from this, pedagogical guidelines were created at both federal and regional levels to be used in preparing students for examinations in the future.² It is expected that the guidelines, which are aligned with national learning standards, will be used by teachers and students in their choice of topics to study. Teachers and students can access a databank containing items similar to those used in the examination as well as items from past years’ examinations.³ Schools with poor Unified State Examination results are subjected to closer inspection in an effort to identify problems and ways to address them. Some regional ministries also establish funding priorities based on the examination results. It has been noted that schools with the highest potential risk of poor examination performance (for example, rural schools) seem to have focused

the most on examination guidelines; this has been judged to have severely narrowed the content and cognitive skills taught in those schools. Tyumeneva (2013) notes that the implementation strategy to use examination results seems to have failed to link information about examination performance in the various content and skill areas to preservice teacher education programs.

In several education systems in which public examinations are firmly established, general or detailed analyses of how candidates performed in examinations are provided. This is intended to help schools identify weaknesses in their delivery of the curriculum so these can be addressed in future years. Box 6.1 lists general recommendations of the chief examiner based on an analysis of candidate performance on the Irish Leaving Certificate Examination in English in 2013.

Although it may not be feasible to provide information about the performance of individual schools, reports based on the whole cohort of examinees can be used by schools in evaluating the performance of their own students. Where more sophisticated computer systems are

BOX 6.1

General Recommendations, Chief Examiner, English

- Teachers and candidates should pay more attention to knowledge and control of the formal aspects of language (register, paragraphs, syntax, punctuation, and spelling).
- Candidates should adopt a “process” approach to writing (involving planning, drafting, redrafting, and editing).
- Candidates should read more widely from a diverse selection of texts in a wide range of genres.
- Candidates would benefit from a greater acquisition of the concept and terminology of visual literacy (including films).
- Candidates should be encouraged to assess the validity of assertions in texts, to challenge ideas, and to form independent views.

Source: Ireland State Examinations Commission 2013.

available, it may be possible to provide schools with statistical data for comparing the performance of their students with national and regional averages. For example, the National Examinations Center in Lithuania offers schools, for a small charge, the opportunity to receive statistical data that can be used to compare their performance with the performance of similar schools.

Other modes of providing feedback to schools include the publication of examination papers and putting marking schemes in the public domain, showing teachers exactly what examiners were looking for, making examination scripts available after marking, and providing a detailed analysis of candidates' performance on an examination. In Sub-Saharan Africa, the West African Examinations Council's chief examiners' reports for each subject area provide brief general comments on overall student performance and list perceived candidate weaknesses and strengths. The reports, which are made available on a website,⁴ list examination questions and give detailed observations on students' responses to individual questions. Box 6.2 presents the comments on one health education question. The Mauritius examination syndicate also produces timely detailed reports for each subject (Bethell 2016).

BOX 6.2

General Recommendations, Chief Examiner, Health Education

Health Education Paper 2, May–June 2017

Question 2

- a. Explain the term community health. [3 marks]
- b. State three characteristics of work place health. [3 marks]

Observation

This question was attempted by many candidates, and their performance was poor.

In part (a) and (b) the candidates could not correctly define the term community health nor state the characteristics of work place health.

(continued)

BOX 6.2**General Recommendations, Chief Examiner, Health Education (*continued*)**

The expected answers are as follows:

a. EXPLANATIONS OF COMMUNITY HEALTH

This is the comprehensive health program, that includes safe environment or provision of potable water or public health facilities or environmental sanitary facilities for the benefit of the individual and community at large.

b. CHARACTERISTICS OF WORKPLACE-BASED HEALTH

- Meant for workers and their families only
- Prompt medical services are provided
- Medical services provided are subsidized
- The sick bay or clinic could be located within the workplace or outside the work place

Source: WAEC 2017.

In Hong Kong SAR, China, at the end of the examination year, each school receives an overall results summary and a comparison of its results with other schools, as well as item-level analysis data to help it identify its strengths and weaknesses. Photo 6.1 depicts a forum organized by the Hong Kong Examinations and Assessment Authority to discuss the effective use of results.

The *KCPE Newsletter*, published by the Kenya National Examinations Council, is an example of a systematic approach to informing teachers about how students performed on an examination. It contained item analysis data for each examination paper, identifying key topics and skills that were causing difficulty. Examples of poor, average, and good answers to selected questions were presented and implications for teaching identified. Examples of how and why essays received marks have been found to be particularly useful (Somerset 1987). A sample essay with marker's comments is presented in box 6.3.

PHOTO 6.1

School Principal Explaining How to Use Public Examination Results and Other Assessment Data to Help Improve Teaching and Learning at a Forum Organized by the Hong Kong Examinations and Assessment Authority



Photograph © Hong Kong Examinations and Assessment Authority. Reproduced with permission from the Hong Kong Examinations and Assessment Authority; further permission required for reuse.

BOX 6.3

Kenya Certificate of Primary Education: Sample Essay and Marker's Comments

That evening my father came home looking unusually happy. He told me he had something to tell us, but before he could do that, he asked me to go and call Heri. When we both came in we found that everyone else was waiting for the two of us. My father as usual was seated on his armchair with a cigarette in his mouth. My mother was standing by him and my two other sisters were seated at the dining table. We all had an idea of what the good news was but we were unsure.

When everybody had seated down, my father began talking. There was suddenly a cry of joy and everybody came rushing towards me. It was so sudden that I almost fell off my seat. Everybody was hugging me and doing

(continued)

BOX 6.3**Kenya Certificate of Primary Education: Sample Essay and Marker's Comments (*continued*)**

all sorts of funny things. I just could not believe my eyes. What I had been waiting for so anxiously was now revealed. It was too good to be true. I do not know how I made it, but passing my examination was the best news I had ever heard. None of us could control ourselves. Some of us almost jumped high enough to touch the ceiling with our heads. All that was ringing in my head was that I had passed with flying colors and to which school I would be chosen to go.

My father asked everyone to be silent since he had not finished saying what he wanted. Once resettled, he told us that I was among the top people in the school and that I would get an award for that. He also told us that he was to go and check which school had chosen me to be one of their students. I was overjoyed in such a way that I could hardly move from where I was now seated. I could just remember the way I was so nervous while doing the exams and all the words of good luck from my friends and relatives. I was glad that all the messages in the cards had been fulfilled and also the fact that I did not disappoint my parents. This had happened before with Heri, but I had challenged him and beaten him. I could remember all the promises made to me by my uncles and aunts and I felt as though my heart was smiling inside.

After having supper that night, I was called to my parent's room and asked what I wanted to have as a present. I named almost everything one could think of and this made them laugh since they know the situation I was in and in which they had also gone through. Since they could not buy me everything, I was promised the most essential things. I would never forget that day because it was the happiest day of my life and having the feeling inside me makes me feel that in future I will also do well.

Marks Awarded: 37

Marker's Comments:

Although this piece of writing was not the very best, it did represent the top-quality writing, which is not very easy to come by. Candidates who reached this standard were actually very few. Except for such errors as the usage of "seated" instead of the simple past "sat" and such flaws as "after having supper that night"—a candidate such as this would even have scored higher marks had he been more interesting and had he used a greater variety of structures in his writing.

(*continued*)

BOX 6.3**Kenya Certificate of Primary Education: Sample Essay and Marker's Comments (*continued*)**

From the brief analysis of errors in the compositions included in this newsletter and in other compositions written by candidates in 1985 KCPE we discovered that certain kinds of errors were being made repeatedly. Incorrect verb tenses, wrong usages of words, spelling errors, and errors in syntax were widespread.

Examples:

Incorrect verb tenses:

We all sit down and listen to him. (We all sat down and listened to him.)

My father did not agreed. (My father did not agree.)

They took the money and leave him crying. (They took the money and left him crying.)

Numerous similar errors were made by candidates.

Source: Kenya National Examinations Council 1986, cited in Appendix 3, Kellaghan and Greaney 1992.

In order to have a strong impact, examination reports for teachers on the general performance of candidates should meet the following criteria. They should

- Provide subject-specific information, preferably in separate documents
- Contain quantitative data and qualitative observations about overall examination performance
- Identify items that showed particular strengths and weaknesses in candidate performance
- Provide samples of students' work demonstrating good answers and common weaknesses and misconceptions
- Be produced as soon as possible after the examination
- Be automatically distributed to schools rather than "on demand"
- Be provided free of charge⁵

INCENTIVE FEEDBACK

As part of accountability movements that have attained increasing importance in many countries in response to political, social, and economic pressures, interest has grown in the use of examination results as indicators of the effectiveness of schools or of individual teachers. This has often involved linking results to rewards and penalties. In perhaps the first example of payment by results, in 1444 the town of Treviso in Northern Italy tied the schoolmaster's salary to the performance of students on an oral examination based on the grammar of that day (Aries 1962). In 1862 the British government introduced a form of payment by results that resulted in part of teacher payment being based on examinations conducted by the school inspectorate (Midgley 2016).⁶ Both it and similar systems introduced in Australia and in Ireland were later abandoned because of their perceived negative impacts on teaching and learning (Coolahan 1977; Madaus and Greaney 1982; Musgrave 1968; Pawsey 1994). In Russia, examination results are used by regional ministries of education and municipalities to put pressure on different levels of the education system. This has occurred contrary to recommendations of federal agencies and despite the fact that the Unified State Examination suffers from some credibility issues among the general public and education practitioners. Official pressure was achieved through the school accreditation process, school rankings, and the publication of school results. In addition, economic incentives were awarded to teachers and schools, without taking into account the socioeconomic context in which schools operated (Tyumeneva 2013).

In the examination reform of the 1970s in Kenya, in addition to guidance information, district and school orders of merit based on performance on the primary school examination were published (Somerset 1987). These national school rankings, or league tables, were banned because some schools and districts were found to be manipulating the system by presenting only their best students for examination (Akers, Migoli, and Nzomo 2001). Despite the ban, ranking continued at the provincial and district levels (Amunga, Amadalo, and Maiyo 2010). Similarly, in Ghana, the introduction of

school league tables in 2004 led to a sharp increase in malpractice in the examinations for the Senior Secondary School Certificate. A more limited release of district-level examination results at the end of the basic level of schooling has been associated with a positive impact in a few districts (Akuffo-Badoo 2017).

Several specific arguments are used to support the attachment of high stakes for teachers and schools to student performance on examinations. First, it encourages teachers to internalize the norms, values, and expectations of stakeholders (a government ministry, parents) and to accept responsibility for conforming to them. Second, it supports the operation of market mechanisms in the education system, involving competition, contracting, auditing, pay for performance based on objective data, and showing the government is getting value for its considerable education budget. Third, it serves to focus teacher and student endeavors on the goals of instruction and to provide standards of expected achievement that students and teachers can aspire to, thus creating a system of measurement-driven instruction. At a very practical level, the argument in favor of using public examination results for these purposes is supported by the fact that in most education systems public examinations are the only commonly available measure of student achievement. Fourth, examination results can be collected, analyzed, and easily reported from existing databases. Furthermore, presenting results in the familiar format of league tables gives parents simple understandable evidence about their children's schools; parents can examine the rank of their child's school, based on examination results, in a published national league table of all schools that took a particular examination (such as the General Certificate of Secondary Education in England).

On the other hand, there are compelling arguments against the use of examination results as measures of school or teacher effectiveness without giving due consideration to the context in which they were obtained (Greaney and Kellaghan 1995; Kellaghan and Greaney 2001; Kellaghan, Greaney, and Murray 2009). These arguments question the advisability of using examination results to incentivize teachers, either by making public students' examination performance or by providing financial rewards to teachers for high student achievement. It may be because of such considerations that school league tables were

abolished in Northern Ireland and Wales in 2001, followed by Scotland in 2003. Japan's Central Council of Education (2005) advised that results should be used to benefit student learning and that school ranking should be avoided. Singapore, which had used school rankings to help parents make informed school-choice decisions, abandoned the practice in 2012 because of its perceived negative effects, especially on low-ranking schools, and the need to promote a culture that valued such aspects of education as sports, the arts, and character development in addition to academic achievement (Sim 2014).

More recently, England replaced its school-ranking system based on examination results with one that considered progress in eight subject areas. The new system, which came into effect in 2017, compares schools to other schools with similar intakes and has resulted in a substantial change in secondary-school rankings.² Despite the limitations of school league table data, including the inappropriateness of using examination results to compare school progress over time or to consider the statistical errors of measurement associated with individual ranks, each year the media tend to publish overly simplistic information about levels of school achievement (see figure 6.1).

There are several reasons why raw or unadjusted school rankings based on student achievement ("league tables") in an examination or test provide a poor method for comparing schools' performance. First, raw examination results do not separate the aspects of achievement that can be attributed to teachers or schools from other factors that affect achievement: student characteristics (including their earlier achievements); teacher characteristics (including those of teachers in earlier grades); conditions in which students live (including family and community resources and support); education policies and supports provided by relevant public authorities, including curricula and teacher preparation; and school conditions and resources. Statistical adjustments to examination results on the basis of students' prior achievement levels, home background, or level of teachers' qualifications could produce a very different result (Goldstein and Spiegelhalter 1996).

Second, focusing on examination results ignores many other important outcomes of schooling, which may lead to a narrowing of the taught curriculum and the neglect of students thought unlikely to be successful (see chapter 12).

FIGURE 6.1**England and Wales: School League Table Headlines****Secondary school league tables 2019: Find out how your London school did in GCSE and A-Level standings***—Evening Standard, January 25, 2019***Best secondary schools in London***—Sunday Times, November 26, 2017***Best and worst schools in England and Wales revealed - is your child's on the list?***—Mirror, August 15, 2018***BAD EDUCATION England's WORST schools revealed—full list of 2018's under-performing secondary schools***—Sun, January 24, 2019***Top class! These are the best schools in Yorkshire***—Doncaster Free Press, November 23, 2018***Top grammar school in the UK according to GCSE league tables***—Telegraph, April 15, 2019***One in EIGHT secondary schools in England falls below minimum standard—but is YOUR children's school among the best or worst?***—Daily Mail, January 25, 2018*

Third, rankings can vary depending on the school outcome that is examined (Guskey and Kifer 1990). A school's ranking based on the percentage of students getting an A grade in the examination may be quite different from its ranking based on the percentage of students getting at least a C grade. A school that does not do well in examinations might do very well on some other criterion. In this context, it is worth bearing in mind that different types of school-performance measures may serve different policy purposes. If the purpose is to provide information relating to school choice, an argument might be made for adjusting for prior achievement with separate data for different student groups (girls or boys). If, however, the focus is to inform school improvement initiatives, adjustment will be required for factors likely to affect growth in achievement (for example, student background, context, input, or processes) (Muñoz-Chereau and Thomas 2016).

Fourth, rankings can vary depending on when they are computed. A series of studies in England showed that many schools in the top quarter of examination results were in the bottom half seven years later; it concluded that school league tables provided little useful information to help parents select schools (Leckie and Goldstein 2009a, 2009b, 2011).

Fifth, whether or not adjustments are made to school output measures such as examination performance, errors of measurement on which school rankings are based are seldom taken into account when judgments of merit are being made by ministries of education, the media, or the public. As a result, fine distinctions among schools on the basis of achievement-test data are statistically invalid, as they do not take into account the level of uncertainty in these rankings (Foley and Goldstein 2012; Leckie and Goldstein 2011, 2016).

Sixth, schools can manipulate pass rates by such practices as student retention and pressure on students to leave school before reaching the examination (Madaus and Greaney 1982). Seventh, the publication of results may lead to schools that are perceived to be doing well attracting students of high levels of ability and motivation, while those that are perceived to be doing badly may be avoided by such students. Finally, publicizing results can lead to the transfer of more able teachers, low morale in schools, the creation of ghetto schools, and in some instances school closure due to declining enrollment.⁸

The circumstances in which teachers and schools operate are taken into account in three approaches to estimating their effectiveness (Muñoz-Chereau and Thomas 2016).

Contextualized Achievement. Achievement is represented in a multilevel statistical model in which account is taken of socioeconomic, family, and personal data together with school processes. The models seek to use statistical procedures to permit comparisons between schools and teachers, taking into account that schools and teachers may be catering to quite different populations of students.

Value Added. The quality of a school is estimated in a longitudinal design in which student achievement at an earlier stage is taken into account in estimating the school's contribution to students' progress or gain score over a period of time

Contextualized Value Added. This approach controls for both prior achievement and compositional or contextual factors. Account is taken not only of individual student characteristics, but also of peer group effects and the collective and individual influence of socioeconomic factors, as well as political and cultural dimensions of the school.

Contextualized value-added models typically provide the best fit of data (in terms of total variance explained). However, their use was discontinued in the national school indicator system in England because of government perception that they may lower expectations for school and student performance (Muñoz-Chereau and Thomas 2016).

Advocates of value-added approaches to judging teacher effectiveness based on test-score performance point to fact that policy makers and others grasp the simplicity of the concept of judging teachers based on test-score gains. They also point to studies showing that teacher effect sizes are large when considered with other factors related to student achievement and that the influence of effective teachers persists into adulthood (Murphy 2012). Some critics, on the other hand, point to a range of methodological issues relating to value-added studies (notably lack of random assignment of students to teachers) and to the need to administer the same content material from year to year, which is problematic in the context of current public examinations (Murphy 2012). From the perspective of giving incentive feedback to teachers, studies have pointed to the negative impact of high-stakes value-added measures, including teachers “gaming” the system by adopting practices such as nonpromotion of weaker students, cutting back on subjects not examined, teaching to the test to ensure good marks, and resorting to various forms of examination malpractice (see chapter 11) (Madaus and Greaney 1985; Rothstein 2004; Spielman 2017). Educational policy makers considering giving incentives or awarding teachers based on student gain scores might reflect on research literature in economics and management that has highlighted the adverse effects of performance-related incentive schemes (Rothstein 2009). Examples of value-added modeling are provided in box 6.4.

BOX 6.4**Value-Added Modeling**

In Singapore, student scores in the Primary School Leaving Examination were used to predict General Certificate of Education O-level performance four or five years later. The top 15 value-added schools, that is, those deemed to have most helped their students improve academically, were formally recognized.

In Wales, student scores on a test taken at age eleven were used to predict performance on the General Certificate of Secondary Education, taken at 16-plus years of age. Predictions were compared with actual results to produce a form of value-added score for each school. Schools with positive scores were considered to have candidates who exceeded expectations, while those with negative scores were deemed to have candidates who performed below expectations and to have performed less well than their earlier test scores predicted. A further analysis compared results within groups of schools with similar levels of family income and special needs.

Sources: Sim 2014; Jenkins 2011a, 2011b.

The linking of financial rewards to students' scholastic performance can give rise to a range of issues. Problems associated with the practice were identified in two recent studies. In one of these, in response to Portugal's poor performance in international assessments such as the Programme for International Student Assessment (PISA), the single pay scale for teachers was broken into two performance-related scales. Automatic progression was no longer assured. Progression depended on students' academic performance and feedback from parents, along with other criteria, including teachers' attendance record, attendance at training sessions, fulfillment of management and pedagogical duties, and involvement in research projects. Several negative outcomes were observed. The new program actually led to a decrease in student achievement. Teachers eager for promotion awarded high internal assessment marks (which carried considerable weight in students' final marks). The policy also inadvertently encouraged competition and reduced cooperation among teachers and lowered teachers' job satisfaction (UNESCO 2014, box 6.9).

In an experiment in primary schools in Kenya, teachers were rewarded for good student test scores and were penalized if students

did not take end-of-year examinations. While test scores and the number of students taking examinations increased, there was evidence that teachers focused on preparing students for the test (“teaching to the test”). At any rate, test scores did not increase in subject areas that were not taken into account in the teacher-pay formula. Wider anticipated benefits, such as reducing teacher absenteeism and lowering the student dropout rate, did not materialize (Glewwe, Ilias, and Kremer 2010; UNESCO 2014).

CONCLUSION

The type of activities described in this chapter should serve to underline the potential for using examination results to help improve the quality of classroom teaching. Teachers can get valuable feedback, which can help them pay particular attention to the most common errors made by students. Unfortunately, boards and agencies do not seem to have the time nor the resources to analyze the rich information provided by examinations in order to enable them to provide feedback to schools and also to curriculum personnel and providers of preservice and in-service teacher education. Results have been widely used to provide accountability measures of school effectiveness in the form of school ranking based on examination marks and, to a much lesser extent, to carry out value-added types of analyses. Perceived limitations of school rankings have caused them to be banned in some educational systems, partly because of the ability of schools to manipulate results and their impact on low-performing schools. These results underline the need for examination agencies to have access to adequate resources and personnel with high-level and diverse skills to analyze results and provide useful, timely, and accurate information to the educational system and to the wider society. Some of these tasks are basically technical and can be addressed by the provision of adequate funding and training. Others may be more challenging because of their philosophical and political implications. In this context, examination activities should be open to a review of the opinions and beliefs of many stakeholders about the nature of achievement, how it should be assessed, and the uses that are made of information about examinees’ performance.

NOTES

1. Comparable data on feedback are not provided in more recent publications of *Education at a Glance*.
2. Restrictions on access to the Unified State Examination (USE) database for reasons of confidentiality have severely curtailed capacity to analyze the data to inform policy. To address this issue, the Committee on Education of the Public Chamber has initiated discussions about possible access to a depersonalized USE database (Tyumeneva 2013).
3. <http://www.fipi.ru/content/otkrytyy-bank-zadaniy-ege> (in Russian).
4. <https://waeonline.org.ng/e-learning/index.htm>.
5. Following introduction of a charge, the newsletter of the Kenya National Examinations Council reached only about half of the schools (Wasanga and Somerset 2013).
6. Proponents of payment by results, notably Lowe, argued in 1862 that “pecuniary incentives operated for teachers in the same way that they did for men in other trades and professions, and that they would stimulate teachers to greater effort in their work” (Sylvester 1974, 72). In response one critic noted that “to pay teachers by results was like paying physicians only for the ‘patients they have cured’ [Hume 1862, 26]. Society would never adopt this policy for its physicians so why should it for its teachers” (Sylvester 1974, 72).
7. The previous system had ranked schools based on the percentage of students in a school who had earned a grade of C or higher in five General Certificate of Secondary Education subjects. Its replacement is a value-added measure based on students’ results in each of eight subjects compared to the actual achievement of other students with similar prior attainment.
8. On this latter point, US evidence suggests the strategy of closing schools to remedy student achievement in low-performing schools has not been successful (Sunderman, Coghlan, and Mintrop 2017).

REFERENCES

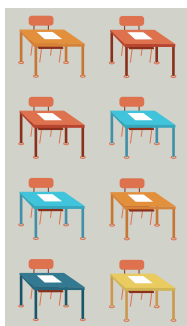
- Akers, J., J. Migoli, and J. Nzomo. 2001. “Identifying and Addressing the Causes of Declining Participation Rates in Kenyan Primary Schools.” *International Journal of Educational Development* 21: 361–74.
- Akuffo-Badoo, F. 2017. “Impact of District Performance League Tables on Performance at the Basic Education Certificate Examination in Ghana.”

- Journal of Educational Assessment in Africa* 12: 74–83. <https://www.umalusi.org.za/docs/misc/2018/AEAA-Journal-2017.pdf>.
- Amunga, J. A., M. M. Amadalo, and J. K. Maiyo. 2010. "Ranking of Secondary Schools and Students in National Examinations: The Perception of Teachers and Students." *Problems of Education in the 21st Century* 20: 10–24. http://www.scientiasocialis.lt/pec/files/pdf/vol20/10-24.Amunga_Vol.20.pdf.
- Aries, P. 1962. *Centuries of Childhood: A Social History of Family Life*. New York: Vintage.
- Bethell, G. 2016. *Mathematics Education in Sub-Saharan Africa*. Washington, DC: World Bank.
- Coolahan, J. 1977. "Three Eras of English Reading in Irish National Schools." In *Studies in Reading*, edited by V. Greaney, 12–26. Dublin: Educational Company.
- Foley, B., and H. Goldstein. 2012. *Measuring Success: League Tables in the Public Sector*. London: British Academy Policy Centre.
- Glewwe, P., N. Ilias, and M. Kremer. 2010. "Teacher Incentives." *American Economic Journal: Applied Economics* 2: 205–27.
- Goldstein, H., and D. J. Spiegelhalter. 1996. "League Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 159: 385–443.
- Greaney, V., and T. Kellaghan. 1995. *Equity Issues in Public Examinations in Developing Countries*. Washington, DC: World Bank.
- Guskey, T. R., and E. W. Kifer. 1990. "Ranking School Districts on the Basis of State-Wide Test Results: Is It Meaningful or Misleading?" *Educational Measurement: Issues and Practice* 9: 11–16.
- Hume, M. 1862. "Brief Comments on the Revised Speech of the Rt. Hon. Robert Lowe on the Revised Code, Feb 13, 1862." London. Cited in D. W. Sylvester, *Robert Lowe and Education* (Cambridge, UK: Cambridge University Press, 1974), 72.
- Ireland State Examinations Commission. 2013. *Leaving Certificate Examinations 2013. English. Chief Examiner's Report*. Athlone, Ireland: State Examinations Commission.
- Japan Central Council of Education. 2005. "Redesigning Compulsory Education for a New Era." Ministry of Education, Culture, Sports, Science and Technology, Tokyo. www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/toushin/06051511.pdf.
- Jenkins, C. 2011a. "Secondary School Performance in Wales Revealed." *BBC News*, May 20. <http://www.bbc.com/news/uk-wales-13456819>.
- . 2011b. "Welsh Schools' GCSE Performances Revealed." *BBC News*, May 23. <https://www.bbc.com/news/uk-wales-13478504>.

- Kellaghan, T., and V. Greaney. 1992. "Using Examinations to Improve Education: A Study in Fourteen African Countries." Technical Paper 165, Africa Technical Department Series, World Bank, Washington, DC.
- . 2001. *Using Assessment to Improve the Quality of Education*. Paris: International Institute for Educational Planning, United Nations Educational, Scientific and Cultural Organization.
- Kellaghan, T., V. Greaney, and T. S. Murray. 2009. *Using the Results of a National Assessment of Educational Achievement*. Washington, DC: World Bank.
- KNEC (Kenya National Examinations Council). 1986. *K.C.P.E. Newsletter 1987*. KNEC, Nairobi. Cited in T. Kellaghan and V. Greaney, "Using Examinations to Improve Education: A Study in Fourteen African Countries," Technical Paper 165, Africa Technical Department Series (World Bank, Washington, DC, 1992), Appendix 3.
- Leckie, G., and H. Goldstein. 2009a. "The Limitations of Using School League Tables to Inform School Choice." *Journal of the Royal Statistical Society: Series A* 172: 835–51.
- . 2009b. "School League Tables: Are They Any Good for Choosing Schools?" *Research in Public Policy: Bulletin of the Centre for Market and Public Organisation* 8: 6–9.
- . 2011. "Understanding Uncertainty in School League Tables." *Fiscal Studies* 32: 207–24.
- . 2016. "The Evolution of School League Tables in England, 1992–2016: 'Contextual Value-Added,' 'Expected Progress,' and 'Progress 8.'" Working Papers in Education 02, University of Bristol, Bristol, UK.
- Madaus, G. F., and V. Greaney. 1982. *Competency Testing: A Case Study of the Irish Primary Certificate Examination*. Cambridge, MA: National Consortium on Testing, The Huron Institute.
- . 1985. "The Irish Experience in Competency Testing: Implications for American Education." *American Journal of Education* 9: 268–94.
- Midgley, H. 2016. "Payment by Results in Nineteenth-Century British Education: A Study in How Priorities Change." *Journal of Policy History* 28: 680–706.
- Muñoz-Chereau, B., and S. Thomas. 2016. "Educational Effectiveness in Chilean Secondary Education: Comparing Different 'Value-Added' Approaches to Evaluate Schools." *Assessment in Education: Principles, Policy and Practice* 23: 26–52.
- Murphy, D. 2012. "Where Is the Value in Value-Added Modeling?" White paper, Pearson Education, Bloomington, MN. http://images.pearsonassessments.com/images/tmrs/Where_is_the_Value_in_Value-Added_Modeling.pdf.

- Musgrave, P. W. 1968. *Society and Education in England since 1800*. London: Methuen.
- OECD (Organisation for Economic Co-operation and Development). 2008. *Education at a Glance 2008: OECD Indicators*. Paris: OECD.
- Pawsey, M. 1994. "The Introduction of Payment by Results into Victorian Schools." *History of Education Review* 23: 1–17. Cited in C. Hooper, "Payment by Results: Teachers' Work, Remuneration, and Student Assessment," *Dictionary of Educational History in Australia and New Zealand (DEHANZ)*, March 25, 2018 (<http://dehanz.net.au>).
- Rothstein, R. 2004. *Class and Schools: Using Social, Economic, and Educational Reform to Close the Black-White Achievement Gap*. Washington, DC: Economic Policy Institute.
- . 2009. "The Perils of Quantitative Performance Accountability." In *Teachers, Performance Pay, and Accountability: What Education Should Learn from Other Sectors*, edited by S.J. Adams, J. Heywood, and R. Rothstein, 67–107. Washington, DC: Economic Policy Institute.
- Sim, C. 2014. "School Ranking." Singapore, National Library Board. http://eresources.nlb.gov.sg/infopedia/articles/SIP_512_2005-01-03.html.
- Somerset, A. 1987. *Examination Reform: The Kenya Experience*. Washington, DC: World Bank.
- Spielman, A. 2017. "HMCI's Commentary: Recent Primary and Secondary Curriculum Research." UK Office for Standards in Education, London. <https://www.gov.uk/government/speeches/hmcis-commentary-october-2017>.
- Sunderman, G. L., E. Coghlan, and R. Mintrop. 2017. "School Closure as a Strategy to Remedy Low Performance." National Education Policy Center, University of Colorado, Boulder. <http://nepc.colorado.edu/publication/closures>.
- Sylvester, D. W. 1974. *Robert Lowe and Education*. Cambridge, UK: Cambridge University Press.
- Tyumenewa, Y. 2013. *Disseminating and Using Student Assessment Information in Russia*. Washington, DC: World Bank.
- UNESCO (United Nations Educational, Scientific and Cultural Organization). 2014. *Global Monitoring Report 2013/4: Teaching and Learning—Achieving Quality for All*. Paris: UNESCO.
- WAEC (West African Examinations Council). 2017. "Health Education Paper 2 May/June." West Africa Examinations Council, Accra, Ghana. <http://waeconline.org.ng/e-learning/Health/HSc226mq2.html>.
- Wasanga, P., and A. Somerset. 2013. "Examinations as an Instrument for Strengthening Pedagogy: Lessons from Three Decades of Experience in Kenya." *Assessment in Education: Principles, Policy and Practice* 20: 385–406.

CHAPTER 7



THE VALIDITY OF PUBLIC EXAMINATIONS

INTRODUCTION

The research literature related to methods of objective educational measurement has tended to focus on the development and validation of standardized achievement, psychological traits, aptitude, intelligence, and personality tests. Comparatively little attention has been given to applying theories and established methods related to the measurement (psychometrics) of these tests and the field of public examinations, despite the fact that performance on these latter assessments probably has greater significance in the lives of students in many parts of the world. This chapter considers validity issues relating to public examinations, with a particular emphasis on content validity. The chapter draws on psychometric studies related to validity to illuminate issues in public examinations, in particular issues related to the appropriateness of the inferences made from examination results, as well as the uses and consequences of these examinations.

Although psychometric approaches used in the construction of standardized tests may be applied to public examinations, major differences between the two should be borne in mind (Baird and Opposs 2018). Standardized tests are usually multiple choice and typically select items that serve to discriminate among students

and to rank-order them. Items that fail to discriminate (such as very easy or very difficult items), even though they measure important curriculum topics, tend to be dropped at the pilot test development stage (Greaney 1980). Items are expected to consistently measure a particular student trait or characteristic that is hidden (such as ability in geometry) but is reflected in test performance. Standardized tests are not normally released after administration as they are usually required for similar purposes on many occasions. Psychometric approaches tend to be supported by sophisticated statistical analytical methodologies. In contrast, public examination items or questions are generally used only once and thus can be released to students and teachers following administration. Public examinations hail from a long-established school or university assessment tradition (see chapter 3). They place a higher priority on criterion-referenced testing and favor the use of short and extended response-type questions that match the official curriculum (Baird and Opposs 2018). They are not, however, generally associated with the high-level statistical analytical techniques associated with standardized tests.

Validity is paramount among principles of educational measurement as it relates to the appropriateness of the inferences, uses, and consequences of an assessment (AERA, APA, and NCME 2014; Cambridge Assessment 2017). It speaks directly to the extent to which a claim about an examinee based on data obtained in an examination is justified. For example, does an examinee's performance indicate that the examinee has "mastered" the curriculum? Can you infer that a candidate with a score of 56 percent is "better" than one with a score of 54 percent? In this context, it should be noted that neither standardized tests nor public examinations are merely statements about particular observations in and of themselves. Both involve claims about an attribute or quality of a person (Cronbach 1970; Lissitz 2009a; Messick 1989; Mislevy et al. 2003).

This chapter considers the validity of public examinations from a number of perspectives: content, criteria, construct, and prediction. It describes threats to validity related to characteristics of the examination (lack of alignment with the syllabus, method of measurement, choice, overpredictability, administrative conditions, aspects of administration), factors associated with examinees (personal characteristics, test-taking strategies), scoring procedures and aggregation of

scores, and examination impact. Based on this review, the chapter offers a number of recommendations to improve the validity of public examinations.

CONTENT VALIDITY

The claim to validity of external public examinations rests primarily on their content (Lissitz 2009b). Syllabi will provide more or less detailed guidance regarding content to be covered and skills and knowledge to be acquired. The key validity issue concerns the extent to which an examination adequately represents the content, knowledge, and skills specified in the syllabus. Thus, the table of specifications that defines the extent of the subject matter coverage of an examination (the domain) is paramount (see photo 7.1). This notion

PHOTO 7.1

Ireland: Examination Paper Setter Checking Content Levels and Cognitive Skills Being Considered for an Economics Examination



Photograph © Ireland, State Examinations Commission. Reproduced with permission from the Ireland State Examinations Commission; further permission required for reuse.

of validity fits well with the function of an examination as certifying the knowledge and skills of students. Professional judgment determines the extent to which an examination adequately represents the syllabus.

Since the knowledge and skills specified in curricula and syllabi are generally very numerous, the only way an examination agency can assess how well a student has learned is to measure the student's performance on a selected sample of topics. How the student responds to the topics, however, is of interest only insofar as it enables the examination authority to infer how well the student has acquired the knowledge and skills prescribed in the area, domain, or sphere of knowledge being assessed. A number of conditions have to be met if tasks from a domain are to be accepted as providing an estimate of performance in the domain as a whole. First, the domain should have been accurately defined with reference to content, cognitive level, and item type. Second, judges should be in agreement that observed performance can be considered representative of the domain. Third, procedures should be in place to control random and systematic errors. Finally, performance should be evaluated appropriately and fairly (Kane 2006).

Given the constraints under which examinations are administered, it is extremely difficult to design ones that would require candidates to exhibit the level of complexity that characterizes the domain they are assessing (for example, the ability to interpret and evaluate complex matters). It is for this reason that new forms of assessment in public examinations have been proposed and are being implemented in many education systems.

OTHER ASPECTS OF VALIDITY

The criterion-related validity of a public examination is usually based on the degree of empirical relationship between examination scores and criterion scores (such as independent standardized test scores or ratings) expressed in terms of correlations or regressions. While *content validity* may be viewed as a singular property of a test or

examination, an examination or test could be assessed in the context of many criterion-related validities. This might involve comparing performance in a job with performance on a test that assessed the knowledge and skills required for the job. Performance on a school leaving examination might also be compared with performance in later education (considered below under predictive validity), or with other performance measures of what students know and can do. Efforts in developing countries to move examination systems from a norm-referenced approach in which achievement was graded in ways that depended on the performance of other students to a more criterion-referenced approach were articulated by the World Conference on Education for All in 1990. This strategic shift is an important consideration when the primary focus in an examination is on the selection of students for higher levels of education. It represented a concern with what students were learning as a result of their experience in schools (Lewin 1997).

Construct validity refers to how well a particular test or examination measures what it claims to measure. If a test has adequate construct validity, students' scores on the test should correlate well with scores on other tests that assess the same construct. A study in Ireland in which the performance of students on a public examination was related to their performance on the Organisation for Economic Co-operation and Development Programme for International Student Assessment (PISA) tests in reading, mathematics, and science might be taken as providing evidence on the construct validity of the examination. The PISA tests, it is claimed, measure competencies that students will need in later life, rather than the outcomes of exposure to any particular curriculum. Because programs of study in Irish schools also have the general long-term objective of preparation for later life, if the PISA tests are valid measures of competencies needed for later life (and that is open to question), a consideration of students' performance on these tests compared to their performance on the examination is of interest. The relationships between PISA and external examination performance in all three domains were in fact moderately strong (correlations were either 0.73 or 0.74), suggesting that, despite differences in the context, content, and methods of the

assessments, there was considerable overlap in the achievements measured (Shiel et al. 2001).

Studies have also provided evidence of the *predictive validity* of public examinations. For example, scores on public examinations used to select students for teacher training have been found to correlate with students' academic performance in training (Greaney, Burke, and McCann 1999). Public examinations have also been found to predict performance in university studies better than scholastic aptitude tests (Choppin and Orr 1976; O'Rourke, Martin, and Hurley 1989; Powell 1973).¹ Other studies, however, have found the relationship between examination performance, on the one hand, and completing a university degree, the quality of the degree obtained, or the performance after graduation, on the other hand, as inconclusive (European Parliament 2014). Issues related to predictive validity studies include possible limitations in both the predictor and the criterion variables. Predictor variables, for instance, may be limited to grades in a minimum number of subjects, use rank position, or use differential subject weighting; criterion variables might be based on first-year grade average score, or level of final degree awarded across subjects, or marks in a final graduation examination. Predictor studies involving highly competitive degree programs (such as law and medicine in many countries) are also likely to encounter problems of restricted predictor score variance because successful students will have scored highly on the examination used to select entrants; this limits the possibility of establishing a statistical relationship with a criterion variable (such as the quality of the final degree).

An overall evaluation of an assessment procedure, such as a public examination, should reflect on the consequences of the procedure. Discussions of the consequences of assessments usually distinguish between intended positive consequences and unintended negative consequences (see chapter 11). What is positive for one observer, of course, may be negative for another. For example, narrowing the curriculum may be positive for those who want to see greater emphasis on basic skills, but may be negative for those who say that access to important content areas, as well as to creative and enjoyable activities, may be restricted (Berliner 2011).

THREATS TO VALIDITY

Examination Characteristics

Lack of Alignment or Underrepresentation of the Content, Knowledge, and Skills of the Domain Being Assessed

Lack of alignment occurs when parts of the target domain (such as the mathematics syllabus) are not included in the assessment or given inadequate weight in examination tasks. Examinations are often too short or too limited in time to achieve an adequate degree of coverage of the knowledge and skills in the domain being assessed. Some multiple-choice tests have too few items, while essay and performance items are particularly vulnerable to the charge of underrepresentation of the domain or curriculum area being examined. It has been argued that when a domain has been set out comprehensively, many hours of testing would be required to assess it with any accuracy (Black 1996).

Lack of alignment with the curriculum also occurs when examinations are limited to written tasks taken in large-group settings in which all examinees take the same tasks under similar conditions. While this approach will improve reliability, it will have the effect of reducing the areas of a domain or curriculum content that are assessed. Extending the areas that are assessed in the interest of improving validity may, on the other hand, reduce reliability.

The assessment of practical skills in a public examination is particularly problematic. In some systems, practical skills are simply ignored. In others, paper-and-pencil² surrogates are used. A problem with this approach is that students do not behave in the same way in practical systems as they do in paper-and-pencil surrogates. In systems in which practical skills are actually assessed, the time allotted may be too short to allow for the assessment of important skills (Black 1996).

Another factor that may contribute to underrepresentation becomes an issue when the selection function of an examination (such as for university admission) takes precedence over its certification function. In this situation, pitching the difficulty of the examination at a level at which maximum discrimination is required may result in excluding the perceived less difficult sections of the domain. In some instances, the contents of tertiary selection examinations may

have little in common with the curriculum covered by the students (World Bank 2018) and may focus instead on other attributes.³

The overall effect, for whatever reason, of limiting an examination to which high stakes are attached to a subset of the target domain is that it is likely to lead to instruction and test preparation that are aimed specifically at the expected content and form of the examination, making test performance unrepresentative of performance in the target domain (curriculum) as a whole (Kane 2001).

Method of Measurement

A considerable portion of variance in test or examination scores may be due to the form or method of assessment used as well as to the individual characteristics (traits, achievements) that the test or examination was designed to measure (Campbell and Fiske 1959). As a consequence, a candidate who has grown up in a modern information technology environment and who has tended to type rather than write may be at a disadvantage when faced with a traditional paper-and-pencil examination. The use of multiple measurement methods will help control the construct irrelevant variance⁴ created by the forms of an examination by averaging method effects (Kane 2006; Messick 1989). Multiple indicators will also provide an opportunity for individuals who are disadvantaged by the use of one assessment method to offer alternative evidence of their expertise.

Choice

Choice in an examination can be considered appropriate if syllabi are framed in such a way that teachers are allowed a degree of latitude in covering a subject. Choice may also be provided on the basis that it allows examinees to select a curriculum area or task they are familiar with, which may lead to better performance than if they had been assigned a different task. In the French Baccalauréat, choice is allowed in several subjects. In the United Kingdom, public examinations have traditionally been rich in options (less so today) as teachers regarded them as a bulwark against the examinations' dominance of the secondary-school curriculum (Newbould and Massey 1979).

A number of problems have been identified in relation to choice. First, allowing a choice in the selection of subjects, level of subject,

components of subjects, or in responding to questions may mean that examinees avoid important sections of the curriculum. Second, comparing examinee responses can be problematic as they are derived from different tasks. Third, options may not be marked to the same standard. Fourth, it has been argued that choice, in addition to measuring an examinee's proficiency, may also measure the examinee's ability to pick the option that will yield the highest marks. If this is indeed the case, choice will again introduce a source of construct-irrelevant variance (Kane 2006). These issues have implications for the validity of some interpretations, for reliability, for comparability of student responses (since they are derived from different tasks), and for the use of performance on examinations to select students for further education.

Arguments in favor of choice are strongest when the focus of an examination is on the assessment of examinees' ability to select and organize evidence and craft a coherent argument on a familiar topic. Choice, however, would be inappropriate if the purpose of the examination was to assess examinees' knowledge of facts or events, such as historical events or the ability to carry out basic operations in mathematics (Lane and Stone 2006).

Overpredictability of Examination Questions

The contents of examinations have to be predictable to some extent. An examination syllabus (and possibly exemplar examination papers) are usually available and will indicate to students and teachers what to expect (chapter 4). However, an examination that is predictable in detail so that, for example, a question on exactly the same topic appears in the same form in every examination will reinforce the practice of "teaching to the test," as well as limit examiners' scope to sample the curriculum domain. If, on the other hand, examination-setting practice dictates that a particular topic will not be featured if it was included in the previous year's examination, the domain will be similarly limited. Armenia's university selection examination provides a fairly extreme example of item predictability. The examination authority, in an effort to accommodate the large number of candidates who took the examination over a 22-day period, administered four different versions of the mathematics examination each day,

using items from a previously published item pool. Items were not repeated and, once used, were released, thereby giving a distinct advantage to candidates who took the examination during the latter days of the examination period (Bethell and Harutyunyan 2008). In Pakistan, it has been noted that the reoccurrence of questions in examinations has reinforced the tendency to focus on limited areas of the curriculum (Christie and Afzaldal 2005).

Taking Ancillary Competencies for Granted

An examination may take for granted the fact that all examinees possess certain competencies that are not the object of the assessment but that candidates require to respond to the examination. For example, it may be assumed in a history or mathematics examination that all examinees have a certain level of competence in ancillary areas, such as in the language of the examination or in reading ability. If, however, that is not the case, and examinees lack competence in reading, for instance, to an extent that it adversely affects their scores on the history or mathematics test, their competence in the ancillary areas (for example, reading) would be considered a source of construct-irrelevant variance. Construct-irrelevant variance occurs when one or more irrelevant constructs are being assessed in addition to the one that the examination is supposed to be measuring. A low level of competency in an ancillary skill for all examinees could be a source of systematic error or bias (Kane 2006).⁵

Inappropriate Conditions for Administration

Threats to validity arising from aspects of administration may be due to inappropriate conditions for assessment. This would be the case, for example, if the examination hall was too hot or too cold or if the supervisor had not carried out the required duties in a satisfactory manner, allowing too little time (creating negative bias), providing candidates with assistance, or allowing others to do so (leading to enhanced candidate performance). In the past, Romanian teachers, upon receiving a master examination copy, wrote the questions on the board; student ability to read the examination questions would have depended to some extent on the clarity of the teacher's writing, on the shape of the room, on distance from the board, and on vision acuity (Bethell 2003).

Candidate Characteristics

General Characteristics

Candidates may not be motivated to do well. They may, for example, think that they have little success because of lack of ability or because they did not receive additional tuition. If this is the case, performance cannot be taken as indicative of their knowledge and skills. Assessment anxiety, which has been defined as the antithesis of low motivation, will also provide a misleading picture of what an examinee might be able to do under less anxious conditions (Crooks, Kane, and Cohen 2008). Other characteristics that may result in an inappropriate inference about a student's ability include the student's health and ability to write quickly and neatly.

Use of Test-Taking Strategies

Test-taking strategies are usually associated with multiple-choice items, but they can also be used in the preparation of students for essay-type examinations. Their use can lead to a situation in which teaching can distort the inferences that can be made from an examination. This phenomenon is illustrated in a comparison of essays written by students in a state examination (which has since been discontinued) at the end of primary school in Ireland. Each year, students were given a choice of five essays, one of which required them to write about a day in their lives. Evidence provided by a school inspector suggests that students in at least one rural area were prepared by memorizing a series of stock sentences that could be used with a variety of prompts (box 7.1). The similarity of student opening responses to three different topics from the three different years is striking. In this instance, a high score on the examination could not be regarded as an indication of candidates' composition skills. Rather it was a measure of the ability to use memorized material to draft standard opening and closing paragraphs (Madaus and Greaney 1985). Performance could not be equated with achievement.

Scoring

Scoring Procedures

A number of factors associated with scoring can threaten the validity of score interpretation. These include lack of intra- and interrater

BOX 7.1**Opening Paragraphs of Student Essays****A Bicycle Ride (1946)**

I awakened early, jumped out of bed and had a quick breakfast. My friend, Mary Quant, was coming to our house at nine o'clock as we were going for a long bicycle ride together.

It was a lovely morning. White fleecy clouds floated in the clear blue sky and the sun was shining. As we cycled over Castlemore bridge we could hear the babble of the clear stream beneath us. Away to our right we could see the brilliant flowers in Mrs Casey's garden. Early summer roses grew all over the pergola which stood in the middle of the garden.

A Day in the Bog (1947)

I awakened early and jumped out of bed. I wanted to be ready at nine o'clock when my friend, Sadie, was coming to our house. Daddy said he would take us with him to the bog if the day was good.

It was a lovely morning. The sun was shining and white fleecy clouds floated in the clear blue sky. As we were going over Castlemore bridge in the horse and cart we could hear the babble of the clear stream beneath us. Away to our right we could see the brilliant flowers in Mrs. Casey's garden. Early summer roses grew all over the pergola which stood in the middle of the garden.

A Bus Tour (1948)

I awakened early and sprang out of bed. I wanted to be ready in good time for our bus tour from school. My friend, Nora Green, was going to call for me at half-past eight as the tour was starting at nine.

It was a lovely morning. The sun was shining and white fleecy clouds floated in the clear blue sky. As we drove over Castlemore bridge we could hear the babble of the clear stream beneath us. From the bus window we could see Mrs Casey's garden. Early summer roses grew all over the pergola which stood in the middle of the garden.

Source: Madaus 1988, 94.

consistency in the performance aspects they consider, the standards they set, or the marks they award (interrater reliability). Although rater agreement can be increased by restricting the areas of the domain that are assessed and by using more objective scoring criteria, doing these things can negatively affect the validity of the examination. The practice in some countries (such as India and Pakistan) of awarding “grace marks,” that is, additional marks to students near key cutoff decision points to get them to the next level, can also have a

negative impact on the validity of the examination. Scoring may also affect validity if it is so analytical that the richness of the performance is not captured or if it is so holistic that key areas of performance are ignored (Crooks, Kane, and Cohen 2008).

Aggregation of Scores from Individual Examination Tasks and Components

Examinations may be made up of many components, each of which will be allocated a mark loading for the component on a scale of achievement common to the examination as a whole. These loadings indicate the *intended* weightings of components. Thus, within a subject, a theory paper (component A) might be allocated 200 marks and a practical or oral paper or teachers' assessments of students (component B) might be allocated 100 marks on the basis that the theory is considered to be twice as important as the other component or, at any rate, that it merits twice the marks. When the marks obtained by students have been aggregated, the overall rank of examinees is determined, and the influence of the component on the rank order will be represented by its *achieved* weight. It might be assumed that the outcome will reflect the intended weighting. However, this may not be the case. Even when fewer marks are available for component B, adding them to the marks for component A, which has a greater number of marks, can dramatically change the order of merit of the examinees. The final rank will depend to a great extent on the *dispersion* of the marks for each component (Murphy 1982; Delap 1994; Millar, Kellaghan, and Mac Aogáin 2006).

Analysis of six Leaving Certificate Examinations in Ireland showed that there is a large reduction in variance in nonwritten components when compared with written components. Statistically significant differences were found between intended and achieved weights for all subjects.⁶ Furthermore, the weight achieved by written papers, with one exception, was greater than intended (Millar, Kellaghan, and Mac Aogáin 2006). Similar results were found for between-subject analysis when marks obtained on candidates' six best subjects were aggregated to provide a score used in selection to third-level education. If all subjects were accorded an equal weight, one would expect an intended weight of 0.167 for each subject. However, analysis of common combinations of subjects in which

intended and achieved weights were compared showed that the achieved weights varied from 0.127 to 0.205 (Mac Aogáin, Millar, and Kellaghan 2011). Thus, subjects did not contribute equally to the overall ranking used for selection.

The results of both within-subject and between-subject analyses point to a reduction in the validity of a public examination insofar as the weightings do not operate as specified. A candidate's ranking on a component or subject with a highly varied set of scores (large standard deviation) and a high correlation with other components will contribute more to the overall ranking than a component or subject with a less differentiated set of scores and lower correlation values. To address this issue in the case of within-subject components, raw scores on the components could be converted to a common scale, following which the intended weights could be applied to the standardized marks. The procedure would be acceptable if the same candidates had taken the components. It would not be appropriate if components or subjects were taken by different groups of candidates.

Impact

It is generally accepted that high-stakes examinations have an impact on educational processes and outcomes. The impact will be considered beneficial if examinations serve to guide teaching, directing attention to important topics and skills specified in the curriculum. The impact will not be considered beneficial if there is a focus on teaching test-taking strategies that subvert the intentions of the curriculum. In this case, the examination will distort teaching, which, in turn, will distort the inferences that can be made on the basis of candidates' performance on the examination. This was the view of Donald Campbell when he claimed, in what came to be known as Campbell's Law, that "the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor" (Campbell 1976, 49). In a specific reference to test scores, he pointed out that when such scores become the goal of the teaching process, "they both lose their value as indicators of educational status and distort the educational process in undesirable ways" (Campbell 1976, 52).

Improving Examination Validity

The recommendations presented in box 7.2 offer some suggestions for reviewing examination procedures, in the interest of improving the overall validity of public examinations.

BOX 7.2

Recommendations to Improve the Validity of Examinations

1. Decide whether separate examination systems (for example, academic, vocational, professional) or separate levels of examination are required to accommodate the varying abilities and needs of candidates, which increase as participation rates increase.
2. Ensure that certification, as well as selection, functions of an examination (if it serves both functions) are adequately reflected in the design of the examination by selecting items with difficulty levels appropriate for both purposes.
3. Construct a table of specifications (blueprint) with content areas listed on the horizontal axis and intellectual skills listed hierarchically on the vertical axis, to help ensure an adequate representation of the content and skills of the domain being assessed (see chapter 4).
4. Use a variety of methods to help average method effects. Multiple-choice items seem most appropriate for lower taxonomic levels of knowledge (factual information) but more difficult to construct for assessing higher levels. Essay-type questions tend to be more appropriate for assessing higher-order knowledge and skills (such as making inferences).
5. If an examination contains only multiple-choice items, include a sufficient number of items to provide an adequate sampling of the assessed domain.
6. Include competencies in areas of achievement that cannot be assessed in a paper-and-pencil examination (for example, oral fluency in language, construction of an object in woodwork or metalwork, ability to carry out an experiment in science).
7. Make some questions compulsory to ensure that candidates do not avoid answering questions on elements of the curriculum that are considered essential, even where a choice of questions is permitted. Divide the examination paper into sections, with a requirement that candidates respond to a question or questions in each section (based on the blueprint).
8. In a multilingual environment, ensure that the language of examinations does not create problems for particular groups of examinees.
9. When an individual examination is made up of a number of components, marks may have to be rescaled so that achieved weights reflect intended weights.

CONCLUSION

While lip service may be paid to the idea that validity (in particular, content and predictive validity) should be a key concern in a public examination, few examination authorities provide evidence relating to the issue. At the very least, examination authorities should be aware that certain common practices lead to incorrect inferences or interpretations being made from student examination scores. As these practices pose threats to validity, examination authorities should identify these threats in their own systems and take actions to mitigate them.

Authorities should also be sensitive to the consequences—in particular, possible negative consequences—of using the results obtained in examinations. Positive consequences may be said to occur if the examination fulfills the functions for which it was designed (that is, certification, selection, motivation, and the control of teaching and learning in schools). The most obvious of these is perhaps selection (for example, the extent to which the examination provides information that can be used to make adequate and equitable decisions in selecting students for further education). But consequences also relate to other facets. These may not all be positive: there is evidence that examinations to which high stakes are attached can have negative, if unintended, effects on teaching, on students' cognitive development, and on their motivation to achieve (see chapter 5).

NOTES

1. The difference between aptitude and achievement tests has become less clear than previously understood, partly in recognition of the fact that aptitude scores can be improved by tuition. In 2016, the Educational Testing Service replaced its widely used Scholastic Aptitude Test (SAT) with a new SAT (Scholastic Achievement Test), which contains more achievement test items than the earlier SATs (Grove 2018).
2. The term “paper-and-pencil instruments” refers to a general group of assessment tools that require students to read items or questions and respond in writing.

3. For example, Saudi Arabia's tertiary-level admission test assesses "reading comprehension, logical relations, problem-solving behavior, inferential abilities, inductive abilities" (Saudi Arabia National Center for Assessment, n.d., 4).
4. This occurs when an examination or test measures factors or variables that are not part of the overall construct (such as knowledge of history) that is being assessed. The variables are irrelevant to the construct the examination or test is supposed to be measuring.
5. Systematic error, in contrast to random error, is not determined by chance.
6. The formula for the achieved weight is based on the product of the standard deviation of the marks for a particular subject and the correlation between the marks for that subject and the total marks, divided by the standard deviation of the total marks (Adams and Murphy 1982).

REFERENCES

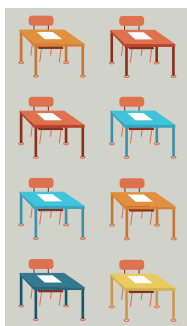
- Adams, R. M., and R. J. L. Murphy. 1982. "The Achieved Weights of Examination Components." *Educational Studies* 8: 15–22.
- AERA (American Educational Research Association), APA (American Psychological Association), and NCME (National Council on Measurement in Education). 2014. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Baird, J.-A., and D. Opposs. 2018. "The Standard Setting Project: Assessment Paradigms." In *Examination Standards: How Measures and Meanings Differ around the World*, edited by J.-A. Baird, T. Isaacs, D. Opposs, and L. Gray, 2–25. London: University College London, Institute of Education Press.
- Berliner, D. 2011. "Rational Responses to High-Stakes Testing: The Case of Curriculum Narrowing and the Harm That Follows." *Cambridge Journal of Education* 41: 287–302.
- Bethell, G. 2003. "Examination and Assessment Reform: An Overview of Experiences in Central and Eastern Europe." Education Support Program, Open Society Foundations, Budapest. https://www.opensocietyfoundations.org/sites/default/files/examinations_20030701.pdf.
- Bethell, G., and K. Harutyunyan. 2008. "Assessment and Examinations in Armenia." *Assessment in Education: Principles, Policy and Practice* 15: 107–19.
- Black, P. 1996. "Commentary." In *Examining the Examinations*, edited by E. D. Britton and S. A. Raizen, 19–21. Dordrecht, Netherlands: Kluwer.

- Cambridge Assessment. 2017. *The Cambridge Approach to Assessment*. Cambridge, UK: Cambridge Assessment. <http://www.cambridgeassessment.org.uk/Images/cambridge-approach-to-assessment.pdf>.
- Campbell, D. T. 1976. "Assessing the Impact of Social Change." Occasional Paper Series 8, Public Affairs Center, Dartmouth College, Hanover, NH.
- Campbell, D. T., and D. Fiske. 1959. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." *Psychological Bulletin* 56: 81–105.
- Choppin, B., and L. Orr. 1976. *Aptitude Testing at Eighteen-Plus*. Windsor, Berks, UK: National Foundation for Educational Research.
- Christie, T., and M. Afzad. 2005. "Rote Memorization of Secondary School Examination Achievement in Pakistan: An Empirical Investigation of a Widespread Assumption." Paper presented at the 31st Annual Conference of the International Association for Educational Assessment, Abuja, Nigeria.
- Cronbach, L. J. 1970. "Test Validation." In *Educational Measurement*, 2d ed., edited by R. L. Thorndike, 443–507. Washington, DC: American Council on Education.
- Crooks, T. J., M. Kane, and S. Cohen. 2008. "Threats to the Valid Use of Assessments." In *Student Assessment and Testing*, edited by W. Harlen, vol. 2, 150–71. London: Sage.
- Delap, M. R. 1994. "The Interpretation of Achieved Weights of Examination Components." *Statistician* 43: 505–11.
- European Parliament, Directorate-General for Internal Policies, Policy Department B. 2014. *Higher Education Entrance Qualifications and Exams in Europe: A Comparison*. Brussels: European Parliament.
- Greaney, V. 1980. "Developing National Measures of Attainment: A Study of Irish Experience." *International Review of Education* 2: 3–16.
- Greaney, V., A. Burke, and J. McCann. 1999. "Predictors of Performance in Primary-School Teaching." *Irish Journal of Education* 30: 22–37.
- Grove, A. 2018. "Differences between the SAT and ACT Exams." ThoughtCo.com. <https://www.thoughtco.com/differences-between-sat-and-act-exams-788714>.
- Kane, M. T. 2001. "So Much Remains the Same: Conception and Status of Validation in Setting Standards." In *Setting Performance Standards: Concepts, Methods, and Perspectives*, edited by G. C. Cizek, 53–88. Mahwah, NJ: Erlbaum.
- . 2006. "Validation." In *Educational Measurement*, 4th ed., edited by R. L. Brennan, 17–64. Westport, CT: American Council on Education and Praeger.
- Lane, S., and C. A. Stone. 2006. "Performance Assessment." In *Educational Measurement*, 4th ed., edited by R. L. Brennan, 387–431. Westport, CT: American Council on Education and Praeger.

- Lewin, K. 1997. "Test Development. Designing Tests and Presenting Results." In *Improving Test Design: Volume 1—Constructing Test Instruments, Analysing Results and Improving Assessment Quality in Primary Schools in Africa*, edited by V. Bude and K. Lewin, 35–55. Bonn: German Foundation for International Development, Education, Science and Documentation Centre.
- Lissitz, R. W. 2009a. *The Concept of Validity: Revisions, New Directions, and Applications*. Charlotte, NC: Information Age.
- . 2009b. "Introduction." In *The Concept of Validity: Revisions, New Directions, and Applications*, edited by R. W. Lissitz, 1–15. Charlotte, NC: Information Age.
- Mac Aogáin, E., D. Millar, and T. Kellaghan. 2011. "Transparency and Discriminatory Power of the Leaving Certificate Examination." *Irish Journal of Education* 39: 74–87.
- Madaus, G. F. 1988. "The Influence of Testing on the Curriculum." In *Critical Issues in Curriculum: Eighty-Seventh Yearbook of the National Society for the Study of Education*, part 1, edited by L. N. Tanner, 83–121. Chicago: University of Chicago Press.
- Madaus, G. F., and V. Greaney. 1985. "The Irish Experience in Competency Testing: Implications for American Education." *American Journal of Education* 93: 268–94.
- Messick, S. 1989. "Validity." In *Educational Measurement*, 3d ed., edited by R. L. Linn, 13–103. New York: American Council on Education and Macmillan.
- Millar, D., T. Kellaghan, and E. Mac Aogáin. 2006. "A Study of the Intended and Achieved Weights of Components in the Leaving Certificate Examination." Educational Research Centre, Dublin.
- Mislevy, R. J., M. R. Wilson, K. Ercikan, and H. Chudowsky. 2003. "Psychometric Principles in Student Assessment." In *International Handbook of Educational Evaluation*, edited by T. Kellaghan and D. L. Stufflebeam, 489–531. Dordrecht, Netherlands: Kluwer Academic.
- Murphy, R. J. L. 1982. "A Further Report of Investigations into the Reliability of Marking of GCE Examinations." *British Journal of Educational Psychology* 52: 58–63.
- Newbould, C. A., and A. J. Massey. 1979. "Comparability Using a Common Element." Occasional Publication 7, Test Development and Research Unit, Cambridge, UK.
- O'Rourke, B., M. Martin, and J. J. Hurley. 1989. "The Scholastic Aptitude Test as a Predictor of Third-Level Academic Performance." *Irish Journal of Education* 23: 22–39.
- Powell, J. L. 1973. *Selection for University in Scotland*. London: University of London Press.

- Saudi Arabia National Center for Assessment. n.d. *General Aptitude Test* (English Version). Riyadh: National Center for Assessment.
- Shiel, G., J. Cosgrove, N. Sofroniou, and A. Kelly. 2001. "Ready for Life? The Literacy Achievements of Irish 15-Year-Olds in an International Context." Summary report. Educational Research Centre, Dublin. <http://www.erc.ie/documents/p00summaryreport.pdf>.
- World Bank. 2018. *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC: World Bank.

CHAPTER 8



THE RELIABILITY OF PUBLIC EXAMINATIONS

INTRODUCTION

The issue of reliability arises from the fact that human performance (physical or mental) is variable. In the case of examinations or tests, reliability is concerned with estimating the extent to which an examinee's performance is consistent or inconsistent over a limited time span. When performance on a public examination is used for selection, two issues arise. First, the examination should effectively distinguish between candidates of different underlying levels of achievement. The second issue relates to replicability. Would the same subset of individuals be selected if they took a parallel form of the examinations, if they took the examination on another occasion, or if their examinations were graded by a different examiner (Cresswell 1995; Haertel 2006)?

In the case of psychometric tests, such as standardized achievement or intelligence tests, it is possible to repeat administration of an instrument to establish reliability. This is not possible in the case of public examinations. Some studies of the reliability of examinations have involved research studies in which procedures in the administration of examinations were copied with some modifications. However, the ecological validity of such studies, that is, the extent to which findings can be generalized to the "real world," may be questioned because

conditions in which they were carried out could have differed considerably from examination conditions. For example, in a research study reliability might be calculated on the basis of scores awarded independently by two examiners, although in the actual administration of an examination discrepancies between examination markers would probably have been resolved before the assignment of grades. An alternative approach to studying reliability involves generalizability theory, which uses analysis of variance to quantify the proportion of the variability of marks from different sources (Kim and Wilson 2009).

Early empirical studies of examination reliability (Edgeworth 1890; Hartog and Rhodes 1935; Starch and Elliott 1912, 1913; Valentine 1932) identified considerable disagreement between different scorings as a serious issue in the marking of essays.¹ In France, the findings of studies of the reliability of marking in the Baccalauréat in the 1930s reported similar findings (Zarrouati 2008). More recent studies have also recorded differences between examiners in their grading of essay-type examinations (for example, in English General Certificate of Education examinations and the Irish Leaving Certificate Examination) (Good and Creswell 1988; Hewitt 1967; Madaus and Macnamara 1970; Meadows and Billington 2005; Murphy 1978, 1982; Newton 1996; Tisi et al. 2013). Evidence presented to a British House of Commons enquiry revealed that less than 1 percent of grades were changed across General Certificate of Secondary Education examinations and A levels in 2016 (Turner 2017).

Factors contributing to unreliability can be related to examinees, examiners, and the subject being examined and how it is scored (Baird, Greator, and Bell 2004; Feldt and Brennan 1989; Strong 1995; Wiliam 1996). Some of these factors are common to all types of assessments, while others are more likely to arise in the case of free-response assessments, such as essay-type examinations. Such factors as examiner experience and the complexity of the marking process may interact to affect the accuracy of marking (Meadows and Billington 2005).

FACTORS ASSOCIATED WITH EXAMINEES

Three examinee-related factors can affect reliability. First, examinees may perform differently on different occasions for a variety of reasons

relating to health, motivation, concentration, lapse of memory, or carelessness. This leads to a situation that should be distinguished from one in which students consistently perform poorly under examination conditions (for example, an anxious student), introducing systematic error to the examination process. Second, fluctuations in external conditions can affect an examinee's performance. These may be physical (for example, the examination hall may be too hot) or they may be psychological (for example, an examinee may have experienced a traumatic event recently, such as the death of a friend or relative). Third, variation in the specific tasks required in the assessment may affect the examinee's performance. Because an examination consists of questions or tasks that represent only a sample from a larger domain, the specific tasks required might unintentionally favor one examinee on one occasion and not on another occasion. Thus, the results that are obtained will depend on an examinee's familiarity with the sample of tasks included in an examination.

FACTORS ASSOCIATED WITH EXAMINERS

Several factors associated with examiners that affect the reliability of scoring have been identified from the research literature (see box 8.1). These relate to examinations that require essay or open-ended responses, rather than tests that require the candidate to choose from a list of possible options.

FACTORS ASSOCIATED WITH THE SUBJECT AND HOW IT IS SCORED

A number of characteristics of the subject being examined, of the marking situation, and of activities surrounding marking affect the reliability of examinees' scores (Murphy 1978; Murphy 1982; Newton 1996; Tisi et al. 2013). First, reliability varies with the subject area being examined. Disagreement between examiners is likely to be greater when imagination and style are important qualities to be assessed than in a mathematics or science examination in which content will be more important.

BOX 8.1**Human Factors That Affect Marking**

General bias	Some examiners tend to give high scores (leniency) while others tend to give low scores (severity), introducing systematic bias. Inexperienced markers tend to be more severe. ^a
Scale of shrinkage	Shrinkage occurs when an examiner does not use the extreme categories or marks on a scale. This is a common issue in school-based assessments. In some cultures there is a general reluctance to opt for extreme values on rating scales. ^b
Inconsistency or random fluctuation	Inconsistency occurs when examiners score erratically or along different dimensions because of their different understandings and interpretations of scoring rubrics. This form of random fluctuation is separate from any scorer bias that may exist.
Halo effect	An examiner's impression from one characteristic of an examinee's response (for example, an essay) is generalized to the examination paper as whole and affects scores awarded on subsequent questions.
Stereotyping	An examiner may have a predetermined impression about a particular group that influences the examiner's judgment of individuals in the group.
Perception difference	A grading experience immediately before the examination influences an examiner's current grading judgment. An examinee's work tends to be rated more favorably when it follows work of a lower standard than when it precedes it. Conversely, poor quality work is assessed more severely when it follows work of high quality.
Rater drift	Individual examiners or groups of examiners tend to apply scoring criteria inconsistently over time. In one study, the unreliability associated with the marks assigned by an individual examiner to a set of papers on two different occasions was scarcely lower than that associated with the marks of two different examiners to this set (Macnamara and Madaus 1970).

Source: Zhang 2013.

a. The extent of severity can be estimated, if examination scripts have been randomly allocated to examiners, by subtracting the average of a set of marks assigned by a particular examiner from the average mark awarded by the entire group of examiners to the same questions or papers.

b. As noted in chapter 7, scores on examination components that are not differentiated (have relatively low standard deviations) will not contribute to the final rank order when combined with components with more differentiated scores.

Second, reliability is affected by the type of examination question and marking scheme. The highest levels of marker agreement are found in examinations made up of highly structured, analytically marked questions while the lowest levels are found in examinations that place most dependence on essay-type questions. When analytic marking is used, that is, when marking schemes are broken down to classify precisely why each mark is being awarded, reliability can be expected to be high. The highly detailed marking schemes that normally accompany mathematics papers help explain the high degree of reliability obtained in the marking of examinations covering this subject. In a language examination that requires essay-type responses, on the other hand, it is not always possible to specify precisely how each mark will be allocated.

Improved reliability may require additional training of markers to make them aware that more than one response may be correct. Sreekanth cites the case of a candidate's response to the question "why sangai deer do not get shelter during the rainy season." As the student's answer, "due to cutting and felling of the trees," did not exactly match the expected answer ("deforestation"), the response was not credited (Sreekanth 2016). In Pakistan, an examination board saw fit to advise its examiners not to penalize candidates for "out-of-textbook" correct answers (Greaney and Hasan 1998).

Third, reliability increases as the number of components in an examination increases. Finally, multiple marking of scripts improves reliability. However, increasing the number of markers beyond two may have little effect (Kim and Wilson 2009). In Slovenia, all essay-type questions were marked twice and the mark assigned a candidate was the mean of the two marks awarded (Gabršček 1999). All answers were also double-marked in Lithuania, with discrepancies being resolved by a senior marker (Bethell and Zabulionis 2012). Because of cost, time, and logistical constraints, multiple marking is normally only feasible for a sample of scripts. It may be noted that the combination of double or multiple marks to produce a final score is an acknowledgment that legitimate differences in opinion can exist between examiners and is fundamentally different from a system in which the most senior examiner's decision is all that counts (Tisi et al. 2013).

IMPROVING EXAMINATION MARKER RELIABILITY

The recommendations in box 8.2 offer pointers for improving marker reliability in public examinations.

CONCLUSION

The requirement for high reliability in examinations has had profound effects on assessment practice. The downside to standardized conditions of administration—which involves the use of detailed marking schemes and marker standardization in attempts to secure uniform criteria and standards—is that the kind of knowledge, skills, and understanding that can be assessed has been generally limited to lower cognitive levels (Cresswell 1995). Alternative approaches in Britain to address this issue by having teachers assess students' work using

BOX 8.2

Recommendations to Improve Marker Reliability

1. Specify the conditions of administration, such as tasks and procedures. Note that items that are objectively marked and that greatly restrain how candidates must respond are associated with greater reliability.
2. Provide opportunities for examiners to discuss marking standards in coordination or in critique meetings, during which sample answers may be used to illustrate standards.
3. Match marker characteristics to item types, for example, by using more experienced markers for items that are complex to mark.
4. Use item-level marking, so that more than one marker contributes to a candidate's overall mark, thereby reducing the effects of individual examiner idiosyncrasies.
5. Use onscreen marking. In this procedure, candidate scripts are scanned into digital format and sent to examiners for marking on screen (see chapter 4).
6. Use multiple marking (several markers) to produce a final mark for items that are difficult to mark (Pinot de Moira 2013; Tisi et al. 2013).
7. Carry out periodic studies of factors that may affect marker reliability.

nonstandard tasks, taken under nonstandard conditions, and supervised by one who knows the students well have run into some problems relating to bias, reliability, and comparability (Cresswell 1995).

Examinations with poor marker reliability can result in quite different scores for the same examination task. An examination board or agency that establishes that the marking of a paper is unreliable should consider whether it is appropriate or ethical to allow marks for this paper to contribute to the overall results of certification or selection examinations.

Examination users, such as employers and admissions officers, can reasonably expect that a national public examination will assess the key objectives of the curriculum. It can be quite difficult, however, to provide reliable measurement for some objectives, as in the case of a language or that of a candidate's ability to communicate in extended prose or display imagination and style (Newton 1996). The open-ended prose-based questions required to assess these abilities are the hardest to mark reliably. Although reliability will be improved by the use of more structured questions, this will probably be achieved at the expense of narrowing the definition of the attribute and, in the case of language, result in the exclusion of factors such as imagination, style, precision, and organization of thought. Portfolios or projects that cross conventional curriculum boundaries may broaden the range of attributes examined, but increase the tension between validity and reliability. If either is low, the information from the examination will be meaningless. What tends to happen in practice when high stakes are attached to performance on an examination is that the range of achievements measured is restricted through standardization to those that can be assessed reliably. However, this will probably result in a shrinkage of the universe of generalizations (validity) that can be made on the basis of performance on the examination (Gipps and Stobart 2003).

NOTE

1. Edgeworth, a widely published economist and statistician, observed in 1890, "I find the element of chance in these public examinations to be such that only a fraction from a third to two thirds of the successful

candidates can be regarded as safe, above the danger of coming out unsuccessfully if a different set of equally competent judges had happened to be appointed” (cited in Linacre 2011, 7).

Valentine (1932) identified several factors associated with lack of reliability that may give an appreciably different result (for example, a different examination paper of the same type, or the same answer paper marked by a different examiner). However, he does not seem to have appreciated the distinction between validity and reliability. For instance, as evidence of unreliability, he reported that his studies found no relationship between the order of merit in an entrance examination for secondary school and the order of merit at the end of secondary schooling.

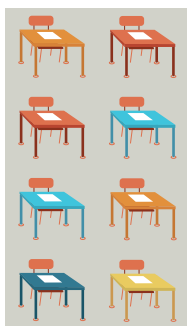
REFERENCES

- Baird, J.-A., J. Grotorex, and J. F. Bell. 2004. “What Makes Marking Reliable? Experiments with UK Examinations.” *Assessment in Education: Principles, Policy and Practice* 11: 331–48.
- Bethell, G., and A. Zabulionis. 2012. “The Evolution of High-Stakes Testing at the School–University Interface in the Former Republics of the USSR.” *Assessment in Education: Principles, Policy and Practice* 19: 7–25.
- Cresswell, M. J. 1995. “Technical and Educational Implications of Using Public Examinations for Selection to Higher Education.” In *Admission to Higher Education: Issues and Practice*, edited by T. Kellaghan, 38–48. Princeton, NJ: International Association for Educational Assessment and Dublin: Educational Research Centre.
- Edgeworth, F. 1890. “The Element of Chance in Competitive Examinations.” *Journal of the Royal Statistical Society*. Cited in J. M. Linacre, “Constructing Valid Performance Assessments: The View from the Shoulders of the Giants” (University of Michigan, June 2011; <https://www.rasch.org/memo85.pdf>).
- Feldt, L. S., and R. L. Brennan. 1989. “Reliability.” In *Educational Measurement*, 3d ed., edited by R. L. Linn, 105–46. New York: American Council on Education and Macmillan.
- Gabršček, S. 1999. “The Right to See and Know: Slovenian Experience.” Paper presented at the 25th Annual Conference of the International Association for Educational Assessment, Bled, Slovenia, May 23–28.
- Gipps, C., and G. Stobart. 2003. “Alternative Assessment.” In *International Handbook of Educational Evaluation: Part Two—Practice*, edited by T. Kellaghan and D. L. Stufflebeam, 549–75. Dordrecht, Netherlands: Kluwer Academic.

- Good, F. J., and M. J. Cresswell. 1988. "Grade Awarding Judgements in Differentiated Examinations." *British Educational Research Journal* 14: 263–81.
- Greaney, V., and P. Hasan. 1998. "Public Examinations in Pakistan: A System in Need of Reform." In *Education and the State: Fifty Years of Pakistan*, edited by P. Hoodbhoy, 136–76. Karachi, Pakistan: Oxford University Press.
- Haertel, E. 2006. "Reliability." In *Educational Measurement*, 4th ed., edited by R. L. Brennan, 65–100. Westport, CT: American Council on Education and Praeger.
- Hartog, P., and E. C. Rhodes. 1935. *An Examination of Examinations*. London: Macmillan.
- Hewitt, E. 1967. "Reliability of GCE O-Level Examinations in English Language." Joint Matriculation Board, Manchester, UK.
- Kim, S. C., and M. Wilson. 2009. "A Comparative Analysis of the Ratings in Performance Assessment Using Generalizability Theory and the Many-Facet Rasch Model." *Journal of Applied Measurement* 10: 408–23.
- Linacre, J. M. 2011. "Constructing Valid Performance Assessments: The View from the Shoulders of the Giants." University of Michigan. <https://www.rasch.org/memos.htm>.
- Macnamara, J., and G. F. Madaus. 1969. "Marker Reliability in the Irish Leaving Certificate." *Irish Journal of Education* 3: 5–21.
- Madaus, G. F., and J. Macnamara. 1970. *Public Examinations: A Study of the Irish Leaving Certificate*. Dublin: Educational Research Centre.
- Meadows, M., and L. Billington. 2005. "A Review of the Literature on Marking Reliability." Manchester, UK: Assessment and Qualifications Alliance.
- Murphy, R. J. L. 1978. "Reliability of Marking in Eight GCE Examinations." *British Journal of Educational Psychology* 48: 196–200.
- . 1982. "A Further Report of Investigations into the Reliability of Marking of GCE Examinations." *British Journal of Educational Psychology* 52: 58–63.
- Newton, P. E. 1996. "The Reliability of Marking of General Certificate of Secondary Education Scripts: Mathematics and English." *British Educational Research Journal* 22: 405–20.
- Pinot de Moira, A. 2013. *Features of a Levels-Based Mark Scheme and Their Effect on Marker Reliability*. Technical report. Manchester, UK: AQA Centre for Education Research and Policy.
- Sreekanth, Y. 2006. "Public Examinations: Means or Ends of Evaluation." Paper presented at the meeting of the Asia-Pacific Educational Research Association, Hong Kong SAR, China, November 28–30.

- Starch, D., and E. C. Elliott. 1912. "Reliability of Grading High School Work in English." *School Review* 20: 442–57.
- . 1913. "Reliability of Grading Work in Mathematics." *School Review* 21: 254–59.
- Strong, G. 1995. "A Survey of Issues and Item Writing in Language Testing." *Thought Currents in English Literature* 68: 281–312.
- Tisi, J., G. Whitehouse, S. Maughan, and N. Burdett. 2013. *A Review of the Literature on Marking Reliability Research*. Report for Ofqual. Slough, Berkshire, UK: National Foundation for Educational Research.
- Turner, M. 2017. "Oral Evidence: Integrity of Public Examinations, HC 345." Education Committee, House of Commons, November 28. <https://www.parliament.uk/business/committees/committees-a-z/commons-select/education-committee/inquiries/parliament-2017/inquiry2/>.
- Valentine, C. W. 1932. *The Reliability of Examinations: An Enquiry*. London: University of London Press.
- Wiliam, D. 1996. "Standards in Examinations: A Matter of Trust?" *Curriculum Journal* 7: 293–306.
- Zarrouati, M. 2008. "The Battle of the *Baccalauréat*: The Long Forgotten Story of a Divided Committee." In *An Atlantic Crossing: The Work of the International Examination Inquiry, Its Researchers, Methods and Influence*, edited by M. Lawn, 99–118. Oxford, UK: Symposium.
- Zhang, M. 2013. "Contrasting Automated and Human Scoring of Essays." *R and D Connections* 21. Educational Testing Service, Princeton, NJ.

CHAPTER 9



STANDARDS IN PUBLIC EXAMINATIONS

INTRODUCTION

Standards in education have become a major concern in recent years. In some cases, as in the United States, the concern reflects a desire to bring some coherence to a system that has no national curricula¹ and in which there is huge variation across the country, often associated with racial, ethnic, and socioeconomic groups (Barton 2009). More generally, across the world, there is concern that students are not being adequately equipped with twenty-first-century skills, variously described as including abilities related to collaboration and teamwork, creativity, critical thinking, and problem solving, as well as technological and socioemotional skills. The development of national standards can be seen as an attempt to set appropriate goals for students and then to provide a means for monitoring, explaining, and responding to students' progress. Standards have been set in many curriculum areas, with prominence being given to literacy and numeracy.

In the context of examinations, the term "standard" is used in two ways. Content standards describe the level of knowledge, skills, and understanding represented in an examination; standards of achievement or performance describe candidates' results based on examination performance.

This chapter deals with standards of achievement. The chapter first considers evidence relating to the comparability of grades in different subjects taken in the same examination or in the same subjects administered in different years or by different examination authorities. Second, the chapter considers change in standards over time, in particular the claim of grade inflation. Third, the chapter describes approaches examination authorities adopt to deal with the lack of comparability. Although the issue of grade comparability is an important one, it does not involve an examination of what a particular grade represents in terms of student achievement. That topic is addressed in the final section of the chapter.

COMPARABILITY OF GRADES IN DIFFERENT SUBJECTS OR EXAMINATIONS

The issue of standards is particularly relevant in considering the comparability of grades that a public examination agency or board might award to examinees in different subjects taken in the same examination or of grades in examinations in the same subject administered in different years or by different examination authorities. One might assume that grades in different examinations would be of comparable standards and that, for example, obtaining an A or a C in a language should be no easier or harder than obtaining the same grade in mathematics.

Similarly, it seems reasonable to assume that an examination in the same subject administered by different examination authorities could be considered comparable, as in the United Kingdom, where there are a number of examination boards; in Germany and Australia, where examination systems vary by state; and in South Asia, where Bangladesh, India, and Pakistan have multiple boards. However, to cite but one example, differences between German states in the percentages of candidates awarded varying grades in the Abitur and in the percentages qualifying for university entrance would suggest that standards vary from state to state (Hawkins, Gandal, and Britton 1996; Noah and Eckstein 1990). The issue acquires a particular significance when examination grades are treated as of equal value when selections for further education are made.

Comparing student grades in different subject areas remains a difficult problem. A number of studies of comparability between subjects, despite limitations in the procedures used, provide at the very least *prima facie* evidence that comparability issues exist in the grading of subjects. In one such study using *subject pairs analysis*, the mean grade of examinees, converted into numerical scores (A=100, B=90, and so on), was compared to the mean grade achieved by the same examinees on a comparison subject or group of comparison subjects. The reported differences in the mean scores (mean grade difference) were attributed to differences in grading standards (Nuttall, Backhouse, and Willmott 1974). Results may differ when analyses are computed for subgroups in an examination cohort. For example, a study of examination results in England found that mathematics and English were graded in a similar manner for male examinees, but for females English appeared to have been more leniently graded than mathematics (Newton 1997).

An alternative approach to examining the comparability of examination grades uses a measure of a student's *overall performance on the examination*, which is interpreted as an indication of "general scholastic ability." For example, grades might be converted to numerical scores and an overall index obtained for each examinee by summing scores for each of their six best subjects. It would then be possible to estimate the overall "ability" of examinees taking a particular subject by computing, for example, the percentage of examinees taking a subject that exceeded the median score on the overall performance scale. Such analyses revealed substantial differences between the "general scholastic ability" of candidates taking different subjects in the Irish Leaving Certificate Examination (Kellaghan and Millar 2003).

CHANGE IN STANDARDS OVER TIME

The public debate on the issue of examination standards changing over time tends to be driven by perception, which is often based on anecdotal evidence. From time to time, the media feature headlines such as "Standard of education is 'dropping' say employers," based on the views of business leaders, even when there is evidence to the contrary (Gilbert 2008). National assessments, such as the US National

Assessment of Educational Progress carried out over a number of decades and featuring linked or common items, can help address this issue in the case of reading and mathematics (US National Center for Education Statistics 2013). Comparison of standards based on changes in examination results is more problematic because of, among other things, changes in curricula and question-and-answer formats and in emerging market economies, to changes in school access, retention rates, and candidate characteristics.

Many commentators regard stability over time in examination performance standards a matter of concern. Would a grade A be awarded now for work that would previously have been awarded a grade B? The issue is important when candidates whose qualifications have been obtained in examinations taken in different years are in competition for a higher education place or for a job. It may also be of societal importance insofar as examination performance is accepted as an indicator of standards in the education system.

The increase over time of the proportion of high grades awarded in examinations in a number of countries (such as England, Germany, and Ireland) has been interpreted as indicative either of improving standards (as a result of students working harder or improved teaching) or of examinations getting easier. However, comparing performance on public examinations over time, in an attempt to resolve this issue, is generally regarded as problematic, if not impossible, given the variety of factors depicted in box 9.1 that could affect student achievement and the awarding of grades (Cambridge Assessment 2010a, 2010b).

In an effort to control the influence of at least some of these factors, Tymms, Coe, and Merrell (2005) carried out a study of General Certificate of Secondary Education (GCSE) and A-level examinations in England. For both examinations, the performance of sample groups of public examination candidates was matched with their performance on an ability or aptitude test to determine (a) how candidates of the same “ability” performed on GCSE examinations from 1997 to 2006 and (b) how candidates of the same “ability” performed on A-level examinations from 1988 to 2004.² For GCSE subjects as a whole, students of comparable ability achieved the same grades regardless of the year in which they took their examination. Science and mathematics candidates, however, delivered higher grades in

BOX 9.1**Factors That Might Affect Student Examination Grades over Time**

- Key characteristics of candidates changing
- The number of students staying at school increasing
- The content and structure of syllabi being modified
- Students working harder or getting “smarter” about preparing for examinations
- Teaching improving
- Examination questions differing from year to year
- External circumstances, such as civil and political conditions, being altered
- Awarding bodies expanding the range of ways candidates can demonstrate their knowledge

Sources: Cambridge Assessment 2010a, 2010b.

2004 than in 1997. At A-level over the period in question, the grades awarded students of the same level of ability significantly increased in all six subjects (biology, English literature, French, geography, history, and mathematics). The increase of about three grades was particularly high in mathematics. The results were interpreted as providing evidence of falling standards, that is, examinations were being graded more leniently, with the result that more candidates achieved higher grades each year.

A problem with the design of this study is that no single reference measure, such as an ability or aptitude test, can allow appropriately for achievement in every subject. There will always remain an unexplained unique contribution within each subject that will vary depending on the content of the reference test. Furthermore, the results of reference-measure-type analyses, like those from subject pairs, are wholly dependent on the particular population of students sampled (Goldstein and Cresswell 1996).

A 2018 report on standard setting in 12 different public examination systems highlights the limitations of using examination results

to monitor changes in achievement levels. The study, which was carried out by examination officials (“examination insiders”), noted that eight of the systems aimed to maintain achievement standards constant over time (Baird et al. 2018). Cut scores marking difference in grade levels were adjusted to ensure that the percentages being awarded each subject grade remained similar to those awarded the previous year; in one other system the marking system was revised to help ensure comparability of achievement standards (McManus 2018).

At a seminar organized by Cambridge Assessment, at which examination experts discussed the issue of examination standards, it was concluded that while the issue of “standards over time” often dominates the public debate, discourse relating to the topic struggles to tell us anything. Comparing standards over time was considered problematic because examinations change over time as priorities change, technology changes, and knowledge changes. At the seminar, Roger Murphy concluded that “our public examination system is not the best way to address standards over time, and this is not its prime purpose” (Cambridge Assessment 2010a, 4).

DEALING WITH COMPARABILITY PROBLEMS

In a study of 30 examination systems, Opposs (2015) reports on a variety of procedures to address intersubject comparability for 8 systems. The procedures involved either the conversion of raw scores (original marks) to scale scores or the adjustment of examinees’ scores to reflect a measure of their overall scholastic achievement.

Converting raw scores to scale scores involves the calculation of z-scores³ for all subjects (in Cyprus) or the use of item response modeling⁴ to convert raw to scale scores (in Taiwan, China). A number of problems are associated with the approach. First, it takes the focus off the knowledge and skills that are being assessed at a time when there is great emphasis on standards-based reform in education and on attempting to specify the knowledge and skills students should acquire. Second, this approach will not register changes in standards over time. Third, when marks are combined, for example, to make a decision regarding selection, no account is taken of the differential

requirements of subjects (such as the difficulty levels of different subjects). Some may require greater student effort and time and may be more highly selective (such as higher-level mathematics) than others. It is easier to justify the procedure when candidates take examinations in the same subjects (as in Singapore; Taiwan; China; and Tasmania in Australia). Finally, if students are allowed to choose the subjects in which they are examined, they may choose ones that have a historical record of being associated with high marks and avoid ones that have traditionally resulted in low marks.

In a number of jurisdictions examinees' performance on individual subjects is adjusted on the basis of their performance on the examination as a whole. For example, Hong Kong SAR, China, calculates a "group ability index" on the basis of a candidature's results in four core subjects (HKEAA 2011). In Australia, grades awarded to candidates in a subject are compared with performance on all other subjects in order to estimate the difficulty of a subject, following which grades are adjusted (McGaw, Gipps, and Godber 2004).

A number of difficulties associated with the use of a measure of general academic ability to adjust candidates' grades in individual subjects can be identified. First, the assumption underlying the approach that a general factor underlies a variety of achievements and is equally appropriate for all would be difficult to sustain (Goldstein and Cresswell 1996). Second, subjects that contribute to the overall score will vary from candidate to candidate, unless all take the same examinations.

The Opposs (2015) study finds no evidence that most jurisdictions implemented procedures to improve intersubject comparability. However, although explicit procedures may not have been reported, and may not even have existed, it is possible that grading procedures in individual subjects were based on implicit beliefs about the nature of candidates. Such a conclusion seems warranted on the basis of an analysis of grading practices in the Leaving Certificate Examination in Ireland (one of the countries that did not adjust grades in the Opposs study). In this study, grades awarded in individual subjects reflected the overall ability of candidates based on an index of overall student performance on an earlier Junior Certificate public examination (Kellaghan and Millar 2003).⁵

Statistical adjusting of student grades to deal with intersubject comparability can be puzzling for students. For example, students may be surprised when they see candidates with the same raw score receiving different scale scores because they had taken different subjects. Critical comments are not limited to students, but also may emanate from other stakeholders, including members of the public (as reported in Australia, Cyprus, and Fiji) and the media, as well as from professional and academic sources (Opposs 2015).

DEFINING PERFORMANCE STANDARDS

The previous discussion of grade comparability does not address the important issue of what knowledge, skills, or achievements are represented by students' performance. The general description of the content of a subject area will usually have been specified in a curriculum document or syllabus, though this may require further specification when used to construct an examination. An additional task in standard setting is to match threshold scores or cut scores (that is, minimum scores associated with individual grades or levels) on a total mark scale with associated knowledge and skills. It can be argued that examination systems do this to some extent in their categorization of examinee performance into grades or levels. However, the terms used are not very different from the terms we use in everyday language (see box 9.2). They do not make explicit what a student knows, understands, or can do.

Examination boards and agencies around the world use various methods to set standards (Cizek 2012; Hambleton 2001; Loomis and Bourque 2001; Opposs and Gorgen 2018; Raymond and Reid 2001). Major steps include

- Identifying and selecting experienced panelists.
- Choosing between a standard-setting method that may be test centered (based on a review of assessment material and scoring rubrics) or examinee centered (in which judgments are made about examinees' actual work) or a combination of the two.
- Reviewing and revising judgments made by panelists. Statistical methods, including the use of item response theory, have been

BOX 9.2**Selected Countries: Academic Grades in Examinations**

Cambodia ^a A Excellent B Very good C Good D Satisfactory E Limited achievement, fair F Fail	Denmark ^b 12 Excellent 10 Very good 7 Good 4 Fair 2 Meets minimum requirements 0 Does not meet minimum requirements
France ^c 16–20 Pass with distinction 14–15.9 Pass with merit 12–13.9 Good pass 10–11.9 Pass 0–9.9 Fail	India ^d 60–100% First Division 45–59% Second Division 33–44% Third/Pass 0–32% Fail
Anglophone West African Countries WASSCE ^e A1 75–100% Excellent B2 70–74% Very Good B3 65–69% Good C4 60–64% Credit C5 55–59% Credit C6 50–54% Credit D7 45–49% Pass E8 40–45 % Pass F9 0–44% Failure	Zambia ^f 1–2 Distinction 3–4 Merit 5–6 Credit 7–8 Satisfactory 9 Unsatisfactory

Note: WASSCE = West African Senior School Certificate Examination.

a. For data on Cambodia, see <https://shelbycearley.files.wordpress.com/2010/06/education-in-indo-china.pdf>.

b. For data on Denmark, see https://en.wikipedia.org/wiki/Academic_grading_in_Denmark.

c. For data on France, see <https://about-france.com/primary-secondary-schools.htm>.

d. For data on India, see <http://wenr.wes.org/2012/05/wenr-may-2012-converting-secondary-grades-from-india> (applies in many states).

e. For data on Anglophone West African countries, see <https://waecdirect.blogspot.ie/2017/02/waecdirect-grading-system.html>.

f. For data on Zambia, see <https://www.scholaro.com/pro/Countries/zambia/Grading-System>.

used in some instances. In addition, those charged with determining cutoff points or ratings may consider the consequences of panelists' decisions (for example, if the ratings resulted in a large number of "failing" students, questions might be raised about the extent to which procedures might have been flawed or unrealistic).

There have been numerous efforts in British public examinations to identify the qualities of candidates' work appropriate to different grade levels, and many descriptions of specific standard-setting grade criteria are available (for example, Christie and Forrest 1981; Opposs and Gorgen 2018; Orr and Nuttall 1983; SEC 1985). However, serious difficulties were encountered in their application. For example, some examination performances that would have merited a particular grade using conventional procedures would not have merited the same grade using the grade-related criteria, although examiners agreed that they should. It seems that the criteria could not accommodate the multidimensional nature of achievement by, for example, specifying the weight that should be attached to different aspects of performance when judging an individual examinee's work. Neither could the application of concise sets of explicit written criteria replicate holistic value judgments, that is, where an entire written response rather than individual elements is evaluated as a whole by qualified judges (Cresswell 1996).

These problems have not deterred authorities in other jurisdictions from attempting to establish standard-setting grade criteria. For example, examination reform in New South Wales, Australia, has shown a preference for standard-based (criterion-referenced) assessment to reinforce a broad revised school curriculum (K–10 in English, mathematics, science, history, and geography). The Higher School Certificate Examination offers syllabi that set expectations of what students must learn and measures student performance against those standards. A student's mark in each course is reported against descriptive performance bands that show what the student knows, understands, and can do.⁶ Essential elements of the achievement standards are (a) a summary of knowledge, skills, and understanding typically demonstrated by students at a given standard; (b) the tasks candidates were given in an examination; and (c) samples of responses exemplifying each standard (Bennett 2009).

Even where there is strong administrative support for modifying existing standard-setting grade criteria, examination boards may encounter opposition to efforts to change the status quo. The review of standard-setting practices in 12 established public examination systems, referred to earlier, concluded that examinations have been

operationalized differently in their respective systems; they tend to be culturally bound and deeply embedded within individual educational systems (Baird et al. 2018). Bringing about changes in current standard-setting practices is likely to prove challenging because of teacher and student expectations, as well as political, financial, and time constraints. Factors that could prompt the desired changes include strong political leadership, support from the media, crises in the examination system, and major cultural and contextual changes (Isaacs and Gorgen 2018), as experienced in Chile (Osses and Varas 2018), Georgia (Andguladze and Mindadze 2018), and South Africa (Sibanda 2018).

PUBLIC CONFIDENCE

Some commentators regard the notion of comparability across subjects as meaningless when one considers that courses on which examinations are based differ in difficulty, demands they make on students, and their motivational effects. Furthermore, students opting for different subject areas differ in their characteristics (Cresswell 1996; Goldstein 1986; Kellaghan and Millar 2003; Newbould and Massey 1979; Newton 1997; Wiliam 1996). In light of these considerations, it has been suggested that what is really important is whether or not grades are accorded the same value by certificate users. The issue then becomes one of maintaining public confidence in the use of an examination for selection purposes (Cresswell 1995), which will depend in great part on the extent to which the public considers the examination process fair in terms of its design, administration, and scoring, as well as devoid of cheating or other forms of malpractice.

The situation today would seem to be that it is worthwhile having general, but not specific, descriptions of the achievement worthy of a grade. This view recognizes the subjective nature of the judgments involved, which have been compared to those made in evaluation of a work of art. Acceptance of this view may be a cause for concern to some, given the decisions that are based on results. However, while the judgments or decisions of examiners may not be amenable to empirical verification, even though they may be supported by empirical data,

this does not mean that they are capricious or unreliable in the sense that they are difficult to replicate. On the contrary, they can be considered the results of a rational process, which can be supported by reason (Cresswell 1996, 2000; Goldstein and Cresswell 1996). Contributing to this procedure are statistical data on examinees' performance, the experience and skills of examiners in categorizing responses appropriate to different grade levels and setting grade boundaries, and procedures to review judgments and achieve consensus.

CONCLUSION

Educational standards related to examinations include both content and achievement standards. They facilitate the setting of learning goals, assessment of student learning, and reporting of student achievement. Content standards summarize the knowledge, skills, and understandings expected of examination candidates. Examination achievement standards, on the other hand, describe how the candidate performed on the examination in summary terms (such as Grade A, B, . . . or Excellent, Very Good, . . .).

Although members of the public and policy makers may use examination results to monitor changes in achievement levels over time it is inadvisable to do so, as many factors affect examination results, some of which are unrelated to a student's true achievement level. Few examinations, for instance, contain link items that would support comparing levels of student examination performance from year to year.

Approaches toward standard setting differ considerably around the world. Many systems, for a variety of historical and political reasons, appear to be loathe to alter their current approaches. This reluctance may also be reflected in the tendency to select cutoff scores in each subject area that ensure that grade distributions match those of previous years.

Researchers with some public and international examination agencies continue to focus on the problem of intersubject comparability in grading examinations. They are likely to build on recent efforts in defining, categorizing, and making explicit the knowledge, skills, or

achievements (performance standards) represented by students' performance in the examination. These efforts may be regarded as worthwhile if they add to the transparency of examinations and if this, in turn, has a positive impact on student learning.

NOTES

1. The Advanced Placement tests might be considered an exception. These tests are based on a national curriculum developed by the College Board and offer a specific population of high-achieving final year high school students an opportunity to demonstrate their readiness for college (Baker 2018; Morgan 2018).
2. The ability and aptitude tests had been administered as part of a system a school could join to obtain information that would allow the school to compare the progress of its students with that of students in other schools.
3. Indicates the number of standard deviations a particular examination score is from the mean or average.
4. Item response modeling is a statistical approach that is widely used (including in national and international assessments) in designing, analyzing, scoring, and comparing tests. It supports placing marks on a common scale. See Crocker and Algina 2006 and Shiel and Cartwright 2015.
5. The research team obtained results of Junior Certificate Examinations taken three years earlier and of Leaving Certificate Examinations. Within each subject the "academic ability" of students as measured by the Junior Certificate Examination increased as the Leaving Certificate Examination grade increased (David Millar, personal communication, December 12, 2018).
6. For information on aligning marks and standards in New South Wales, see <http://educationstandards.nsw.edu.au/wps/portal/nesa/11-12/hsc/about-HSC/moderation>.

REFERENCES

- Andguladze, N., and I. Mindadze. 2018. "Standard Setting in Georgia: The Unified National Examinations." In *Examination Standards: How Measures and Meanings Differ around the World*, edited by J.-A. Baird, T. Isaacs, D. Opposs, and L. Gray, 133–51. London: University College London Institute of Education Press.

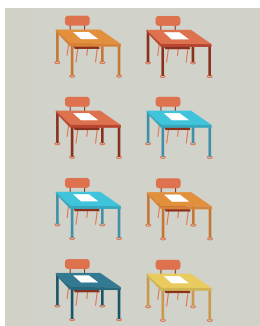
- Baird, J.-A., T. Isaacs, D. Opposs, and L. Gray, eds. 2018. *Examination Standards: How Measures and Meanings Differ around the World*. London: University College London Institute of Education Press.
- Baker, E. L. 2018. "Context and Change in Standard Setting." In *Examination Standards: How Measures and Meanings Differ around the World*, edited by J.-A. Baird, T. Isaacs, D. Opposs, and L. Gray, 279–80. London: University College London Institute of Education Press.
- Barton, P. 2009. *National Education Standards: Getting beneath the Surface*. Princeton, NJ: Educational Testing Service.
- Bennett, J. 2009. "Country Case Study: Australia (New South Wales)." In *Secondary School External Examination Systems*, edited by B. Vlaardingerbroek and N. Taylor, 55–78. Amherst, NY: Cambria.
- Cambridge Assessment. 2010a. *Exam Standards: The Big Debate—Report and Recommendations*. Cambridge, UK: Cambridge Assessment.
- . 2010b. *School Exams: Have Standards Really Fallen? Report of a Seminar*. Cambridge, UK: Cambridge Assessment.
- Christie, T., and G. M. Forrest. 1981. *Defining Public Examination Standards*. London: Schools Council.
- Cizek, G. J., ed. 2012. *Setting Performance Standards: Foundations, Methods, and Innovations*. 2d ed. New York: Routledge.
- Cresswell, M. J. 1995. "Technical and Educational Implications of Using Public Examinations for Selection to Higher Education." In *Admission to Higher Education: Issues and Practice*, edited by T. Kellaghan, 38–48. Princeton, NJ: International Association for Educational Assessment and Dublin: Educational Research Centre.
- . 1996. "Defining, Setting and Maintaining Standards in Curriculum-Embedded Examinations: Judgemental and Statistical Approaches." In *Assessment: Problems, Developments and Statistical Issues*, edited by H. Goldstein and T. Lewis, 57–84. Chichester, UK: Wiley.
- . 2000. "The Role of Public Examinations in Defining and Monitoring Standards." In *Educational Standards* (Proceedings of the British Academy 102), edited by H. Goldstein and A. Heath, 69–120. Oxford, UK: Oxford University Press.
- Crocker, L., and J. Algina. 2006. *Introduction to Classical and Modern Test Theory*. Pacific Grove, CA: Wadsworth.
- Gilbert, N. 2008. "Standard of Education Is 'Dropping' Say Employers: Company Directors Call into Question Government Figures for Education Standards and Report Poor Numeracy and Literacy in New Employees." *Guardian*, August 11. <https://www.theguardian.com/education/2008/aug/11/falling.standards>.

- Goldstein, H. 1986. "Models for Equating Test Scores and for Studying the Comparability of Public Examinations." In *Assessing Educational Achievement*, edited by D. L. Nuttall, 168–84. London: Falmer.
- Goldstein H., and M. J. Cresswell. 1996. "The Comparability of Different Subjects in Public Examinations: A Theoretical and Practical Critique." *Oxford Review of Education* 22: 435–42.
- Hambleton, R. K. 2001. "Setting Performance Standards on Educational Assessment and Criteria for Evaluating the Process." In *Setting Performance Standards: Concepts, Methods and Perspectives*, edited by G. J. Cizek, 89–116. Mahwah, NJ: Erlbaum.
- Hawkins, S., M. Gandal, and E. Britton. 1996. "Examination Systems in Seven Countries." In *Examining the Examinations*, edited by E. D. Britton and S. A. Raizen, 219–49. Dordrecht, Netherlands: Kluwer Academic.
- HKEAA (Hong Kong Examinations and Assessment Authority). 2011. "Grading Procedures and Standards-Referenced Reporting in the HKDSE Examination." HKEAA, Hong Kong SAR, China. http://www.hkeaa.edu.hk/DocLibrary/Media/Leaflets/HKDSE_SRR_A4_Booklet_Jun2011.pdf.
- Isaacs, T., and K. Gorgen. 2018. "Culture, Context and Controversy in Setting National Examination Standards." In *Examination Standards: How Measures and Meanings Differ around the World*, edited by J.-A. Baird, T. Isaacs, D. Opposs, and L. Gray, 307–30. London: University College London Institute of Education Press.
- Kellaghan, T., and D. Millar. 2003. "Grading in the Leaving Certificate Examination: A Discussion Paper." Educational Research Centre, Dublin.
- Loomis, S. C., and M. L. Bourque. 2001. "From Tradition to Innovation: Standard Setting on the National Assessment of Educational Progress." In *Setting Performance Standards: Concepts, Methods, and Perspectives*, edited by G. J. Cizek, 175–217. Mahwah, NJ: Erlbaum.
- McGaw, B., C. Gipps, and R. Godber. 2004. "Examination Standards: Report of the Independent Committee to QCA." London: UK Government, Qualifications and Curriculum Authority.
- McManus, H. 2018. "Standard Setting in Ireland: The Leaving Certificate." In *Examination Standards: How Measures and Meanings Differ around the World*, edited by J.-A. Baird, T. Isaacs, D. Opposs, and L. Gray, 157–77. London: University College London Institute of Education Press.
- Morgan, D. L. 2018. "Setting Standards in the United States: The Advanced Placement Programme." In *Examination Standards: How Measures and Meanings Differ around the World*, edited by J.-A. Baird, T. Isaacs, D. Opposs, and L. Gray, 257–78. London: University College London Institute of Education Press.

- Newbould, C. A., and A. J. Massey. 1979. "Comparability Using a Common Element." Occasional Publication 7, Test Development and Research Unit, Cambridge, UK.
- Newton, P. E. 1997. "Measuring Comparability of Standards between Subjects: Why Our Statistical Techniques Do Not Make the Grade." *British Educational Research Journal* 2: 433–49.
- Noah, H. J., and M. A. Eckstein. 1990. "Trade-Offs in Examination Policies: An International Perspective." In *Changing Educational Assessment: International Perspectives and Trends*, edited by P. Broadfoot, R. Murphy, and H. Torrance, 84–97. London: Routledge.
- Nuttall, D. L., J. K. Backhouse, and A. S. Willmott. 1974. "Comparability of Standards between Subjects." Schools Council Examinations Bulletin 29, Evans/Methuen Educational, London.
- Opposs, D. 2015. "Is It Valid to Treat Assessment Grades from Different Subjects the Same?" Paper presented at the 41st Annual Conference of the International Association for Educational Assessment, Lawrence, Kansas, October 11–15. www.iaea.info/documents/paper_3fc72a4a9.pdf.
- Opposs, D., and K. Gorgen. 2018. "What Is Standard Setting?" In *Examination Standards: How Measures and Meanings Differ around the World*, edited by J.-A. Baird, T. Isaacs, D. Opposs, and L. Gray, 54–76. London: University College London Institute of Education Press.
- Orr, L., and D. L. Nuttall. 1983. "Determining Standards in the Proposed Single System of Examinations at 16+." Schools Council, London.
- Osses, A., and M. L. Varas. 2018. "Standard Setting in Chile: The Preuba de Selección Universitaria." In *Examination Standards: How Measures and Meanings Differ around the World*, edited by J.-A. Baird, T. Isaacs, D. Opposs, and L. Gray, 78–95. London: University College London Institute of Education Press.
- Raymond, M. R., and J. B. Reid. 2001. "Who Made Thee a Judge? Selecting and Training Participants for Standard Setting." In *Setting Performance Standards: Concepts, Methods, and Perspectives*, edited by G. J. Cizek, 119–57. Mahwah, NJ: Erlbaum.
- SEC (Secondary Examinations Council). 1985. "Reports of the Grade-Related Criteria Working Parties." SEC, London.
- Shiel, G., and F. Cartwright. 2015. *Analyzing Data from a National Assessment of Educational Achievement*, vol. 4 of National Assessments of Educational Achievement. Washington, DC: World Bank.
- Sibanda, E. 2018. "Standard Setting in South Africa: The National Senior Certificate." In *Examination Standards: How Measures and Meanings Differ around the World*, edited by J.-A. Baird, T. Isaacs, D. Opposs, and L. Gray, 212–27. London: University College London Institute of Education Press.

- Tymms, P., R. Coe, and C. Merrell. 2005. "Standards in English Schools: Changes since 1997 and the Impact of Government Policies and Initiatives—A Report for The Sunday Times." CEM Centre, University of Durham, Durham, UK.
- US National Center for Education Statistics. 2013. *The Nation's Report Card: Trends in Academic Progress 2012*. Washington, DC: National Center for Education Statistics.
- Wiliam, D. 1996. "Standards in Examinations: A Matter of Trust?" *Curriculum Journal* 7: 293–306.

CHAPTER 10



EQUITY ISSUES IN PUBLIC EXAMINATIONS

INTRODUCTION

The concepts of equity and meritocracy are closely tied to the institution of examinations. In Imperial China, a prime consideration was that the “all-important business of government must not be left to the accidents of either birth or wealth” (Hu 1984, 7). This idea was maintained when examinations were introduced to European and other countries. Following the French Revolution, the principle of equality implied that all citizens would be in a position to compete for qualifications, employment, and wealth, and that advancement would be based on demonstrated achievement in school and in examinations rather than on the social position of one’s family (Eckstein and Noah 1993).

Similar motivation is evident in Britain and in India, where examinations were introduced to replace practices of patronage and of buying positions in the civil service and armed forces (Montgomery 1965). In Japan also, the competitive examination system that was introduced in the latter half of the nineteenth century was designed to open educational opportunities to the whole (male) population (Amano 1990). However, the reality may not always have matched the aspiration. Examinations were,

for example, perceived by some commentators to defend privilege and impede the development of “popular” education at a time when secondary education was expanding (Judd 1935). Questions were also raised about possible bias in examinations, which may have limited the chances of particular groups (boys, girls, ethnic groups, language groups, urban vs. rural candidates, public vs. private school candidates) (see Mathews 1985).

This chapter considers evidence relating to inequalities in the examination success of identifiable groups. In particular, it asks whether there is evidence that the chances of success are loaded unfairly in favor of or against students of a particular gender, socio-economic background, place of residence, type of school attended, or ethnic or language group membership. The most frequently used statistical procedure to address this issue (originally called “item bias”) is differential item functioning (DIF) analysis. Item bias occurs when individuals from different major subgroups with the same latent trait (ability), as measured by the examination or test, have a different probability of giving a certain response to items in an examination or test (Camili 2006; Cartwright 2015; Osterlind and Everson 2009). If the examination authority is interested in, for instance, the possibility or the extent of gender bias in a particular question, the results of a DIF analysis would indicate whether the question is biased in favor of boys or girls *after taking into account the overall examination score gender difference*. There are two main approaches to the detection of DIF. One is based on item response theory; the other is non-parametric and uses the chi square statistic (Mantel-Haenszel odds ratio). Different analytical methods have been shown to produce relatively consistent but not identical results (Gierl, Khaliq, and Boughton 1999).

It is important to bear in mind that the fact that a difference in performance is detected does not tell us whether it can be attributed to characteristics of the examination. It could reflect a real difference in achievement. As DIF analyses are susceptible to sampling error (Cartwright 2015), the results should be interpreted with caution; they provide a good basis for discussion and can be considered a necessary but not sufficient condition to conclude that bias exists. Other forms of investigation are required to determine that. One such form would involve a review of examination

items by subject-matter specialists and members of the subgroups represented in the analysis. Another form would compare the predictive validity of the examinations for each of the subgroups of examinees.

GENDER

Early public examinations in China and Great Britain were not open to females. Perhaps there would not have been much point in females taking the examinations, since most of the occupations on which examination performance depended were not available to women. The situation, however, was changing in the mid-nineteenth century, and the Oxford and Cambridge “Locals” were opened to females in 1865 (Roach 1971). In France, although the Baccalauréat was awarded to a female in 1861, it was not until 1924 that the examination was fully open to girls (Cros 2009). Opening examinations to females led to debate over whether there should be separate systems of examinations for male and female students. The *Global Monitoring Report* (UNESCO 2017) indicates that gender parity has been achieved in the world as a whole at all educational levels with the exception of tertiary. However, disparity at the expense of girls is still pronounced in low-income countries, where economic factors and cultural bias serve to restrict female access to education and reduce their likelihood of taking public examinations; rural location, religion, ethnicity, age, disability, and socioeconomic status have tended to affect female participation in education more than male (World Bank 2018).

Findings on differences in performance on examinations relating to gender are not consistent. In some countries, boys, on average, perform better than girls; in others, girls do better, but little or no difference has been recorded in some countries. In general, although the finding is not universal, girls tend to do better in language and literary subjects, and boys in science and technical subjects. Differences, when found, can be persistent. For example, in Malaysia (Ahmad et al. 2017; Government of Malaysia 2015) and Oman, girls (see photo 10.1) perform better than boys on examinations at every level (Oman Ministry of Education and World Bank 2012). Indeed, the gap in

PHOTO 10.1

Oman: Female Examination Candidates



Photograph © Ministry of Education, Sultanate of Oman. Reproduced with permission from the Oman Ministry of Education; further permission required for reuse.

Malaysia increases over students' lifetimes up to university. In both countries, the effect has been an overrepresentation of females in universities. This has led to a quota selection system in Oman in which males are admitted on lower grade 12 examination scores.

Several possible reasons for performance differences have been advanced: cultural beliefs regarding gender roles; conflicting role expectations for boys and girls; demands to work outside school hours, which differ for boys and girls; the social organization and quality of schooling (use of physical punishment, provision of certain subjects on a gender-biased basis, or predominance of male teachers); and mode of assessment.

Of these possible reasons, the strongest evidence relates to mode of assessment. Studies in a number of countries have found the method of measurement to be related to the performance of males and females. In comparisons of the performance of males and females on multiple-choice and essay-type items, males tended to perform better on the former and females on the latter (Bolger and Kellaghan 1990; Lane and Stone 2006; Mathews 1985; Murphy 1980; Stage 1995). The danger of overgeneralizing about gender differences in achievement across countries is underscored by the findings of the Trends in International Mathematics and Science Study (TIMSS) 2015. Roughly similar percentages of score points were awarded for multiple-choice and constructed response-type items in the grades 4 and 8 tests in both mathematics and science. In slightly more than half of the overall country-by-gender comparisons, boys and girls did not differ in mean achievement scores. In the remaining country comparisons, boys scored higher than girls in some instances, while the reverse was true in others (Mullis et al. 2016).

A decline in the performance of females on Joint Matriculation Board public examinations in England from 1963 to 1980 was attributed to an increase in the use of objective tests and a decrease in the use of essays (Mathews 1985). A similar finding was reported on the basis of analyses of the performance of students on the primary school leaving examination (11 plus) in Trinidad and Tobago following a change from multiple-choice to constructed responses and the removal of science and social studies from the examination. Gender differences, favoring girls, increased (De Lisle et al. 2012).

A number of possible explanations of the relationship between gender and method of measurement have been put forward. It may

be that variables unrelated to content (for example, quality of writing or expectations of readers) affect the score awarded to an essay test, while males, compared to females, tend to guess more in answering multiple-choice questions. When essays were evaluated using an analytic rubric, domain scores for conventions and sentence formation were more affected by gender than domain scores for content or organization and style (Gabrielson, Gordon, and Engelhard 1995). Differences in self-regulatory functions, such as confidence and persistence, may also contribute to gender differences (Martinez 1999).

Gender differences in examination performance are difficult to interpret because of differences in the numbers of males and females taking examinations. Gender differences in examination entry, which reflect differences in subject choice and affect candidates' postexamination options, continue to this day.

Each examination system should conduct its own analyses to identify questions, items, or tasks that are gender biased. The evidence from one or a group of countries should not be used to make general statements about the superiority of one gender in specific subjects or aspects of subjects areas tested.

SOCIOECONOMIC STATUS

The relationship between students' achievements, measured in a variety of ways, and their socioeconomic background is one of the most strongly established findings of empirical research (Kellaghan 2015). The findings have given rise to concern about the extent to which education systems reflect and reproduce social class structure. Research in emerging market economies has been limited and has produced contradictory findings. Some studies have replicated the findings in so-called developed economies; others have not. Failure to find a relationship may be due to limited variation in the variables representing parental education and occupation, the use of inappropriate measures of home background, or the absence of

class differentiation that characterizes these latter economies. When indicators of the material, social, and psychological aspects of home backgrounds are used (for example, parental values and pressure for the child to achieve), one is more likely to find a relationship between home background and school achievement (Greaney and Kellaghan 1995).

A number of factors associated with public examinations seem relevant in considering possible inequalities associated with socioeconomic background. First, the opportunity costs¹ of persisting in school up to and including the time of the final examination have to be considered; potential examination candidates may have to forgo further education to assist in the home or farm or earn some income to support the family. Second, students and their families may have to pay examination-related costs, including those for books and examination fees and, in some instances, boarding fees (especially for some rural students). Policies regarding fees vary from country to country. Although examination fees at the primary level are often nominal, they can be sizable at the secondary level and sometimes beyond the means of families. A third element that may result in inequalities associated with socioeconomic factors relates to the provision of tuition outside school (see chapter 11) or for specialist cram-type schools (see box 10.1). Given the high status attached to public examination performance, it is not surprising that in many countries parents who can afford it seek additional private tuition for their children. It is unlikely that tutoring would be popular if it were not attended by some success. Thus, the less well prepared children of parents who cannot afford it are clearly at a disadvantage in the examination context. Fourth, in countries where examination malpractice flourishes, parents from wealthier families are more likely to be able to afford the costs associated with illegal efforts to gain an advantage for their children (see chapter 11).² Fifth, the introduction of computer-based forms of examinations is likely to favor candidates whose families have provided them with opportunities to use computers at home.

BOX 10.1**Examination Town Where Results Really Matter**

Maotanchang, a small town in eastern China, bears witness to the willingness and ability of parents in rural areas to make extreme financial and personal sacrifices to enable their children to achieve success on the all-important *gaokao* examination. Maotanchang's controversial and highly successful examination record is based on an intensive, regimented cram program. The town has a registered population of 5,000 but houses more than 50,000 people during its nine-month school year, including about 10,000 parents, mostly mothers, who live in small apartments or partitioned rooms. Fees including those from its lucrative repeat examination program support relatively high teacher salaries, bonuses linked to examination results, and well-endowed facilities (see photo B10.1.1). The class schedule extends from 6:10 a.m. to 10:50 p.m. In support of the program, the local government has shut down all forms of entertainment, cellphones and laptops are not allowed, and campus visitors are limited to one three-hour period per week (Larmer 2014; Li 2018; Rudolf 2013).

PHOTO B10.1.1**China: Part of Maotanchang Cram School's 165-Acre Campus**

Photograph © Sim Chi Yin/Magnum Photos. Reproduced with permission from Magnum Photos; further permission required for reuse.

PLACE OF RESIDENCE

Differences in levels of educational attainment between urban and rural children, in favor of the former, are found in many education systems throughout the world but are most pronounced in less-developed regions (Graetz et al. 2018). In South Asian countries, students in states with a high proportion of rural schools perform less well on examinations than students in states with fewer rural schools (Dundar et al. 2014). In addition, a lack of secondary schools in developing countries may require travel to a neighboring town or village, which can deter parents from sending their children, especially their daughters, to school, thus depriving them of the opportunity of sitting for the end of a second-level public examination. Such findings are not surprising given the disadvantages frequently associated with particular parts of a country (often remote rural areas): low income, lack of resources in schools, shortage of school places (particularly at the secondary level), lower per capita investment in schools, limited public transport, teacher shortages, rapid turnover of teachers, and the provision of education in a language that is not widely used in the area (see box 10.2).

Two issues arise from the point of view of inequalities associated with examinations. The first relates to access to the examination; the second is whether or not examinations are likely to underestimate the achievements of rural students. Regarding access, students from rural regions tend to be at a disadvantage because of comparatively low levels of

BOX 10.2

China: Location and Likelihood of Performing Well on the Gaokao

"Inequality is more a function of primary and secondary education than *gaokao* [College Entrance Examination] . . . top students and teaching staff flow continuously from less-developed regions to big cities and provincial capitals. Substandard facilities and teachers in rural areas prevent many rural students from even reaching high school, let alone passing the *gaokao*."

Source: China Policy 2019.

parental education, general support for education in the home, and poorer educational provision in the form of access to school, resources, and trained teachers. Persisting in school in the mid-teens has associated high opportunity costs because of the need for farm help and possibly other local low-level work opportunities. In Australia, 15–19-year-old students from “remote areas” were about twice as likely, and those from “very remote areas” about five times as likely, not to participate in school compared to the overall percentage of students who had not completed grade 10 (Australia Bureau of Statistics 2008). Russian rural students were underrepresented in universities when universities held their own entrance examinations. Following implementation of the Unified State Examination at the end of secondary schooling (see chapter 3), students were able to take examinations in their hometowns. The result was an increase of 15 to 20 percent over the period 2001 to 2008 in the number of students from rural areas attending university (Bolotov et al. 2013).

Students from rural areas may still be disadvantaged even if they have the opportunity of sitting the examinations that determine entry to university. For example, questions posed to national college exam candidates in China’s Sichuan Province about the effect of corporations on modern society that included references to Apple Inc. and to micro-blogging were considered unfair to rural students, who at that time would not have seen either an iPhone or an iPad (Abrams 2012).

Problems of inequality relating to location may also arise when examination and tertiary-level admission authorities in a country, instead of using a national merit-based system, operate on a regional basis or when different cutoff points for passing are applied in different regions. They may, for instance, introduce pass marks for different regions to establish a quota system that is designed to overcome differences in examination performance attributed to disadvantage associated with location or ethnicity. Policy makers, by introducing a form of equity to allow greater access for underrepresented student minorities (such as from rural regions), effectively displace candidates who would normally expect to gain admission based on their learning achievement records. In 1972, Sri Lanka introduced a university admission system that used district quotas to award 60 percent of its university places. This practice, however,

resulted in a new form of inequality: Candidates who did not obtain admission to university in one location would have qualified had they resided in another area. Thus, a candidate's likelihood of gaining admission to university may depend on where he or she lives. A consequence of this situation is that students may change their place of residence or their names to improve their chances of selection. It can also result in problems at the tertiary level due to the widely varying levels of student achievement of incoming students from different regions.

In a somewhat similar vein, Nigeria uses a university admission quota system in an attempt to address economic and educational differences among its 36 states. The system is based on merit (45 percent), university catchment area (35 percent), and degree of educational disadvantage at state level (20 percent).³ This gives rise to a situation in which candidates who would qualify for admission based on an index of national merit but are not from the state where the university is located or from an educationally disadvantaged region within the state may not qualify for admission. The quota system has been subjected to criticism for the low weight assigned to merit or achievement (Opara 2017).

Examinations in Conflict Areas

A particular form of inequity occurs when students reside in areas of armed conflict while others are unaffected. In addition to negatively affecting education, conflict can intensify existing ethnic and religious imbalances. Students in conflict regions within a country may have lower-level learning outcomes (Akresh and de Walque 2008), suffer psychological trauma (Hadi and Llabre 1998), and may not have access to schools. In Sub-Saharan Africa, war has had serious negative consequences on educational participation (Poirier 2012). During Sudan's long civil war, close to 70 percent of 1.06 million eligible students were not enrolled in primary schools (Breidlid 2010). In South Sudan, a separate civil war following independence resulted in the destruction of more than 800 schools, the occupation of many other schools, and the displacement of children and contributed to the postponement of national examinations, in part because of

security concerns (Lutheran World Federation 2015; Walters 2016). Photo 10.2 shows young males taking their grade 8 primary school leaving examination in a refugee camp in that country. Sierra Leone's 1991–2002 civil war resulted in tens of thousands of children being killed and in more than 25,000 children being conscripted (Betancourt et al. 2018). In Mali, conflict disrupted the education of more than 700,000 children (Tran 2013).

Running a public examination system in times of conflict can help underline the legitimacy of the state, but it can represent a target for those opposed to the existence of the state or to certain state supports, such as the provision of girls' education. In Kenya, militants attacked a vehicle transporting examination material (Huho, Mashara, and Musyimi 2016), while in Nigeria the Boko Haram kidnapped 270 schoolgirls who had travelled to their school to take their end-of-year high school exams (Human Rights Watch 2016). A Brookings report noted that during the 2012 Israeli offensive

PHOTO 10.2

South Sudan: Primary School Students Taking Their Final Examination at Napata School in Ajoung Thok Refugee Camp



Photograph © Lutheran World Federation South Sudan. Reproduced with permission from Lutheran World Federation South Sudan; further permission required for reuse.

in Gaza, 11 Palestinian students and four teachers were killed, nearly 300 educational facilities were damaged or destroyed, and examinations were disrupted.

Despite considerable challenges, resourceful officials and others developed solutions to help ensure that examinations were successfully carried out in some conflict areas. In Syria, UNICEF facilitated more than 10,000 Syrian children from remote and besieged areas crossing conflict lines to enable them to sit for their national exams, while in Mogadishu examination sites were guarded against attacks by militants (Watt 2017). In the war-torn Republic of Yemen, teams of educators managed to conduct final examinations in each of the governorates in 2014–15 (Moheyddeen 2016). During the Sri Lankan civil war, the government and the opposing Liberation Tigers of Tamil Eelam had an implicit agreement that allowed public examinations to be conducted in areas under the control of the latter.⁴

TYPE OF SCHOOL

Schools differ in a whole host of ways, including by type of management, specialization, or religious affiliation, whether or not they are public or private, selective or nonselective, fee paying or non-fee paying, or residential or nonresidential. Parents often go to considerable lengths to try to ensure that their children are enrolled in schools that are considered to increase the likelihood of success on public examinations and that may also offer social and socioeconomic benefits.

For a variety of reasons, including geographic, parents in many countries tend not to have many options when selecting a postprimary school for their children; this is particularly true in emerging market economies. The contribution of the school to student achievement, relative to that of the home, seems to be much greater in developing countries than in industrial countries (Heyneman and Loxley 1983). This may be due to greater variance in school conditions in developing countries. Some schools may lack trained teachers or teachers who have an adequate command of the language of instruction or students' home language. There may also be differences in the

availability of textbooks, desks, adequate lighting, science equipment, and the range of curriculum options required by students seeking university admission.

It is sometimes claimed that private (unaided) schools in low-income countries offer lower-quality education than aided schools because of more limited resources, less qualified teachers, and less government oversight, resulting in poorer student learning. There is evidence, however, that this is not necessarily the case. In Nepal, for example, once observable student characteristics were taken into account, private schools appeared no worse, and often were better, than public schools (Dundar et al. 2014). In a similar vein, evidence from Organisation for Economic Co-operation and Development member countries suggests that private schools may offer no academic performance advantage after socioeconomic background factors have been taken into consideration (OECD 2010). Differences between private and public schools are often difficult to interpret because of the failure to recognize the huge variance in the facilities of private schools and in their clienteles.

ETHNIC OR LANGUAGE GROUP MEMBERSHIP

Ethnic inequality occurs when one or more racial groups are politically or economically disadvantaged. In education, it occurs when access to resources, including further education, is limited by one's racial or ethnic origin. For example, governments in countries such as Sri Lanka in the past, and more recently Malaysia, have used examination-based quota systems to favor a particular ethnic group.⁵ It can also occur in situations in which there are insufficient qualified teachers from ethnic minorities who speak the local language or when schools in ethnic minority areas are inadequately resourced in terms of buildings and materials.

Support for students receiving at least their primary-level education in their native language is based in part on the idea that students benefit from using their home language, and student's literacy skills acquired in the first language can subsequently transfer to or aid in the acquisition of second language skills (Abadzi 2006;

August and Shanahan 2006; Cummins 2000; Ellis 2009; UNESCO 2016). However, it is not unusual for children in many emerging market economies and immigrants to be taught from the beginning in second or third languages in which they may have limited or no proficiency. In Mauritius, for example, most children speak Creole, a lingua franca that is not written; others of Asian ancestry speak Bhojpuri, Tamil, Hindi, Marathi, Urdu, Telugu, Odia, and Chinese, while teaching and learning is carried out in French and English (Hollop 2004). India has more than 19,500 languages or dialects; 121 are spoken by 10,000 or more people (Press Trust of India 2018). A 2002 study shows that most primary schools taught in the most common regional mother tongue, but 87 percent of secondary schools taught in another language (National Council of Educational Research and Training 2002). In Sub-Saharan Africa, relatively few countries use an African language as a medium of instruction at the secondary-school level. This is partly due to the multiplicity of local languages (500 alone in Nigeria), the lack of a dominant language in many areas, a lack in teachers with local-language competency (see, for instance, Muthwii 2004), and established policies of using European languages for secondary and higher education (Bamgbose 2004). Secondary-school examinations are generally offered in colonial languages (English, French, or Portuguese) (Bamgbose 2004; Kellaghan and Greaney 1992).

Are students further disadvantaged by having to take examinations in a language other than their home language? If the purpose of an examination is to determine students' level of achievement in a subject other than the language itself (for example, science or history), the answer would appear to be "yes." From an early stage, second-language learners, especially those from poor countries or regions, face disadvantages (Abadzi 2006). As learning to read is related to children's early home literacy experiences (Greaney 1996; Guthrie and Greaney 1991; Mullis et al. 2017) it is likely that students who enroll in schools in which they face linguistically different environments face additional problems in learning to read accurately and quickly and acquiring basic comprehension skills. Furthermore, they are placed at a disadvantage and are likely to require assistance in mastering the language conventions that typically are used in

everyday communication across subjects, subject-specific academic vocabularies, and forms of expression (Ballantyne and Rivera 2014).

International studies of achievement have revealed that grade 8 mathematics and science students, as well as grade 4 reading literacy students, who “sometimes” or “never” spoke the language of the test at home scored lower than their peers who spoke it more frequently (Martin et al. 2016; Mullis et al. 2016). In the 2011 international reading literacy study, about one-fifth of the students in the Islamic Republic of Iran spoke a language other than the official language (Farsi) at home. Of these, only about half reached the minimum learning standard, compared with more than 80 percent of Farsi speakers (Altinok 2013).

A number of commentators have attributed low levels of achievement in public examinations to the language difficulties of students, pointing out that students with a limited knowledge of a language will inevitably be handicapped in their ability to demonstrate in an examination the knowledge and skills they have acquired (Eisemon 1990). For example, it has been stated that students from minority ethnic groups in Vietnam are disadvantaged because the examinations have to be taken in the Vietnamese language (World Bank 2009). Similarly in the Russian Federation nonethnic students whose first language is other than Russian have been disadvantaged by having to take the Unified State Examination in Russian (Artemev 2014).

In South Africa, the Senior Certificate Examination “is seen as one of the prominent symbols of the inequities of the past” (Berkhout and Beets 2009, 146). In support of this view, it is noted that in 1988, 95 percent of white candidates, compared to 56 percent of black candidates, passed the examination, and 43 percent of white candidates attained matriculation status, compared to 17 percent of black candidates. Four years later, in 1992, the number of black candidates had increased from about 184,000 to about 350,000. However, the inequality of performance increased. Almost all (97 percent) of white candidates, but only 45 percent of black candidates, passed the examination, while 41 percent of white candidates but only 10 percent of black candidates attained matriculation exemption, a requirement for students who did not earn the certificate. The reasons for this state of

affairs were said to be differences in family traditions of schooling, differences in the quality of educational provision (resources and quality of teachers), and the fact that candidates were examined in only 2 of South Africa's 11 languages, English or Afrikaans. Affirmative action and the recognition of prior learning and alternative routes of learning for university admission have been proposed to address this situation.

Differences in performance on public examinations of ethnic minorities in Great Britain have given rise to the issue of special consideration in examinations in an attempt to address the difficulties ethnic groups might experience (Mathews 1985). One of these difficulties, lack of facility in the language of an examination, raises validity issues, which are considered in chapter 7.

CONCLUSION

A variety of factors associated with examinations may create inequities for some students. These inequities may be related to the costs of examination fees or of private tutoring. Some students may be placed at an advantage or disadvantage because of culturally inappropriate items, having to take the examination in a relatively unfamiliar language, or different forms of malpractice (see box 10.3).

Over time several national authorities have taken steps to remove some of the inequities that could have a bearing on participation and performance in second-level public examinations. These steps include introducing laws and regulations that have raised the legal working age and the minimum school leaving age; supporting funding to enable girls to complete secondary education; covering fees and supplies, including books, for needy students; abolishing or modifying admission quota systems; and providing special facilities for candidates with special needs. Other steps adopted by some examination authorities to improve equity include training paper setters to address question bias, addressing issues related to examination malpractice, and introducing identification numbers to ensure that marking of examination papers is not affected by considerations of gender, socioeconomic status, location, or ethnicity (see chapter 11).

BOX 10.3**Practices Associated with Examinations That May Create Inequities**

- Use of culturally inappropriate questions
- Use of questions containing content more familiar to one gender than to another
- Examinations set in a language with which examinees may be relatively unfamiliar
- Inadequate provision for students with special needs
- Requirement that candidates pay fees
- Excessive influence of private tuition
- Malpractice (for example, bribery, intimidation, impersonation)
- Quota systems to deal with differences in performance associated with location, ethnicity, and language group
- Ranking of schools on the basis of student examination performance

It is often difficult to separate inequalities that might be due to examinations from more general inequalities and inequities in education and broader societal systems. The distinction between “inequality” and “inequity” is an important one: the former describes disproportionate conditions; the latter implies unjust differences. For example, inequalities in participation rates in higher education by different social groups may reflect differences in fitness to benefit from such education, however undesirable such differences may be. If in this situation examinations contribute to the selection of students for higher education, they may be considered to be a confirmatory and publicly defensible judgment of inequalities, which may not be inequitable (Brimer et al. 1978). We may have to accept that “examinations are no more or less fair than schools are fair, or society is fair” (Mathews 1985, 136). However, care in the construction, administration, and scoring of public examinations may go some small way toward addressing these inequities.

NOTES

1. Opportunity costs refer to the alternative options (such as loss of income) not taken advantage of when a decision is made for a child to attend school.
2. In a somewhat similar situation, wealthy parents in the United States used various strategies, including bribing of officials, to ensure that their children gained admission at some of the most prestigious universities (see chapter 11).
3. Merit is based on examination performance and on a separate standardized test.
4. Harsha Aturupane, personal communication, March 14, 2019.
5. Some educational systems promote policies, such as affirmative action (United States), caste-based reservation (India), or district-based merit (Malawi), to promote educational opportunities for disadvantaged groups. A number of countries, for example, Sweden (De Siene 2006), the United Kingdom (United Kingdom 2010), and the United States (Devins 2003), however, do not permit the use of university admission ethnicity quotas.

REFERENCES

- Abadzi, H. 2006. *Efficient Learning for the Poor: Insights from the Frontier of Cognitive Neuroscience*. Washington, DC: World Bank.
- Abrams, S. 2012. "Is China's *Gaokao* Examination Biased against the Rural Poor?" *China Hearsay* (blog), June 13. <http://www.chinahearsay.com/is-china-gaokao-biased/>.
- Ahmad, N. A., S. A. Hassan, A. R. Ahmad, C. Lay Nee, and N. H. Othman. 2017. "The Trend of Academic Achievement among Malaysian Boys and Girls: Where Are the Boys?" *Global Journal of Business and Social Science Review* 5: 1–8. <https://ssrn.com/abstract=3002369>.
- Akresh, R., and D. de Walque. 2008. "Armed Conflict and Schooling: Evidence from the 1994 Rwandan Genocide." Policy Research Working Paper 4606, World Bank, Washington, DC.
- Altinok, N. 2013. "Performance Differences between Subpopulations in TIMSS, PIRLS, SACMEQ and PASEC." Background paper for *EFA Global Monitoring Report 2013–14*, United Nations Educational, Scientific and Cultural Organization, Paris.
- Amano, I. 1990. *Education and Examination in Modern Japan*, translated by W. K. Cummings and F. Cummings. Tokyo: University of Tokyo Press.

- Artemev, N. 2014. "The Problems of the Unified State Examination in Russia." *Education Policy Talk: International Perspectives and Debates* (blog), May 23. <https://educationpolicytalk.com/2014/05/23/the-problems-of-the-unified-state-examination-in-russia/>.
- August, D., and T. Shanahan. 2006. *Developing Literacy in Second-Language Learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah, NJ: Erlbaum.
- Australia Bureau of Statistics. 2008. "Education across Australia." In *Australian Social Trends, 2008*, 93–98. <http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4102.0Chapter6002008>.
- Ballantyne, K., and C. Rivera. 2014. "Language Proficiency for Academic Achievement in the International Baccalaureate Diploma Program." Center for Equity and Excellence in Education, George Washington University, Washington, DC. <https://www.ibo.org/globalassets/publications/ib-research/language-proficiency-final-report.pdf>.
- Bamgbose, A. 2004. "Language of Instruction Policy and Practice in Africa." United Nations Educational, Scientific and Cultural Organization, Paris. http://www.unesco.org/education/languages_2004/languageinstruction_africa.pdf.
- Berkhout, S. J., and A. D. Beets. 2009. "Country Case Study: South Africa." In *Secondary School External Examination Systems*, edited by B. Vlaardingerbroek and N. Taylor, 135–52. Amherst, NY: Cambria.
- Betancourt, T. S., S. Zuilkowski, E. Coles, K. Collet, and M. Jambai. 2018. "Adolescents and Armed Conflict: War, Conflict and Child Soldiers." In *Handbook of Adolescent Development Research and Its Impact on Global Policy*, edited by J. E. Lansford and P. Banati, 196–216. Oxford, UK: Oxford University Press.
- Bolger, N., and T. Kellaghan. 1990. "Method of Measurement and Gender Differences in Scholastic Achievement." *Journal of Educational Measurement* 27: 165–74.
- Bolotov, V., G. Kovaleva, M. Pinskaya, and I. Valdman. 2013. *Developing the Enabling Context for Student Assessment in Russia*. Washington, DC: World Bank.
- Breidlid, A. 2010. "Sudanese Images of the Other: Education and Conflict in Sudan." *Comparative Education Review* 54: 555–75.
- Brimer, A., G. F. Madaus, B. Chapman, T. Kellaghan, and R. Wood. 1978. *Sources of Difference in School Achievement*. Windsor Berks, UK: National Foundation for Educational Research in England and Wales.
- Camili, G. 2006. "Test Fairness." In *Educational Measurement*, 4th ed., edited by R. L. Brennan, 220–56. Westport, CT: American Council on Education and Praeger.

- Cartwright, F. 2015. "Item and Test Analyses." In G. Shiel and F. Cartwright, *Analyzing Data from a National Assessment of Educational Achievement*, 123–266. Washington, DC: World Bank.
- China Policy. 2019. "Gaokao: Reform and Reality Clash." *CP Signal*, January 25. <https://policycn.com/?r=https%3A%2F%2Fpolicycn.com%2F%2F19-01-24-gaokao-reform-and-reality-clash%2F&login=true>.
- Cros, F. 2009. "Country Case Study: France." In *Secondary School External Examination Systems: Reliability, Robustness and Resilience*, edited by B. Vlaardingerbroek and N. Taylor, 15–28. Amherst, NY: Cambria.
- Cummins J. 2000. *Language, Power and Pedagogy: Bilingual Children in the Crossfire*. Clevedon, UK: Multilingual Matters.
- De Lisle, J., P. Smith, C. Keller, and V. Jules. 2012. "Differential Outcomes in High-Stakes Eleven Plus Testing: The Role of Gender, Geography, and Assessment Design in Trinidad and Tobago." *Assessment in Education: Principles, Policy and Practice* 19: 45–64.
- De Siene, M. 2006. "Uppsala Discriminated against Swedes." *Local*, December 21. <https://www.thelocal.se/20061221/5871>.
- Devins, N. 2003. "Explaining Grutter v. Bollinger." *University of Pennsylvania Law Review* 152: 347–83.
- Dundar, H., T. Béteille, M. Riboud, and A. Deolalikar. 2014. *Student Learning in South Asia: Challenges, Opportunities, and Policy Priorities*. Washington, DC: World Bank.
- Eckstein, M. A., and H. J. Noah. 1993. *Secondary School Examinations: International Perspectives on Policies and Practice*. New Haven, CT: Yale University Press.
- Eisemon, T. O. 1990. "Examination Policies to Strengthen Primary Schooling in African Countries." *International Journal of Educational Development* 10: 69–82.
- Ellis, R. 2009. *The Study of Second Language Acquisition*. Oxford, UK: Oxford University Press.
- Gabrielson, S., B. Gordon, and G. Engelhard. 1995. "The Effects of Task Choice on the Quality of Writing Obtained in a Statewide Assessment." *Applied Measurement in Education* 8: 278–90.
- Gierl, M., S. N. Khaliq, and K. Boughton. 1999. "Gender Differential Item Functioning in Mathematics and Science: Prevalence and Policy Implications." Paper presented at the Symposium "Improving Large-Scale Assessment in Education," Canadian Society for the Study of Education, Sherbrooke, Quebec, Canada, June 7–11.
- Government of Malaysia. 2015. "National Education for All Review Report: End of Decade Review." United Nations Educational,

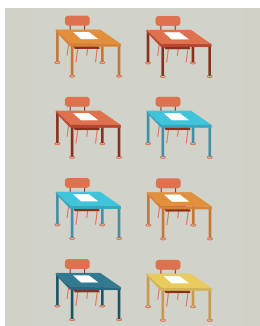
- Scientific, and Cultural Organization, Paris. <http://unesdoc.unesco.org/images/0022/002297/229719E.pdf>.
- Graetz, N., J. Friedman, A. Osgood-Zimmerman, R. Burstein, M. H. Biehl, C. Shield, J. F. Mosser, et al. 2018. "Mapping Local Variation in Educational Attainment across Africa." *Nature* 555: 48–53. doi:10.1038/nature25761.
- Greaney, V. 1996. "Reading in Developing Countries: Problems and Issues." In *Reading in Developing Countries*, edited by V. Greaney, 5–38. Newark, DE: International Reading Association.
- Greaney, V., and T. Kellaghan. 1995. *Equity Issues in Public Examinations in Developing Countries*. Washington, DC: World Bank.
- Guthrie J. T., and V. Greaney. 1991. "Literacy Acts." In *Handbook of Reading Research*, edited by R. Barr, M. L. Kamil, P. Mosenthal, and P. D. Pearson, 68–96. New York: Macmillan.
- Hadi, F. A., and M. A. Llabre. 1998. "The Gulf Crisis Experience of Kuwaiti Children: Psychological and Cognitive Factors." *Journal of Traumatic Stress* 11: 45–56.
- Heyneman, S. P., and W. A. Loxley. 1983. "The Effect of Primary-School Quality on Academic Achievement across Twenty-Nine High- and Low-Income Countries." *American Journal of Sociology* 88: 1162–94.
- Hollop, O. 2004. "Educational Policies, Reforms and the Role of Teachers Unions in Mauritius." HiT Publication 8/2004, Telemark University College, Porsgrunn, Norway. https://www.researchgate.net/publication/37686886_Educational_policies_reforms_and_the_role_of_teachers_unions_in_Mauritius.
- Hu, C. T. 1984. "The Historical Background: Examinations and Control in Pre-modern China." *Comparative Education* 20: 7–26.
- Human Rights Watch. 2016. *"They Set the Classrooms on Fire": Attacks on Education in Northeast Nigeria*. Washington, DC: Human Rights Watch. <https://www.hrw.org/report/2016/04/11/they-set-classrooms-fire/attacks-education-northeast-nigeria#>.
- Huho, J. M., J. N. Mashara, and P. K. Musyimi. 2016. "Profiling Disasters in Kenya and Their Causes." *Academic Research International* 7: 290–305.
- Judd, C. H. 1935. "The Third Conference of the International Inquiry on School and University Examinations." *School Review* 43: 653–63.
- Kellaghan, T. 2015. "Family and Schooling." In *International Encyclopedia of the Social and Behavioral Sciences*, 2d ed., edited by J. D. Wright, vol. 3, 751–57. Oxford, UK: Elsevier.
- Kellaghan, T., and V. Greaney. 1992. *Using Examinations to Improve Education: A Study in Fourteen African Countries*. Washington, DC: World Bank.

- Lane, S., and C. A. Stone. 2006. "Performance Assessment." In *Educational Measurement*, 4th ed., edited by R. L. Brennan, 387–431. Westport, CT: American Council on Education and Praeger.
- Larmer, B. 2015. "Inside a Chinese Test-Prep Factory." *New York Times Magazine*, January 4. <https://www.nytimes.com/2015/01/04/magazine/inside-a-chinese-test-prep-factory.html>.
- Li, W. 2018. "Asia's Largest 'Cram School' Sends 15,000 Students to Gaokao Exam." *GBTimes*, June 5. <https://gbtimes.com/asias-largest-cram-school-sends-15000-students-to-gaokao-exam>.
- Lutheran World Federation. 2015. "Final Exams for Primary Students in Ajourng Thok." Lutheran World Federation, Geneva. <https://www.lutheranworld.org/news/final-exams-primary-students-ajourng-thok>.
- Martin, M. O., I. V. S. Mullis, P. Foy, and M. Hooper. 2016. *TIMSS 2015 International Results in Science*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College.
- Martinez, M. E. 1999. "Cognition and the Question of Test Item Format." *Educational Psychologist* 34: 207–18.
- Mathews, J. C. 1985. *Examinations: A Commentary*. Boston: Allen and Unwin.
- Moheyddeen, K. 2016. "Education in Yemen Struggles after More Than a Year of Conflict." *Voices and Views: Middle East and North Africa* (blog), October 7. <http://blogs.worldbank.org/arabvoices/education-yemen-struggles-conflict>.
- Montgomery, R. 1965. *Examinations: An Account of Their Evolution as Administrative Devices in England*. London: Longmans Green.
- Mullis, I. V. S., M. O. Martin, P. Foy, and M. Hooper. 2016. *TIMSS 2015 International Results in Mathematics*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College.
- . 2017. *PIRLS 2016 International Results in Reading*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College.
- Murphy, R. J. L. 1980. "Sex Differences in Objective Test Performance." *British Journal of Educational Psychology* 52: 213–19.
- Muthwii, M. J. 2004. "Language of Instruction: A Qualitative Analysis of the Perceptions of Parents, Pupils and Teachers among the Kalenjin in Kenya." *Language, Culture and Curriculum* 17: 15–32.
- National Council of Educational Research and Training. 2002. *Seventh All India School Education Survey*. New Delhi, India: National Council of Educational Research and Training. Cited in P. Glewwe, M. Kremer, and S. Moulin, "Many Children Left Behind? Textbooks and Test Scores

- in Kenya" (Working Paper 13300, National Bureau of Economic Research, Cambridge, MA, 2007; <http://www.nber.org/papers/w13300>).
- OECD (Organisation for Economic Co-operation and Development). 2010. *PISA 2009 Results: What Makes a School Successful? Resources, Policies and Practices*. Paris: OECD.
- Oman Ministry of Education and World Bank. 2012. *Education in Oman: The Drive for Quality*. Muscat: Oman Ministry of Education and World Bank.
- Opara, J. 2017. "University Quota System—Pushing Fairness or Mediocrity?" *University World News: Africa Edition* 190, February 10. <http://www.universityworldnews.com/article.php?story=20170201142523374>.
- Osterlind, S. J., and H. T. Everson. 2009. *Differential Item Functioning*. Thousand Oaks, CA: Sage.
- Poirier, T. 2012. "The Effects of Armed Conflict on Schooling in Sub-Saharan Africa." *International Journal of Educational Development* 32: 341–51.
- Press Trust of India. 2018. "More Than 19,500 Mother Tongues Are Spoken in India, Says Report." *Business Standard*, July 1. https://www.business-standard.com/article/current-affairs/more-than-19-500-mother-tongues-are-spoken-in-india-says-report-118070100250_1.html.
- Roach, J. 1971. *Public Examinations in England 1850–1900*. Cambridge, UK: Cambridge University Press.
- Rudolf, J. 2013. "China's Cram School from Hell." *China Digital Times*, October 11. <https://chinadigitaltimes.net/2013/10/chinas-cram-school-hell/>.
- Stage, C. 1995. "Gender Differences on Admission Instruments for Higher Education." In *Admission to Higher Education: Issues and Practice*, edited by T. Kellaghan, 105–14. Princeton, NJ: International Association for Educational Assessment and Dublin: Educational Research Centre.
- Tran, M. 2013. "War Denying Millions of Children an Education." *Guardian*, July 12. <https://www.theguardian.com/global-development/2013/jul/12/war-denying-children-education>.
- UNESCO (United Nations Educational, Scientific and Cultural Organization). 2016. *Global Education Monitoring Report: Education for People and Planet; Creating Sustainable Futures for All*. Paris: UNESCO.
- . 2017. *Global Monitoring Report—Accountability in Education: Meeting Our Commitments*. Paris: UNESCO.
- United Kingdom. 2010. "Equality Act 2010." <https://www.legislation.gov.uk/ukpga/2010/15/part/1>.
- Walters, Q. J. 2016. "Conflict Zones: In South Sudan, a Struggle to Get, and Keep, Kids in Schools." *Parallels*, March 20. National Public Radio, Washington, DC. <https://www.npr.org/sections/parallels/2016/03/20/465439724/in-south-sudan-a-struggle-to-get-and-keep-kids-in-schools>.

- Watt, E. 2017. "Courage of Children Who Ignore Conflict and Threats to Take Their Exams." *Their World*, July 14. <https://theirworld.org/news/brave-children-ignore-conflicts-and-threats-to-take-school-exams>.
- World Bank. 2009. *Vietnam: Student Assessment*. SABER Country Report. Washington, DC: World Bank.
- . 2018. *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC: World Bank.

CHAPTER 11



THE INTEGRITY OF PUBLIC EXAMINATIONS AND MALPRACTICE

INTRODUCTION

This chapter examines the extent to which procedures to standardize the conditions under which examinations are prepared, administered, and scored are violated. Only if such procedures are successfully implemented can we be sure that the integrity or soundness of examinations is maintained so that no candidate is placed at an advantage or disadvantage relative to other candidates and that, as a consequence, the marks or grades awarded are directly related to the ability that is being measured. The extent to which these procedures are enforced is a critical factor in determining the validity of students' examination scores as well as the confidence of other users of examination results.

Examination boards clearly recognize that uncontrolled conditions arise from time to time and use a variety of terms in their regulations to describe practices that would interfere with the integrity of the examinations. These terms include "misconduct," "dishonest conduct," "cheating," "unfair practice," "irregularity," "dishonesty," and "corruption." It seems reasonable to subsume all these terms under the term "malpractice," which is the one used in this chapter to

describe any action taken in connection with an examination or test that attempts to gain unfair advantage or, in some cases, to place a candidate at a disadvantage.

FORMS OF MALPRACTICE

Malpractice may occur at any stage in the examination process, from development of the examination through preparation of students, in the actual administration of examinations and marking, and finally in the issuance and use of results (Greaney and Kellaghan 1996). Since some individuals will pay large sums of money for advance information about the content of high-stakes examinations or for assistance to individual candidates during or after examinations, the temptation to engage in malpractice is, and will remain, strong.

Examination Development

Leakage occurs when the content of any part of an examination is disclosed prior to candidates taking the examination. The information can be released verbally, on a computer file, on paper, or on a physical object (such as pre-engraved pencils). Offending personnel may include staff members of examination authorities, printers, proof-readers and messengers, personnel employed to develop the examination (setters) or to determine its suitability (moderators), and school administrators.

Preparation of Students for Tests or Examinations

Test preparation may or may not be regarded as malpractice, depending on conditions surrounding test administration. It is generally accepted that teaching test-taking skills and the content and skills known to be covered in a test is acceptable. However, in the United States, when tests are not released to schools after students have taken them (because items may be used in subsequent assessments), the provision of practice on a parallel form of a test is not regarded as ethical (Haladyna, Nolen, and Haas 1991; Mehrens and Kaminski

1989). In the case of public examinations, in both industrial and developing countries, where examination papers normally enter the public domain after an examination, practice on earlier “parallel” forms is not regarded as unethical.

Administration of Examinations

Impersonation occurs when an individual who is not registered as a candidate takes the place of one who is registered. It is usually difficult for impersonation to take place without the knowledge of the school principal or the collusion of a supervisor. It frequently involves university students or teachers taking the test for monetary reward or as a favor for a girlfriend or boyfriend. It may involve coercion when, for example, a young employee is required by an employer to impersonate another. It can be difficult to counteract, especially in some Muslim cultures in which female faces are veiled or female candidates are exempt from having their photographs taken for identification cards. Private or external candidates who are not known to school principals or staff pose particular identification challenges.

External assistance involves individuals who are not examination candidates giving unauthorized assistance to candidates. Teachers, parents, and other “helpers” might pass information from outside during the course of the examination (see photo 11.1). Invigilators are also frequently involved. They may dictate answers, work answers on a blackboard, circulate sheets of worked-out answers during the course of the examination, or act as couriers of material into the examination center. Modern technology poses severe challenges to examination authorities (Foster 2015; Maynes 2008). In recent years, the mobile phone has been identified as the most common cheating device (Dorff 2013). Candidates have used mobile phones and the more powerful iPhones to access such information sources as files, diagrams, dictionaries, and maps. Pinhole cameras and specially equipped smart or digital glasses (see photo 11.2) can connect candidates to outside helpers who, in turn, can transmit responses through smart watches, phones, or MP3 players as text or as audio messages through tiny earpieces.¹ Audio contact with external helpers can be maintained through smart watches and other two-way devices. Social

PHOTO 11.1

India: Parents and Other Helpers Climbing a School Wall in the East Indian State of Bihar to Help Children Cheat on the High-Stakes Grade 10 Public Examination



Photograph © Press Trust of India. Reproduced with permission from Press Trust of India; further permission required for reuse.

media sites have facilitated the circulation of photographs of examination papers to tutors during the course of an examination, while in-ear microphones have permitted recording of listening portions of examinations.

Smuggling of foreign materials relates to the introduction of unauthorized material (for example, notebooks, “crib notes,” charts, and answer booklets complete with answers) into an examination hall. Material is frequently smuggled in pants, shoes, hems, jewelry, beverage or food items, and bras, or information may be written on parts of the body. Programmable calculators and other modern technological devices have been used to access unauthorized material. Candidates have also used ultraviolet pens to access information written previously in invisible ink on what appears to be scrap paper.

Copying refers to the reproduction of another candidate’s work, with or without permission. Lack of a proper hall, inadequate spacing

PHOTO 11.2

A Police Officer Displaying a Pair of Confiscated Glasses with a Hidden Camera and a Tiny Receiver Attached to a Coin, Both Pieces of Equipment Used for Cheating on Examinations



Photograph © Stringer/Reuters. Reproduced with permission from Reuters; further permission required for reuse.

between candidates, poor-quality furniture, and laxity of supervision all facilitate copying.

Collusion involves unauthorized passing of information between candidates during the course of an examination. Traditional strategies include exchanging notes or scripts, throwing balls of paper containing material that would be of assistance to other candidates, or using codes (such as foot or hand movements) to signal responses to multiple-choice questions. Devices such as pagers, mobile phones, and smart watches allow for more sophisticated and less obvious forms of collusion.

Intimidation occurs when examination officials, including supervisors and markers of papers, are physically threatened by people seeking support for individual candidates or by a candidate signalling possession of a weapon.

Theft involves stealing examination papers or files during examination preparation, printing, or distribution. Modern technology facilitates a form of cyber theft in which external bodies or “hackers” gain

illegal access to examination papers, scoring keys, results data, and examination certificates.

Improper assignment involves the deliberate placing of candidates in centers under the supervision of corrupt officials.

Ghost centers are fictitious centers established with the assistance of corrupt examination officials. In these unsupervised sites (for example, houses), candidates can complete their papers without supervision and with the assistance of books and helpers.

Marking of Examinations and Issuing of Results

Marking malpractice may be initiated by a candidate or it may occur as a result of the initiative of an examination authority staff member. Candidates may take the initiative in contacting markers by providing their addresses and phone numbers on their answer papers, presumably in the expectation that a marker may wish to make contact with them at a later stage for a financial payment. Markers also may identify particular candidates and offer to do deals with their parents. Corrupt officials may take the initiative by revealing the names of markers to parents of candidates, again opening the way for intimidation or bribery. In some cases, parents may intervene directly in the marking process.

Awards and certificates may be forged by an individual or a company specializing in manufacturing fake certificates. One such company has been charged with selling more than 200,000 fake certificates in 197 countries (Walsh 2015). Finally, a candidate's official final ranking or diploma may be unjustly enhanced by an official of the examination authority.

FREQUENCY OF MALPRACTICE

Interpretation of data on the frequency of malpractice in tests or examinations must recognize that probably only a small percentage of those who cheat are caught, while those who admit to cheating in questionnaires might also represent an underestimate. In view of these and other considerations, it is difficult to obtain precise data on the frequency of malpractice.

On the basis of anecdotal evidence, malpractice on public examinations seems to be endemic in some places, such as Bangladesh, India, Nigeria, and Pakistan. According to one observer, in India “schools where cheating is not reported make newspaper headlines” (Moore 1993, A14); a teacher in one school noted that “during yesterday’s exams, every student was cheating” (Moore 1993, A20). In Baluchistan, a head teacher reported that in a typical examination session, 100 percent of students cheat; even those who initially seemed to have no intention of doing so would start when they saw what was going on around them. In Nigeria, the 2013 results of more than 12,000 matriculation candidates were withheld pending possible disciplinary action (*Vanguard* 2013); the introduction of modern technology was said to have “revolutionized examination malpractice” (Onyibe, Uma, and Ibina 2015, 14).

The limited evidence from Organisation for Economic Co-operation and Development member countries suggests that malpractice occurs, though not to the same extent as in some nonindustrial countries. In the United Kingdom, the reported percentage of examinations taken that involved malpractice in 2010 amounted to 0.03 (Patton 2010). More recently, one examination board revealed that thousands of teachers had given improper assistance to candidates (Griffiths and Corke 2018). A study in the far east of the Russian Federation notes that 31 percent of respondents had observed some violations during the administration of the Unified State Examination (Denisova-Schmidt and Leontyeva 2014). In the United States, cheating on standardized tests has been confirmed in 39 states and Washington, DC (Schaeffer 2014); a separate study of 24,000 students in 70 high schools reported that 64 percent admitted to cheating on a test (Plagiarism.org 2017). Also in the United States, in 2019, a scheme involving prominent wealthy parents turned up evidence of malpractice, including impersonation, fake documentation, adjusting of scores, and bribery of officials, which helped buy spots in first-year classes in some of the country’s top universities (Medina, Benner, and Taylor 2019).

REASONS FOR MALPRACTICE

The level of malpractice in public examinations should be viewed against a series of educational, social, economic, political, and

legal realities. Perhaps the most obvious occasion for malpractice is when examination results are used for competitive purposes and can have a profound and immediate impact on the course of an examinee's life.² Reasons for malpractice tend to vary according to economic circumstances; they include the relatively low risk of getting caught, high-level expectations of parents and family members, and the intensity of the pressure to get into competitive areas of tertiary study (such as medicine and law). In addition, when accountability systems put teachers under intense pressure to get good results and on schools to have high rankings, the likelihood of cheating increases (Cizek 2001).

Findings have shown that attitudes toward cheating can vary among countries (Kuehn, Stanwyck, and Holland 1990; Magnus et al. 2002). In emerging market economies, even when the stakes attached to performance seem uniformly high, there would appear to be considerable differences in the extent of malpractice from country to country. The available evidence suggests that other factors, in addition to the high-stakes nature of examinations, should also be looked at in attempting to explain malpractice (see box 11.1) These might include personal factors, such as the propensity of some students to cheat. Situational factors—such as quota systems based on ethnicity, location, and gender—may lead some students to conclude that results are determined by causes beyond their control and may be used as a justification for resorting to unfair means to improve chances of examination success. Ethiopia provides an example of the use of examinations to embarrass the government; as part of a political protest, activists leaked national examination papers before the examination date (De Birhaner 2016). Poor political leadership can also contribute to malpractice, as exemplified in the following statement attributed to a chief minister: “A degree is a degree! Whether fake or genuine, it’s a degree! It makes no difference!”³

STRATEGIES TO DISCOURAGE AND DETECT MALPRACTICE

All examination authorities employ strategies to discourage and detect malpractice. In 2018, one national examination board considered it necessary to issue a press release warning of the existence of

BOX 11.1**Reasons for Examination Malpractice**

- High-stakes nature of examination for students, teachers, schools, and parents
- Personal characteristics of some candidates
- Attitudes toward cheating in school or in society
- Bribes offered by parents, school officials, and others
- Quota systems related to ethnicity, location, and gender
- Inadequate remuneration for examination officials, administrators, and teachers
- Inadequate seating arrangements during examination
- Poor supervision and quality control during examination administration, paper development, scoring, and issuance of results
- Low probability of being caught
- Inadequate legislation or enforcement of existing legislation

fake examination papers (Uganda National Examinations Board 2018). Some countries require candidates to sign a statement declaring that any project they submit represents wholly their own work. A number employ additional supervisors during examinations. Other state bodies, including military and police, may also provide support. For example, in 2016, eight police officers were deployed to each of 96 examination centers in Beijing to monitor cheating (Hernandez 2016).

Candidates may be required to submit photographs at the time of registration, which may be checked at the actual examination. Thumbprints are also used, though it may be noted that fingerprint film has been used to bypass fingerprint scanners. Although school principals and invigilators are often expected to ensure that only legitimate candidates are allowed to sit the examination, it is evident that many fail to perform this duty in a satisfactory manner.

Various approaches have been used after an examination to check for evidence of suspicious behavior that should be followed up to look for evidence of malpractice. For essay-type examinations and ones that

include project or portfolio components, software such as Turnitin will look for matching texts in examination papers or scan responses to see if material has been plagiarized from online or published material. For multiple-choice type examinations, forensic data analysis approaches are available to identify highly improbable response patterns, such as when a series of difficult items are answered correctly by a candidate who had answered easy items incorrectly (Maynes 2009). For items that have been repeated in a number of examinations, a comparison of point-biserial correlations can help detect improper access. Data from within and across sites can be analyzed to determine if response patterns are sufficiently similar to suggest systematic cheating or collusion. Small wireless devices issued to invigilators can help locate candidates who might be using electronic methods to receive and transmit information. However, initial efforts to introduce security systems that restrict mobile and internet access to examination sites have proved to be very expensive. In addition, it should be noted that blocking or jamming mobile signals is illegal in some countries.⁴

When examinations require extensive written responses, the presence of neatly written material unrelated to questions on the examination paper in the middle of relevant hurriedly written answers may suggest that answers were prepared beforehand and smuggled into the examination. Evidence of copying tends to be based on the presence of identical mistakes and peculiarities in scripts answered by candidates sitting close to each other. If collusion among candidates is suspected, reference may be made to the seating arrangement of candidates recorded on the day of the examination.

Malpractice does not cease with the issuing of certificates. Examination authorities have the responsibility of ensuring that their certificates are not interfered with or are not used by individuals who have not earned them. This is not an easy task. As noted previously, fake certificates are widely available (Walsh 2015). Strategies that are used to detect alteration of marks on an official diploma include using a code number on the diploma that contains information on the candidate's performance. For example, if a candidate was awarded 86 on a mathematics test, the code number might read 527648. The original mark can be calculated by subtracting each of the second and fifth digits from 10.

WAYS TO CONTROL AND PREVENT MALPRACTICE

Control of examination malpractice is a never-ending battle. To help limit it, examination authorities should continually monitor their security procedures. They should try to ensure that question papers, completed answers, and other materials are secure during distribution and collection. The use of computer-based examinations does not diminish the responsibility of ensuring that examination material is not compromised during transmission to and from examination sites.

Several procedures serve to reduce the possibility of malpractice during the process of examination development. For example, in Uganda, paper setters are required to set individual questions rather than an entire paper. In other cases, several sets of parallel papers may be produced, from which a final one is selected. In Indonesia, starting in 2011, five parallel versions of an examination have been developed and assigned to schools on a random basis. This reduces considerably the potential for leakage as schools cannot know which version will be assigned to them.

Lack of roads and difficult terrain add to the difficulty of distributing examination material in many emerging market economies. To prevent leakage, papers are usually placed in sealed envelopes within metal or wooden boxes. Transit storage facilities include banks, police stations, and army barracks. To further ensure that the security of the process has not been breached, many examination authorities require packages of examination papers to be opened and answer sheets to be sealed in front of candidates. Examination authorities frequently carry out random unannounced visits to examination sites. To guard against examination paper leakage one Indian Board, working with Microsoft, has piloted a technique of encrypted question papers, which examination centers are required to print half an hour before the exam starts (Microsoft News Center India 2018).

Examination authorities should ensure that candidates are informed about the rules governing behavior during examinations and, in particular, that they are aware of the materials that are, and are not, permitted at examination sites (Cambridge Assessment, n.d.). They should also make explicit the consequences of bringing electronic devices, such as headphones, mobile phones, and iPhones, into

examination settings. It is worth bearing in mind that the level of expertise of supervisors in the use of modern technologies is likely to be inferior to that of candidates.

To address the problem of impersonation, candidates should be required to produce valid forms of identification prior to entering the examination center. These forms of identification may include fingerprint- and iris-recognition data if photos of females are not permitted. Practices related to the use of identification information can vary. Sri Lanka requires that candidates use a national identity card or passport. Although India's Aadhaar card, which is a unique 12-digit national identification number (*Hindustan Times* 2018), could have provided an effective means of tackling impersonation as it contains detailed biometric and demographic data, the Indian Supreme Court in 2018 ruled that examination boards cannot insist on candidates using the card.

Steps should be taken to prevent examination questions from being communicated to other candidates in large countries with several time zones, as has happened in Russia (Parfitt 2013), which has 10 time zones. In 2012, a number of Asian Russian students were found to have used the web to assist their colleagues in European Russia (Artemev 2014). This issue also arises when national examinations are offered in different parts of the world, mainly to emigrant students. One possible solution is to require students sitting for identical examination papers to take the examination at the same Greenwich Mean Time worldwide (Heriot Watt University, n.d.).

Adequate training is seldom provided for invigilators. Such training should cover aspects of basic supervision, such as making sure that the distance between candidates' desks is adequate and checking to ensure that material has not been concealed in desks. Supervisors should also learn how to detect information written on parts of the body and on beverage containers or food items, or hidden in items of clothing. If candidates are allowed to use calculators to carry out computational or graphical tasks, invigilators should be able to determine whether the calculators can be used for tasks other than those permitted during the examination.⁵

Examination authorities need to become familiar with the variety of threats posed by modern technological devices. They also need to work with cybersecurity experts to establish multilayered levels of

security and to block the internet, as well as other incoming and outgoing signals, from examination sites. In 2016, Iraqi authorities cut off fixed-line and mobile broadband services during examination hours to discourage children from smuggling mobile phones into state examinations (Gibbs 2016). In 2017, Ethiopia shut down the internet in an effort to prevent examination papers from being released through social media (Dahir 2017). In a similar vein, in 2018, Algerian authorities closed the internet for a period during each examination day to discourage cheating; they also installed surveillance cameras and phone jammers to prevent collusion among candidates (Bradbury 2018). The possibility that draft examination papers might be accessed or that records, including examination marks or grades, will be hacked or wiped out needs to be anticipated.⁶ Examination agencies using connected digital technologies are likely to experience hacking of their online tests or examinations, or their databases (Trucano 2018). Prompted by the selling of fake matric certificates, public and private bodies in South Africa have been advised to contact officially recognized verification agencies (News24Wire 2018).

Severe penalties for corrupt examination officials in seventeenth- and nineteenth-century China included the death penalty, confiscation of property, and exile (Hu 1984). Today, specific penalties are prescribed for different forms of malpractice in many countries on the assumption that the threat of being caught and punished will deter malpractice (Tittle and Rowe 1973). In contemporary China's selection examination for higher education (the *gaokao*), for example, a new law provides for a jail sentence of up to seven years and a ban from taking a national examination for three years for candidates found guilty of malpractice (Hernandez 2016). In Ireland, candidates who have been found guilty of malpractice in public examinations can be banned from taking such examinations for a number of years (Ireland State Examination Commission 2018). In the United Kingdom, candidates who have cheated may have marks deducted or all of their results cancelled (Griffiths and Corke 2018).⁷ Nigeria's Examination Malpractice Act (1999) stipulates a minimum fine plus a maximum of five years imprisonment for various forms of examination malpractice. Lengthy prison sentences are also specified in Kenya and Malawi. In 2017, courts in Bhutan sentenced an official to

two years in prison for “official misconduct” in the form of leaking an examination paper; it ordered the offender to pay the costs of reassessing the answer sheets of the leaked paper (Palden 2017). In 2015, the Egyptian government introduced a law that imposed prison sentences and substantial fines on those who leaked exams through printing, broadcasting, or any other means of publication (Shams El-Din 2015).

Some countries do not have adequate legal provisions for dealing with malpractice. In this situation, examination authorities may have little legal or political backing to support their efforts to punish students or adults. Even where the legal framework is in place, police and judicial authorities are frequently corrupt or for other reasons may show a marked reluctance to enforce sanctions. Where it is apparent that sanctions are too severe or will not be enforced, the law serves little or no useful purpose. Government regulations and strong persistent leadership can contribute to the security of examinations. In Lithuania for instance, printing and delivery of examination papers are accorded high-level security status, which precludes, among other things, printing by commercial printers.⁸

In some countries, the perennial problem of intimidation of markers by candidates and parents was addressed by centralized marking and, even more effectively, by the use of substitute candidate numbers. In the latter case, original candidate identification numbers are replaced by other numbers and a record of the matching numbers is stored on a computer file. At the end of the marking process, the original and substitute numbers are matched.

Modern technology in the form of optical scanners and computers has the potential to contribute to a decrease in the incidence of malpractice. For example, computerized adaptive testing reduces opportunities for malpractice by requiring examinees to take different sets of multiple-choice items based on strengths and weaknesses they have revealed in their performance on earlier items in an examination (though this may give rise to some validity issues). Optical scanners, when used to correct multiple-choice items, decrease the risk of improper interference at the marking or data-entry stages. Computers, because they reduce the number of people who are directly involved in the processing of results, also limit the opportunity for malpractice,

especially at the results-processing stage. To minimize the number of people who have access to results and the possibility of altering results, some examination authorities transfer data files of results directly to admissions officers in third-level institutions or permit these officers to log into results databases.

Examination authorities should take steps to ensure the security of their own offices through installation of closed circuit television cameras, implementation of control systems in operationally sensitive areas (relating to examination preparation, printing, and packing), and prevention of unauthorized personnel from having access to confidential computer files. They might consider engaging the services of a qualified independent expert to identify weaknesses in their current security procedures.

CONCLUSION

The high rewards attached to success in public examinations make malpractice at various stages of the examination process almost inevitable. In most countries, it is probably fair to assert that the majority of candidates do not resort to the use of “unfair means.” However, examination authorities need to be vigilant in their efforts to ensure the integrity of the system as failure to do so can have far-reaching negative consequences. A candidate who knows that success on a public examination can be engineered by bribing or intimidating a marker, by purchasing a copy of the examination paper, or by using internal or external forms of assistance during the course of the examination has little or no incentive to learn. It is fundamentally unfair when a meritorious student is denied access to a scarce place at tertiary level. It is also particularly difficult for a student of high ability from a poor family to devote time and effort to study when he or she perceives that weak candidates from influential backgrounds routinely resort to unethical tactics to “succeed” in an examination. In such cases, far from being perceived as an instrument of equity, the examination system will serve to undermine students’ belief that achievement and effort are rewarded.

Teachers are also affected by malpractice. There will be little incentive for them to work diligently to cover a broad curriculum if they

have prior knowledge of the content of an examination paper or are aware that many of their students will resort to unfair means in the examination. At the national level, formal qualifications that are based on an examination system that lacks credibility will be viewed as being of little value. As a result, users of results, such as university admissions bodies and especially employers, may have to create their own selection procedures.

Modern technology confronts examination officials with a set of new opportunities and problems. On the positive side, computerization, by limiting the extent of human contact with various aspects of the examination process, can help limit the possibility of malpractice. On the other hand, officials are faced with new challenges, including calculator-size instruments that have the capacity to hold vast quantities of information or that can be used for transferring information from one candidate to another. External assistance can be rendered by radio transmitters placed in pens, wristwatches, and pagers, or linked with tiny hearing devices, by personal stereos loaded with prerecorded tapes, and by programmable calculators packed with data. Hacking of examination databases is likely to become a major threat. It should also be appreciated that some strategies for addressing malpractice, such as using blocking devices and other types of technological checks and providing training to examination administrators, are expensive.

No system is ever going to be completely secure. The key to maximizing security is knowing when there has been a cyber intrusion, establishing what the intruders have done, and taking the necessary defensive steps. Tackling real and potential security breaches will require technical expertise with budgetary support. Despite the fact that technical expertise might appear expensive, the price would probably represent a tiny fraction of the cost of a major leak of examination content or data.

While much of the literature to date on educational assessment has focused on technical aspects of measurement instruments, such as validity, reliability, and comparability, malpractice may be considered to represent as serious a threat as any of those aspects to the credibility of the public examination systems. Furthermore, if malpractice is to be adequately addressed, this will involve more than

dealing with largely administrative and technical issues. As examination systems differ greatly from country to country in terms of traditional attitudes toward malpractice,² format, and administrative approaches, educational systems are likely to have to adopt different approaches to tackle their particular problems taking into account local conditions and practices, and identifying potentially successful interventions. Daunting as this task may be, failure to address the threats posed by malpractice in examinations, whatever their origin, is likely to have very serious negative consequences for the student selection and certification process, for teacher behavior, for the quality of student learning, and, in some instances, for the perceived effectiveness of a state agency. Cheating or malpractice “is no more justifiable than telling a sick patient that he is well and then sending him on his way” (Cizek 2001).

NOTES

1. In Thailand, 3,000 students were caught using cameras embedded in their “reading glasses” to transmit examination material (Bradbury 2018).
2. In some emerging market economies, examination success may offer the only lifeline to a government job or to further education.
3. Raisani, Chief Minister, Baluchistan, *Dawn*, June 30, 2010, cited in Burdett 2017, 51.
4. Including Australia, Canada, EU countries, India, Pakistan, South Africa, and the United States.
5. Irish candidates are prohibited from using calculators that support data banks, dictionaries, language translators, text retrieval, graph plotting, equation solving, symbolic algebraic manipulation, numerical integration, numerical differentiation, and matrix calculations.
6. For instance, a process known as “keylogging” can be used to record keys struck on a keyboard by another person without the knowledge or permission of the latter. This process was used successfully by a US college student (subsequently charged) in 2015–16 to intercept exams and test questions in advance and change his and others’ grades on tests, quizzes, and homework assignments more than 90 times (Vas 2017).
7. Teachers deemed guilty of malpractice received more lenient sanctions.
8. A. Zabullinos, personal communication, November 14, 2016.

9. In some situations, allowing another examination candidate to copy answers is likely to be considered cheating, while in other cultures it might be regarded as a required form of altruism, a case of simply helping a friend (Burton 1976).

REFERENCES

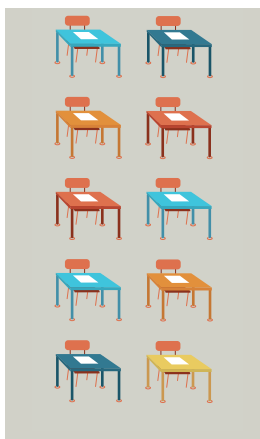
- Artemev, N. 2014. "The Problems of the Unified State Examination in Russia." *Education Policy Talk: International Perspectives and Debates* (blog), May 23. <https://educationpolicytalk.com/2014/05/23/the-problems-of-the-unified-state-examination-in-russia/>.
- Bradbury, D. 2018. "Internet Shut Down in Algeria to Stop Exam Cheats." *Naked Security* (website), June 25. <https://nakedsecurity.sophos.com/2018/06/25/internet-shut-down-in-algeria-to-stop-exam-cheats/>.
- Burdett, N. 2017. "Review of High Stakes Examination Instruments in Primary and Secondary School in Developing Countries." Research on Improving Systems of Education (RISE) Working Paper 17/018, RISE, Oxford, UK. <https://www.riseprogramme.org/publications/rise-working-paper-17018-review-high-stakes-examination-instruments-primary-and->
- Burton, R. V. 1976. "Honesty and Dishonesty." In *Moral Development and Behavior: Theory, Research and Social Issues*, edited by T. Lickona, 173–97. New York: Holt, Rinehart and Winston.
- Cambridge Assessment. n.d. "Information for Candidates," accessed May 21, 2019. <http://www.cambridgeinternational.org/Images/86457-information-for-candidates.pdf>.
- Cizek, G. J. 2001. "Cheating to the Test: What to Do about It." *Education Matters* 1 (1). <https://www.educationnext.org/cheatingtothetest/>.
- Dahir, A. L. 2017. "Blocked Again: Ethiopia Shut Down the Country's Internet to Beat Exam Cheats." *Quartz Africa* (website), May 31. <https://qz.com/africa/994990/ethiopia-shut-down-the-internet-ahead-of-a-scheduled-countrywide-national-exams>.
- Dawn. 2010. "Fake Degree Scandal Roils Pakistani Politics." *Dawn*, June 30. <https://www.dawn.com/news/958083>.
- De Birhaner. 2016. "Ethiopian ESLCE Leaving Cert Exams Reportedly Leaked." *De Birhan*, May 30.
- Denisova-Schmidt, E., and E. Leontyeva. 2014. "The Unified State Exam in Russia: Problems and Perspectives." *International Higher Education* 76: 22–23.

- Dorff, V. 2013. "Cheating [Is] the System." *Huffington Post*, June 24. http://www.huffingtonpost.com/victor-dorff/cheating-is-the-system_b_3488509.html.
- Foster, D. 2015. "The Language of Security and Test Security." Caveon Test Security, Midvale, UT. <https://www.caveon.com/2015/06/24/the-language-of-security-and-test-security-2/>.
- Gibbs, S. 2016. "Iraq Shuts Down the Internet to Stop Pupils Cheating in Exams." *Guardian*, May 18. <https://www.theguardian.com/technology/2016/may/18/iraq-shuts-down-internet-to-stop-pupils-cheating-in-exams>.
- Greaney, V., and T. Kellaghan. 1996. "The Integrity of Public Examinations in Developing Countries." In *Assessment: Problems, Developments and Statistical Issues*, edited by H. Goldstein and T. Lewis, 167–88. Chichester, UK: Wiley.
- Griffiths, S., and J. Corke. 2018. "Thousands of Teachers Caught Cheating in Exams." *Sunday Times*, February 11. <https://www.thetimes.co.uk/article/thousands-of-teachers-caught-cheating-in-exams-j9vcxr5kj>.
- Haladyna, T., S. Nolen, and N. Haas. 1991. "Raising Standardized Achievement Test Scores and the Origins of Test Score Pollution." *Educational Research* 20: 2–7.
- Heriot Watt University. n.d. "Policy on Examination and Other Assessment Held around the World in Different Time-Zones." Heriot Watt University, Edinburgh. <https://www.hw.ac.uk/services/docs/learning-teaching/policies/timezonesexamsched.pdf>.
- Hernandez, J. C. 2016. "China Threatens Jail for College Entrance Exam Cheaters." *New York Times*, June 7, A8. <https://www.nytimes.com/2016/06/08/world/asia/china-exam-gaokao-university-cheating.html>.
- Hindustan Times*. 2018. "Aadhaar Not Needed for Neet, UGC, CBSE and School Admissions." *Hindustan Times*, September 26. <https://www.hindustantimes.com/india-news/aadhaar-not-needed-for-neet-ugc-cbse-and-school-admissions/story-F8mm2labwBrQJ5O93vEQdN.html>.
- Hu, C. T. 1984. "The Historical Background: Examinations and Control in Pre-modern China." *Comparative Education* 20: 7–26.
- Ireland State Examination Commission. 2018. "General Candidate Information: Conduct of Candidates during Examinations." Ireland State Examination Commission, Athlone. <https://www.examinations.ie/index.php?l=enandmc=caandsc=gy>.
- Kuehn, P., D. J. Stanwyck, and C. L. Holland. 1990. "Attitudes towards 'Cheating' Behaviours in the ESL Classroom." *TESOL Quarterly* 24: 313–17.
- Magnus, J. R., V. M. Polterovich, D. L. Danilov, and A. V. Savvateev. 2002. "Tolerance of Cheating: An Analysis across Countries." *Journal of Economic Education* 33: 125–35.

- Maynes, D. 2008. "Moore's Law Favors the Cheater." Caveon Security Insights, January 21. Caveon Test Security, Midvale, UT. <https://www.caveon.com/2008/01/21/moores-law-favors-the-cheater/>.
- . 2009. "Caveon Speaks Out on IT Exam Security: The Last Five Years." Caveon Test Security, Midvale, UT.
- Medina, J., K. Benner, and K. Taylor. 2019. "Actresses, Business Leaders and Other Wealthy Parents Charged in U.S. College Entry Fraud." *New York Times*, March 12. <https://www.nytimes.com/2019/03/12/us/college-admissions-cheating-scandal.html>.
- Mehrens, W., and J. Kaminski. 1989. "Methods for Improving Standardized Test Scores: Fruitful, Fruitless, or Fraudulent?" *Educational Measurement: Issues and Practice* 8: 14–22.
- Microsoft News Center India. 2018. "CBSE Partners with Microsoft to Provide Encrypted Question Papers for Class X and XII Examinations." August 13. <https://news.microsoft.com/en-in/cbse-partners-with-microsoft-to-provide-encrypted-question-papers-for-all-class-x-and-xii-examinations/>.
- Moore, M. 1993. "Indian Children Cheat to Survive." *Washington Post*, April 17, A14, A20.
- News24Wire. 2018. "Umalusi Exposes Websites Selling Fake Matric Certificates." *AllAfrica*, July 2. <https://allafrica.com/stories/201807020748.html>.
- Nigeria. 1999. "Examination Malpractices Act." LawNigeria.com. <http://www.lawnigeria.com/LawsOfTheFederation/EXAMINATION-MALPRACTICES-ACT.html>.
- Onyibe, C. O., U. U. Uma, and E. Ibina. 2015. "Examinations Malpractice in Nigeria: Causes and Effects on National Development." *Journal of Education and Practice* 6: 12–17.
- Palden, T. 2017. "Official Gets 2 Years in Prison for Leaking Exam Paper." *Kuensel*, January 13. <http://www.kuenselonline.com/official-gets-2-years-in-prison-for-leaking-exam-paper/>.
- Parfitt, T. 2013. "Russia Hit by Exam Cheating Epidemic." *Telegraph*, May 28. <http://www.telegraph.co.uk/news/worldnews/europe/russia/10084299/Russia-hit-by-exam-cheating-epidemic.html>.
- Patton, G. 2010. "More Pupils Cheating in Exams, Says Ofqual." *Telegraph*, February 3. www.telegraph.co.uk/education/educationnews/7139336/More-pupils-cheating-in-exams-says-Ofqual.html.
- Plagiarism.org. 2017. "Plagiarism: Facts and Stats: Academic Integrity in High School." June 7. <https://www.plagiarism.org/article/plagiarism-facts-and-stats>.

- Schaeffer, B. 2014. "Confirmed Cases of Test Cheating, 2009–2014." Jamaica Plain, MA: National Center for Fair and Open Testing. <http://www.fairtest.org/sites/default/files/CheatingReportsList.pdf>.
- Shams El-Din, M. 2015. "New Law Imposes Prison Sentences for Leaking Exams." *Mada Masr*, October 9.
- Tittle, C. R., and A. R. Rowe. 1973. "Moral Appeal, Sanction, Threat and Deviance: An Experimental Test." *Social Problems* 20: 488–97.
- Trucano, M. 2018. "What to Do after Your Education System Is Hacked." *EduTech: A World Bank Blog on ICT Use in Education*, July 27. <https://blogs.worldbank.org/edutech/what-do-after-your-education-system-hacked>.
- Uganda National Examinations Board. 2018. "Examination and Security in the Management of Examinations Is My Responsibility." Press statement, October 11. <https://uneb.ac.ug/downloads/2018>.
- Vanguard*. 2013. "JAMB Releases 2013 UTME Results." May 3. <https://www.vanguardngr.com/2013/05/jamb-releases-2013-utme-results-withholds-12110-results/>.
- Vas, L. 2017. "Student Charged by FBI for Hacking His Grades More Than 90 Times." *Naked Security* (website), November 1. <https://nakedsecurity.sophos.com/2017/11/01/student-charged-by-fbi-for-hacking-his-grades-more-than-90-times/>.
- Walsh, D. 2015. "Fake Diplomas. Real Cash. A Net of Made-Up Schools." *New York Times*, May 18, A1.

CHAPTER 12



PUBLIC EXAMINATIONS PROVISION FOR CANDIDATES WITH DISABILITIES OR DIVERSE EDUCATIONAL NEEDS

INTRODUCTION

Few areas of education have experienced such rapid change in recent years as students with disabilities or diverse educational needs. Although definitions of educational disabilities and special needs vary, they tend to include students with physical, sensory, cognitive, intellectual, and psychosocial disabilities. In the past, the population of young people who were considered disabled or handicapped tended to be segregated from the mainstream. Many in fact did not attend school; in 1987, only 55 percent of Chinese students with disabilities were in school compared to an overall Chinese enrollment rate of 97 percent (McCabe 2003). In the United States, as recently as 1975, more than 1 million such children did not attend school, and only about 20 percent were educated in public schools. Many states did not accept into public schools students who were blind or deaf or who suffered from other impairments (US National Council on Disability 2000). Data presented at the 2018 Global Disability Summit suggested that more than half of the 65 million children with disabilities in low- and middle-income countries are not in school (Global Disability Summit 2018).

In the final decades of the twentieth century, the concept of inclusive schooling gained ground and was reflected in a recognition in many educational systems and by international organizations (UN 1993; UNESCO 1994) that persons with disabilities or diverse educational needs should become part of the regular school system. This resulted in many countries enacting equality legislation. International bodies are increasingly prioritizing support for students with disabilities and special educational needs. For example, at the Global Disability Summit, the World Bank committed to ensuring all World Bank education projects and programs would be disability-inclusive by 2025 (Anders 2018). Against this background, it is not surprising to find that the challenge to public examination authorities to assess such students in national public examinations has become much more pronounced in recent years.

One of the strengths attributed to public examinations is that all candidates are treated in the same way. However, in recent years concern that the circumstances of candidates with special needs should be taken into account has been increasing, as they may be prevented from demonstrating their achievements by features of the examinations. This has led to accommodations of examination procedures in certain circumstances to allow these candidates opportunities to demonstrate their knowledge and skills in examination situations that might differ from those of most other candidates (see box 12.1). The challenge is to ensure that any accommodation that is provided does not affect the integrity or security of the assessment.

BOX 12.1**Testing Accommodations**

Testing accommodations have been defined as “changes to the regular testing environment and auxiliary aids and services that allow individuals with disabilities to demonstrate their true aptitude or achievement level on standardized exams or other high-stakes tests.”

Source: US Department of Justice 2014, 2.

In many (but not all) examination systems, candidates with permanent or long-term conditions, including hearing and visual difficulties, or specific learning difficulties they believe will significantly impair performance in the examinations, can apply to the examinations authority for “reasonable accommodation(s)” or adjustment to facilitate their taking the examination (see, for example, Ireland Expert Advisory Group on Certificate Examinations 2000; UK Joint Council for Qualifications 2015). The accommodations are intended (a) to remove, as far as possible, the impact of the disability on the candidate’s performance and (b) to ensure that the arrangements do not give a candidate an unfair advantage over other candidates (Ireland State Examinations Commission 2010, 2019).

IDENTIFICATION OF STUDENTS WITH DISABILITIES OR SPECIAL NEEDS AND IMPLICATIONS OF ACCOMMODATION

Independent evidence of a disability and support needs is normally required before allowing special arrangements. However, the regulations regarding the evidence required of a candidate’s special needs vary from one examination system to another. In some, a score on a standardized test of reading accuracy or comprehension (84 or less) is required if a reader is to be provided (for example, UK Joint Council for Qualifications 2015). However, this procedure will not be available in countries that do not have standardized norm-referenced tests. In some systems, students are required to provide appropriate documentation to support their request for accommodation (see box 12.2).

The final determination of who can be considered a candidate with special needs may not always rest with the examination authority. The US Americans with Disabilities Act specifies that

reports from qualified professionals who have evaluated the candidate should take precedence over reports from testing entity reviewers who have never conducted the requisite assessment of the candidate for diagnosis and treatment. This is especially important for individuals with learning disabilities because face-to-face interaction is a critical component of an accurate evaluation, diagnosis, and determination of appropriate testing accommodations. (US Department of Justice 2014, 7)

BOX 12.2**Document Requirements for Candidates with Special Needs**

All applications should be supported by an up-to-date medical/assessment report, valid throughout the examination period, issued by the Hospital Authority, Department of Health, Education Bureau, a registered medical practitioner, or a professional (for example, audiologist, speech therapist, or psychologist), in accordance with the requirements stated in the Application Guide. Schools should provide information on the special arrangements made for the candidates in the school internal examinations. Recommendations on the special arrangements in the public examinations should also be stated on the application forms.

Source: Hong Kong Examinations and Assessment Authority 2016.

The precise arrangements to be made should be determined on the basis of the disability or impairment established in each individual case and of the particular needs of the candidate in each individual subject area. It is recognized that different subjects and different methods of assessment may make different demands on candidates.

Schools (or examination centers) usually are assigned the primary responsibility of taking the lead in arranging access within their schools or centers. This is because the means by which a candidate normally studies or communicates will largely determine the accommodations that will be allowed. Schools should brief examiners on any dietary or medical needs a candidate may have and keep a record of such briefing. Furthermore, the school or center is usually responsible for (a) determining and implementing the arrangements following prescribed guidelines, (b) ensuring that approved access arrangements are in place for internal school assessments or examinations, and (c) processing applications and holding evidence for possible inspection.

EXEMPTIONS

A candidate's disability or special educational needs may be such that it is not possible for him or her to participate in a particular mode of assessment (such as an aural examination for a candidate with severe

hearing impairment), in which case it should be open to the candidate to apply for exemption from part of the assessment procedure.

When an element or elements of an examination have been waived, so that the purpose of an examination has not been met, or the method of examining has been significantly altered, the grade awarded is determined by the work that the candidate has completed. Details of the content or tasks that have been exempted may or may not be reported. In Greece, where students with visual problems are permitted to take oral tests and students with hearing impairments take written tests, the accommodations are noted on certificates of achievement (Pepper 2007). In South Africa, an Endorsed National Senior Certificate is awarded to successful learners who have registered as having special education needs (South Africa, Umalusi, Council for Quality Assurance in General and Further Education and Training 2013). In Hong Kong SAR, China, details of exempted components are specified, but information on reasons for exemption are not (Hong Kong Education Bureau 2017). In the United States, on the other hand, the release of such information is expressly forbidden on the grounds that it suggests a candidate may be disabled and might impede his or her chances of employment or of pursuing further education (US Department of Justice 2014).

TYPES OF ACCOMMODATION

The provision of accommodations in examination procedures has allowed special needs students, across a range of disabilities, to demonstrate their achievement in high-stakes public examinations. Accommodations relate to the settings in which examinations are administered, the format of examination presentation, methods of responding, and amount of time permitted to complete examination questions or tasks (Florida Department of Education, Bureau of Exceptional Education and Student Services 2015; Hong Kong Education Bureau 2017; Pepper 2007; Scottish Qualifications Authority 2017; UK Joint Council for Qualifications 2015).

Setting

Special provision can involve a variety of situations or settings (see box 12.3).

Presentation

The format and nature of presentation of examination materials will vary depending on the nature and severity of a candidate's particular disability.

- Candidates with visual issues may be offered Braille versions of papers (in addition to the normal versions), screen readers, questions presented in enlarged print, answer sheets with wide spaces between lines, calculators with talking or large display features, and proctors who read aloud individual questions.
- Candidates with hearing issues may use aids such as digital audio players, commonly termed MP3 players, and have a proctor such as a teacher who can relay the examiner's instructions to assist with the administration of the examination in individual or small group sessions.
- Modified questions may be used when diagrams, maps, and photographs are involved to make the content more accessible.

Response Methods

The methods used to respond to examination questions will vary according to the type and level of student disability or educational need. Box 12.4 outlines some of the response accommodations that might be considered.

BOX 12.3

Special Provisions for Special Needs Candidates

- Seating the candidates along with other candidates near or facing the examiner; permitting use of a digital audio player and headphones.
- Conducting the examination individually or in small groups in a separate room or in a hospital where a proctor or support person may read aloud the question papers and record responses.
- Establishing conditions that will make it easier for candidates with special needs to respond, such as special lighting, large desks, or cubicles to allow the use of examination papers set in Braille.

Source: Adapted from Ireland State Examinations Commission 2019.

Timing and Scheduling

Additional time or special scheduling may be granted to help eliminate the role of speed in testing. Arrangements vary from jurisdiction to jurisdiction. The examination authority in Hong Kong SAR, China, has suggested that the normal time allowance for answering questions be extended by 25 percent, which can be adjusted on the basis of professional advice, while in Ireland 10 minutes additional time per hour is recommended in cases in which a candidate is visually impaired or needs the help of a scribe or proctor or in which he or she would otherwise be unable to make adequate use of the mechanical aids provided for recording the answers. Large standardized testing programs have accommodated students with disabilities, especially those with visual impairments. In the United States, formally approved visually impaired students can get 50 percent additional time when taking the SAT if large print format is used (Family Connect, n.d.).

Examination authorities should be aware of the risk that the provision of additional time for special-accommodation students might introduce an element of bias into the system. A 2019 report reveals that students from more affluent US high schools are somewhat more

BOX 12.4

Accommodations for Special Needs Candidates

- An examiner enters the necessary identification information on the candidate's papers, answer sheets, or envelopes.
- The candidate dictates or points to a response, uses a Braille writer, or signs responses to a sign language specialist.
- The candidate responds to multiple-choice questions by writing, circling responses on the question papers (answers to be transferred later by the proctor to the standard answer sheet), or fills in large bubbles or boxes on specially designed answer sheets.
- An examiner or proctor transfers answers to answer sheets or records dictated notes and essays.
- A visually impaired candidate records answers on a digital recorder or uses a computer word processor with large keys to type responses.
- An examiner or proctor records responses in cases such as unclear speech.

Source: Adapted from US Department of Justice 2014.

likely to get special time allowances for tests such as the SAT. Schools determine whether students qualify for special accommodation, often following an external medical evaluation. More affluent school systems and parents are more likely to know of the additional-time option and to be able to afford the associated medical costs (Belkin, Levitz, and Korn 2019).

A guide for test developers recommends that testing of the visually impaired allow for extended time of specific duration, such as 2.5 times for Braille readers and 1.5 times for large print readers (Allman 2009). Other time and scheduling accommodations may include the following:

- Frequent breaks
- Additional time for oral assessment, reading text, and communicating responses in the case of candidates with reading and writing disabilities
- Pronounced pauses between sentences in oral assessments
- Prolonged pauses between sentences in audio presentations in aural assessments

Examination authorities, rather than adhering strictly to rigid guidelines when determining additional time for various forms of disabilities, might also take into consideration the nature and difficulty of the examination questions.

A postexamination adjustment may be made to a candidate's mark or grade to reflect temporary injury, illness, or other indisposition at the time of the examination or assessment. In some examination systems, information from school-based assessments will be available.

ISSUES ARISING FROM THE PROVISION OF ACCOMMODATIONS FOR DISABILITIES OR DIVERSE EDUCATIONAL NEEDS

In this context, accommodations refer to special arrangements, compensations, or conditions that an examination agency may permit for candidates with disabilities or special needs. Allowing such candidates various accommodations during high-stakes public examinations gives rise to a range of concerns. Table 12.1 presents examples of such

TABLE 12.1**Issues Arising from Administering Examinations to Students with Disabilities or Diverse Needs**

Concerns	Possible responses
Students may not have access to professional personnel to testify to their levels of visual or aural impairment or to the nature of other physical or psychological disabilities.	School principal and examination or ministry official use formal protocols to identify nature of disability some months ahead of the examination.
Examination administrators may be unaware that students have to be accommodated.	Examination officials should identify such students well in advance, inform administrators or proctors, and ensure that special arrangements and facilities have been put in place.
Examination administrators may be unfamiliar with the requirements of candidates with disabilities or diverse educational needs.	Administrators should be given specific training prior to the examination. Aides and special education teachers familiar with the types of accommodation needed should give logistical assistance.
Accommodations can disadvantage poor students in cases in which all or a substantial part of the examination fee is borne by the student's family.	Government subsidizes additional costs.
In examinations that require computer use, students may not be familiar with particular computer types and programs, accessing the internet, or using files and software.	Students should be familiar with these technologies and programs and should have used them beforehand as part of the learning process.

(continued)

TABLE 12.1

Issues Arising from Administering Examinations to Students with Disabilities or Diverse Needs (continued)

Concerns	Possible responses
Some students may be unable to concentrate or become highly stressed.	Conduct the examination in the morning. Allow breaks if needed. Do not to permit the student to interact with others taking the same examination paper.
Examination scores might not be valid, as possibilities for malpractice are increased, such as allowing a student to answer fewer examination questions, paraphrase questions, offer additional information, or alter student responses.	Advise examination aides or proctors that examination rules and penalties for malpractice apply. Video record sessions. Examination officials conduct random visits.

concerns and possible responses from examination or school authorities (Salmon, Smarick, and Hogan 2017).

As modern technology develops, some problems may become less pronounced. For example, the requirement to have a proctor record an examinee’s message or transcribe it may diminish when speech-to-text software is available to input responses.

CONCLUSION

Changing attitudes toward disability and special needs, concerns about equality, and increasing participation rates in formal education have combined to draw attention to the need to alter examination procedures to meet the special needs of some students. Relevant issues have been largely addressed in developed education systems. In some emerging market economies, concern about challenges posed by special needs students taking public examinations may not, given other pressing demands, have received much attention. While systems in which participation rates are low may not yet have come to

terms with these issues, they should be prepared to deal with them as participation rates increase.

A variety of responses will be required of examination bodies in establishing the veracity of claims of students seeking accommodation, as access to medical, psychological, and other support services will vary greatly depending on location and level of economic development. Although many relevant issues have been addressed, problems remain, such as additional costs posed by special needs examination candidates, establishing appropriate examination conditions, limiting opportunities for malpractice, and estimating how much additional examination time might be allowed. Examination systems are likely to benefit from sharing the administrative, measurement, legal, and financial lessons learned from their experiences when addressing the unique challenges posed by examination candidates with disabilities or special needs.

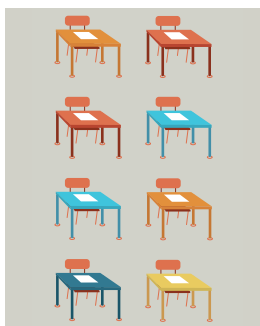
REFERENCES

- Allman, C. 2009. *Making Tests Accessible for Students with Visual Impairments: A Guide for Test Publishers, Test Developers, and State Assessment Personnel*. 4th ed. Louisville, KY: American Printing House for the Blind. <http://www.aph.org>.
- Anders, M. 2018. "Global Disability Summit Yields 170 New Commitments—But What's Missing?" Devex. <https://www.devex.com/news/global-disability-summit-yields-170-new-commitments-but-whats-missing-93172>.
- Belkin, D., J. Levitz, and M. Korn. 2019. "Many More Students, Especially the Affluent, Get Extra Time to Take the SAT." *Wall Street Journal*, May 21.
- Family Connect. n.d. "Taking the SAT or ACT as a Student Who Is Blind or Visually Impaired." <http://www.familyconnect.org/info/browse-by-age/teenagers/education-teenagers/taking-the-sat-or-act/1235>.
- Florida Department of Education, Bureau of Exceptional Education and Student Services. 2015. *Accommodation of Florida's Statewide Student Assessment*. Tallahassee, FL: Department of Education.
- Global Disability Summit. 2018. *Inclusion in Education*. London: International Disability Alliance and Department for International Development. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/725745/DFIDHandouts_InclusionInEducation_V3.pdf.

- Hong Kong Education Bureau. 2017. "Special Arrangements for Internal Examinations for Students with Special Educational Needs." Hong Kong Education Bureau, Hong Kong SAR, China.
- Hong Kong Examinations and Assessment Authority. 2016. "Providing Services to Candidates with Special Needs." Hong Kong Examinations and Assessment Authority, Hong Kong SAR, China.
- Ireland, Expert Advisory Group on State Examinations. 2000. "Arrangements for the Assessment of Candidates with Special Needs in Certificate Examinations." Report to the Minister for Education and Science, Expert Advisory Group on State Examinations, Athlone.
- Ireland, State Examinations Commission. 2010. "Arrangements for Candidates with Physical, Visual, Hearing, and Learning Difficulties." State Examinations Commission, Athlone.
- . 2019. "Reasonable Accommodations at the 2019 Certificate Examinations: A Guide for Students." State Examinations Commission, Athlone. <https://www.examinations.ie/schools/BI-1013-9280184.pdf>.
- McCabe, H. 2003. "The Beginnings of Inclusion in the People's Republic of China." *Research and Practice for Persons with Severe Disabilities* 28: 16–22.
- Pepper, D. 2007. *Assessment for Disabled Students: An International Comparison*. London: UK Qualifications and Curriculum Authority.
- Salmon, K. B., A. Smarick, and L. Hogan. 2017. "Maryland Assessment, Accessibility, & Accommodations Policy Manual: Selecting, Administering and Evaluating the Use of Accommodations for Instruction and Assessment." Maryland State Department of Education, Baltimore. <http://marylandpublicschools.org/programs/Documents/Special-Ed/IEP/MAM508102017.pdf>.
- Scottish Qualifications Authority. 2017. "Assessment Arrangements Explained: Information for Centres." Scottish Qualifications Authority, Glasgow.
- South Africa, Umalusi, Council for Quality Assurance in General and Further Education and Training. 2013. *Directives for Certification: National Senior Certificate (Schools)*. Pretoria, South Africa: Umalusi. https://www.umalusi.org.za/docs/directives/2013/directives_nsc.pdf.
- UK Joint Council for Qualifications. 2015. "Adjustments for Candidates with Disabilities and Learning Difficulties: Access Arrangements and Reasonable Adjustments; General and Vocational Qualifications with Effect from 1 September 2015 to 31 August 2016." Joint Council for Qualifications, London.

- UNESCO (United Nations Educational, Scientific and Cultural Organization). 1994. "The Salamanca Statement and Framework for Action on Special Needs Education." UNESCO, Paris.
- United Nations. 1993. "The Standard Rules on the Equalization of Opportunities for Persons with Disabilities." Resolution 48/96, adopted by the United Nations General Assembly, 48th session, United Nations, New York.
- US Department of Justice. 2014. *Americans with Disabilities Act: Testing Accommodations*. Washington, DC: Department of Justice.
- US National Council on Disability. 2000. *Back to School on Civil Rights: Advancing the Federal Commitment to Leave No Child Behind*. Washington, DC: National Council on Disability. ERIC doc. no. 438632.

CHAPTER 13



THE ROLE OF SCHOOL-BASED ASSESSMENT

INTRODUCTION

Public examinations and internal assessments carried out by teachers have different objectives, procedures, and consequences. For this reason, the two activities have been kept strictly apart in many jurisdictions. However, there also has been a belief that there are ways in which aspects of external (public) and internal (school-based) assessments might be combined in the interest of increased equity, effectiveness, or efficiency. The school-based component is perceived to enhance the validity of the overall public examination by closely aligning curriculum and assessment and assessing student outcomes that cannot be measured using the format of external examinations. School-based assessment (SBA) requires teachers to mark their students' work, and normally involves cooperation among teachers and schools to help ensure comparability of marking criteria.

In the case of public examinations, relationships between external and internal assessments can take a number of forms. In one, the external and internal components interact throughout the examining process. Teachers may propose examination questions (from which the examination authority selects the questions posed in the examination) and again play a role in marking the scripts of

their own students. An example of this can be found as far back as 1862, when in the school examinations set by the University of Cambridge Local Examinations Syndicate and the University of Oxford Delegacy of Local Examinations (see chapter 3), teachers in the examined school were responsible for marking their students' scripts, a situation that is not very different from what happens in the German Abitur. In an alternative (and more common) arrangement, external and internal assessments are independent operations, and the judgments of each assessment of student achievement are subsequently merged to determine the final overall examination mark or grade. The results of the internal or school-based assessment may be adjusted in light of some form of external moderation (including students' performance on the external assessment).

STRATEGIES TO REDUCE THE BURDEN OF EXAMINATIONS

Many efforts at public examination reform over the years that were directed at reducing the burden of examinations on students, while improving the validity of the assessments, have involved proposals regarding SBA. Three broad strategies to achieve these goals can be identified: (a) the abolition of an examination at a particular point in the education system, (b) the allocation of a percentage of examination marks to SBA, and (c) basing certification entirely on SBA.

A reduction in the number of public examinations was achieved in many countries through their abolition at the end of primary schooling, prompted in great part by the fact that they were no longer needed to certify students' achievement or select them for secondary education. Examinations at the end of the junior cycle of secondary school have also been abolished in some countries. Three responses to the abolition of public examinations can be identified. In one, schools-based certification is proposed as an alternative; schools themselves certify that students have achieved the desired standards. In a second response, the results of a census-based national assessment—that is, an external assessment involving nearly all schools (Greaney and Kellaghan 2008)—are used to serve many of the functions of a public

examination (for example, student promotion, school evaluation) following its abolition. This has been the case in The Gambia. The third and most common response to discontinuation of a public examination is to discontinue certification and the other functions served by that examination as well.

Abolition of public examinations at the end of secondary school is less common than abolition at earlier stages in the education system. In fact, the number of education systems that have introduced external public examinations at the end of secondary school in recent years outweighs the number that have cut back on external examining at this point (see chapter 3). However, in several education systems that retain a public examination, procedures are in place or proposals exist to allocate a percentage of the examination marks to assessments carried out by students' own teachers. Not surprisingly, there is wide variation in the nature of this SBA.

A third possible strategy directed toward reducing the burden of external examinations is to base certification entirely on SBA. Certification in some countries (such as Australia, Canada, New Zealand, Scotland, and South Africa), as well as for the International Baccalaureate, relies to a considerable extent on SBA, and in a few systems it is entirely school based (Grima 2003). In the United States, rules for high school graduation vary by educational jurisdiction. Assessments are usually developed and marked by the students' teachers; high schools award graduation diplomas after students pass a minimum set of required courses. Some states have made passing an additional state exit exam a graduation requirement. Students wishing to attend third-level educational institutions also take externally administered aptitude or achievement tests, the results of which are taken into consideration along with the student's high school achievement record.

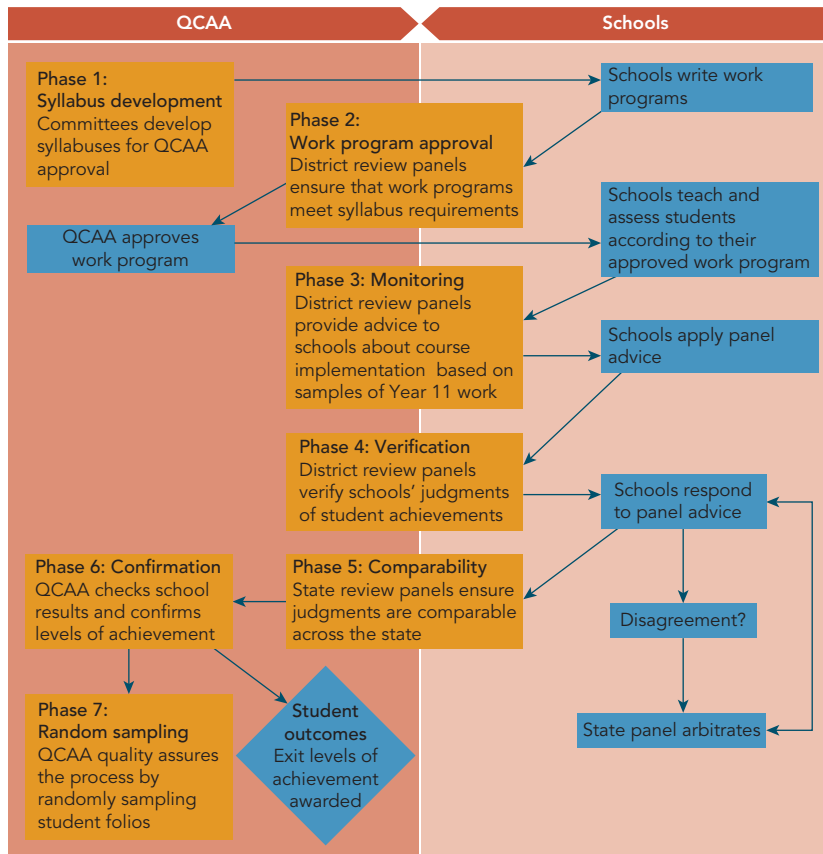
The Australian state of Queensland has based high school graduation solely on the results of SBA (Queensland Curriculum and Assessment Authority 2014). The development of the system, in place since the 1970s, required considerable negotiation with teachers and their representatives (Clarke 1987); professional development, especially in the area of assessment; formal evaluations of its effectiveness (Masters and McBryde 1994); and strong leadership

from the assessment authority, the Queensland Curriculum and Assessment Authority.

Central to the system is an acceptance of the ideas that an examination at the end of secondary school should be an assessment of a student's work in upper secondary school (grades 11 and 12) based on the curriculum and that teaching should amount to much more than "teaching to the test." Teachers have the freedom to develop a curriculum suited to local needs and may use a wide variety of techniques to assess their own students' performance against sets of approved standards based on objectives in the official syllabus (Campbell 2018). Subject panels review and endorse curriculum and assessment plans before students complete the assessments. To ensure a standard level of marking, the assessment authority supports external review and confirmation panels consisting of groups of teachers from around the state, which look at samples of student work and may adjust student grades. Confirmation panels try to ensure that assessment judgments are consistent and comparable across different teachers, programs, and schools (Maxwell 2002). Figure 13.1 summarizes the external moderation process and identifies the specific roles played by the curriculum and assessment authority and by individual schools in the process (Queensland Curriculum and Assessment Authority 2018).

Partly in response to concern over aspects of SBA, starting in 2019, external assessments are to be introduced in Queensland. Results will be based on three internal and one external assessment and will account for 25 percent of subject marks, except in the case of general mathematics and science subjects, in which they will account for 50 percent (Campbell 2018). The external assessment will be quite similar to that currently used in many public examinations systems; it will be administered on the same day throughout the state and will be developed and marked by the curriculum and assessment authority using common marking schemes.

At the end of the process, successful students receive the Queensland Certificate of Education, an official statement that contains a summary of skills and knowledge acquired during upper secondary school. For students seeking entry into tertiary institutions, Queensland Certificate of Education marks are adjusted, through a scaling process

FIGURE 13.1**Queensland: External Moderation Process**

Source: Queensland Curriculum and Assessment Authority 2018. © State of Queensland (Queensland Curriculum and Assessment Authority). Reproduced with permission from the Queensland Curriculum and Assessment Authority; further permission required for reuse.
 Note: QCAA = Queensland Curriculum and Assessment Authority.

based on a sole common test administered to all schools, to give an Overall Position, or rank, which is used by universities and colleges of technical and further education to select students. Beginning in 2020, the Overall Position will be replaced with an Australian Tertiary Admission Rank, bringing it into line with other states and territories (Queensland Curriculum and Assessment Authority 2017).

At present it seems that education systems with strong external examination traditions are reluctant to embrace broad-scale SBA,

although there may be an expectation (or aspiration) that school-based weighting would increase over time.

DIMENSIONS OF SCHOOL-BASED ASSESSMENT

School-based assessment systems as part of a public examination vary in a number of respects (Darling-Hammond and McCloskey 2008; Griffith 2015; Satterly 1981).

Focus on Practical Aspects of Student Performance

In some systems, the school-based element is limited to areas in which a terminal external examination is inappropriate, for example, the assessment of practical or project work or oral fluency, which requires judgment to be made at the point of performance. Or it could be more general and overlap with elements of the external assessment. Many examination systems include practical assessments of the type described in boxes 13.1 and 13.2.

Degree of Formality

Systems differ in the degree to which assessments are based on formal or informal procedures. In informal assessments, teachers’

BOX 13.1

External and School-Based Components for Science

Theory papers: written tests set and marked externally	+	Practical assessment tasks conducted in schools and assessed by teachers according to prescribed criteria	=	Final score derived from theory and practical components
---	---	---	---	--

Source: World Bank 2001.

BOX 13.2**External and School-Based Components for Foreign Language**

Written tests of, for example, grammar, usage, writing, reading comprehension, set and marked externally	+	Listening test: audio tape recorded externally and responses, marked externally	+	Speaking test: conducted in schools and assessed by teachers according to prescribed criteria	=	Final score
--	---	---	---	---	---	-------------

Source: World Bank 2001.

interpretations of observations are made in the context of everyday classroom activity, such as paying attention (or not) in class, working with other students, or carrying out an experiment in science. Teachers may draw on the same sources about students when assessing them for a public examination as when assessing them in the day-to-day routine of instruction. However, when SBA is incorporated into a public examination system, the stakes associated with the assessment obviously go up. This gives rise to the need to document the performance of each student more thoroughly and more formally than is the case in a low-stakes classroom assessment. It may also be necessary for teachers to retain the marking scheme to be used by external monitors of teachers' ratings of student school-based work (Kane 2006).

Formal assessments, in contrast to informal assessments, are often carried out during a period set aside with the specific aim of evaluating students' achievements and competencies. These include students' performance in written exercises, tests (for example, of arithmetic or spelling), assignments, projects, practical work, homework, orals, aural, group discussion, or portfolios. A variety of assessment

techniques may be permitted, giving teachers considerable freedom in choosing what to assess and how to assess it. In some education systems that have assigned a role to SBA in their public examinations, the school-based element is limited to student performance on formal tests or written examinations. It could be argued that such SBA tests or exams can and should serve a formative role as they offer the possibility of timely and relevant feedback to students (Black et al. 2003; Carless 2011); however, in many instance the written SBAs may merely mimic (and add little to) the external component of the examination.

Formal assessment procedures can be differentiated in terms of control. In high-control situations, the awarding body is responsible for task setting. In medium-control situations, the school may select from a number of options provided by the awarding authority. In low-control situations, schools design their own tasks following specified criteria and may receive exemplar tests from the awarding body (United Kingdom, Joint Council for Qualifications 2015).

Portfolios offer an alternative to more formal methods of assessment, including tests or examinations. They also provide a type of performance assessment that comprises a systematic collection of significant samples of student work compiled over a period of time (a month, the school year, or throughout an instructional unit). Typically, portfolios require students to demonstrate a range of cross-curricular skills, including the ability to interpret and reflect on their work (Airasian and Abrams 2003). Clear criteria for assessment of performance are required.

Process or Product

A product such as an essay, a technical drawing, a piece of artwork, a selection of a student's written work, or responses to an arithmetic test may be the focus of the assessment. Or the assessment may focus on processes that cannot be assessed in a product such as a written terminal examination. In this case, the interest may be on skills that apply across subjects (cross-curricular skills), such as communication

skills, data handling, problem solving, use of information and communications technologies, and collaborative working.

Weight Allocated to School-Based Component

The weight allocated to the SBA component in a public examination varies from one education system to another. If schools and teachers enjoy a high degree of trust, the contribution of the school-based component will be correspondingly high. Even in systems in which the school-based element weight is high, the same weight may not be allocated to the school-based element in all subjects in an examination. The school-based element of practical subjects (for example, art or woodwork) tends to attract a higher weight than the school-based element of subjects that do not have a large practical component. In South Africa, for instance, school-based marks in the Independent Examinations Board examinations contribute 50 percent to the marks for subjects with a practical component and 25 percent to the marks for subjects without a practical component (South Africa, Umalusi, Council for Quality Assurance in General and Further Education and Training 2015).

ARGUMENTS IN FAVOR OF SCHOOL-BASED ASSESSMENT

Several arguments support the use of SBA in the context of public (mainly external) examinations (see box 13.3) (Kane 2006; Moss 2003; Shepard 2001; Wolf et al. 1991). First, whereas an external written examination is limited in the range of achievements that can be assessed, which lessens the validity of the exercise, SBA can evaluate student competence in a variety of curriculum topics and skills. These include a wide array of such practical skills as oral fluency, reasoning, capacity to plan and execute projects, and ability to work with other students, all of which are very difficult, some would say impossible, to assess adequately in an external examination. Even oral tests of language administered by an external examiner are recognized as extremely artificial and unlikely to assess accurately the level of an examinee's competence. Second, because SBA is carried

BOX 13.3

Advantages of School-Based Assessment Compared to External Examinations

	Advantages of school-based assessment	Limitations of external exams
Scope	Extends the range and diversity of assessment collection opportunities	Narrower range of assessment opportunities
Authenticity	Assessment by students' own teachers, who know students' competencies	Stressful conditions may lead to students not demonstrating real capabilities
Validity	Improves validity through assessing relevant factors that cannot be included in an external examination	Restricts validity by limiting the scope of the assessment
Positive washback ^a	Ongoing assessment encourages students to work consistently	Examination is summative; effects on teaching and learning may be negative
Teacher and student empowerment	Teachers and students become part of the assessment process Collaboration and sharing of expertise take place within and across schools	Teachers' role in assessment limited
Professional development	Builds teacher assessment skills	Opportunities to develop assessment skills limited

Source: World Bank 2001.
a. Refers to the effects that a high-stakes examination may have on individuals, policies, or practices—within a classroom, a school, an educational system, or society as a whole (Wall 1997).

out in realistic settings over an extended time period by a person who knows students well, it is likely to provide a more valid and reliable appraisal of a student's achievement than can be obtained in a single terminal examination. Third, teachers have access to a rich array of data about students, including their participation in

classroom activities, discussions, and projects. Given this situation, teachers would seem well placed to capture the multidimensional aspects of students' performance; they can take into account the processes through which students develop understanding of subjects and their ability to reflect on the quality of their work. Fourth, SBA can have a positive effect on teaching and learning while at the same time reducing the undesirable negative effects of external examinations. In particular, SBA can be used to promote deeper understanding and higher-order thinking skills. Fifth, if spread over time, SBA can increase the level of student motivation and application throughout the year. Box 13.3 summarizes the advantages of SBAs over external public examinations.

Since assessment by teachers is a crucial component of good teaching and learning, it is argued that every effort should be made to improve teachers' competence in this area (see Carless 2011; Crooks 1988; Daugherty 2011). If SBA becomes part of the certification process, it is likely, as noted earlier in the Queensland situation, that greater effort will have to be invested in improving teachers' general competence in assessment, which, it is argued, should have beneficial effects on teaching and learning.

ARGUMENTS AGAINST SCHOOL-BASED ASSESSMENT

A number of arguments have been advanced against the use of SBA in examinations designed to certify the achievement of students or select them for further education (Crisp and Green 2013; Harlen 1994; Johnson 2011; Keeves 1994; Kellaghan and Greaney 1992; Meadows and Billington 2005; Ofqual 2013; Opposs 2016). Different markers produce different outcomes for the same piece of work, even when relatively tight marking schemes are applied (see chapter 5). Variation in marking standards, both within and between schools, can arise for a variety of reasons. First, there are differences between schools in the characteristics of students who attend them. Second, there are differences between teachers, within and between schools, in the mean level of scores, the spread of scores, and the shape of the

score distributions they assign to students. Teachers may be too “harsh” or too “lenient” and may vary in the extent to which they discriminate between students. Some science teachers may be tempted to refocus instruction when they receive timely confidential instructions from an examination authority to assemble equipment for experiments to be carried out for the practical components of examinations.

Third, it is argued that teacher assessments include many sources of measurement error, arising from variation in the types of assessment themes, topics, and tasks they use; differences in the interpretation and application of performance criteria or marking schemes; and the intrusion of irrelevant contextual information (such as students’ gender or socioeconomic background) in making judgments. Fourth, there may be differences among the courses of instruction that students have studied. Fifth, schools may differ in the characteristics of students that study different subjects within the school.

Sixth, teachers may find it difficult to check the extent of plagiarism, use of online resources, and parental support in completing assignments. Finally, it may be difficult, if not impossible, for a variety of logistical and other reasons, to apply SBA to non-school-based (external) examination candidates.

Other arguments against the use of SBA in the context of public examinations relate primarily to teachers and administration. First, such use of SBA can subject teachers to considerable pressure from parents and students, especially during the periods leading up to and immediately after critical examinations. Bishop (1998) cites a study that found that 30 percent of American teachers felt pressure to give higher grades than students’ work deserved. There are added incentives to award high marks in the school-based component if individual school results are made public. Second, there is scope for students to receive significantly differing levels of assistance from teachers, peers, parents, and others, or to engage in plagiarism. Plagiarism has been identified as a recurring concern in New South Wales (Bennett 2009), and various forms of cheating have been associated with SBA in South Africa (Kerr-Phillips 2007). Third, SBA can change the nature of the relationship between teachers and students, making the

judicial aspect of a teacher's role more prominent. Fourth, there is often a lack of consistency in the assessment demands of schools. Fifth, it is difficult to assess an individual in a task that is carried out collectively by a group of students. Sixth, SBA increases the workload of teachers, requiring them to devote more time to assessment and recording, to storing students' work, and to attending meetings related to school review and moderation procedures. Seventh, SBA gives rise to a variety of administrative problems for schools, such as what to do with non-school-based candidates, or when students are absent for tests, or when students transfer from one school to another, or where to store student work. Eighth, teachers require professional development in how to carry out and implement SBA, in how to grade students' practical work, and in how to record and grade oral performance. Ninth, teachers may use proposals regarding SBA as a bargaining tool in issues related to their salaries and conditions of work. Finally, SBA can require extensive external moderation and in some instances may open the door to bribing of, or by, external moderators.

Costs, which will vary from country to country depending on economic circumstances, are likely to be substantial, especially where external moderation is required. Costs can vary depending on the subject being assessed, the materials that are required, storage of student cumulated work, travel and subsistence expenses, and time required to complete the assessment work. The Caribbean Examination Council incurred some SBA-related savings by assembling its moderators from different territories at a single location for remarking SBA samples (Grima 2003). Opportunity costs, the trade-off cost due to a teacher's absence while carrying out SBAs, could also be a factor as the teacher who has travelled to other schools to serve as a moderator is unable to teach his or her regular students.

Given the level of resources required (especially in science and vocational subjects) and including the need for high-level development of teachers' assessment skills, SBA may not be an option for many examination agencies, including those in emerging market economies.¹ Even in countries with relatively high gross domestic product indexes, SBA standards may not enjoy the same level of

community support as other aspects of the examination system, which are often considered fair, transparent, and objective. England's Office of Qualifications and Examinations Regulation (Ofqual) conducted a review of SBA efforts to focus on constructs that were different from those assessed in written examinations, in order to enhance the validity of the assessment. It concluded that non-examination assessments "should be used only when it is the only valid way to assess essential elements of the subject (Green 2014, 4; Ofqual 2013).

ADDRESSING PROBLEMS IN SCHOOL-BASED ASSESSMENT

A number of procedures have been proposed to anticipate and deal with problems that might arise if SBAs are incorporated into an examination on which performance will have important consequences for a student's educational future. The procedures may be categorized as (a) specification of assessment tasks, (b) detailed specification of assessment criteria, (c) provision of materials to guide teachers, (d) teacher professional development, (e) informing students and parents, (f) moderation by inspection, and (g) statistical adjustment of marks by means of a reference test (Cumming and Maxwell 2004; Gipps and Stobart 2003; Harlen 1994; Keeves 1994). It should be recognized that in the absence of a significant amount of empirical research on the reliability of teacher assessments, proposals to address problems associated with such assessments (including reliability) rely mainly on the application of generally accepted principles of assessment (Johnson 2011).

Specification of Assessment Tasks

Assessment tasks should be clearly described by the examining authority and made available to schools. For example, in the case of a science subject, details should be provided for the project or experiments to be carried out, when they are to be carried out, and how students should report on their work.

Detailed Specification of Assessment Criteria

The skills to be assessed should be specified, as well as the criteria for judging performance. Scoring scales and specific descriptors of what students at different levels of achievement should be able to do may be specified. The weight to be given to various aspects of performance and methods of recording student performance should also be specified.

Provision of Materials to Guide Teachers

Materials to guide teachers in their assessments should be provided: a list of prescribed or recommended tasks, samples of student work exemplifying typical levels of performance, and a form to report student performance. Training will be required in the application of criteria.

Teacher Professional Development

Education or examination authorities will have to support programs of teacher professional development before an assessment is designed in order to standardize the process of assessment and help achieve consensus in marking standards among teachers. These efforts, which are made in the interest of *quality assurance*, include explanations of assessment objectives, procedures, and grading criteria supported by examples; teachers meeting with other teachers in their own schools and possibly with teachers in other schools to discuss students' work; and district seminars to reflect assessment concerns. Moderators can visit schools to explain moderation procedures and help identify variation in assessment practices (Cumming and Maxwell 2004; Western Australia School Curriculum and Standards Authority 2014, 2016).

The adoption of a school-based element as an integral part of a public examination can have implications beyond the field of assessment, affecting the teacher-student relationship. For example, the role of teachers as assessors may conflict with their role as mentors as they strive to ensure that students do not receive inappropriate amounts of assistance from third parties in any work they submit for SBA. Teachers should not provide detailed advice on how to improve drafts, give detailed feedback on errors or omissions, or intervene to improve the content or presentation of work.

Informing Students and Their Parents

Students and their parents should be provided with details of proposed SBA activities, including its rationale, the nature of activities, and any role that may be proposed for parents. It is usual to require students to sign a declaration that the work they are submitting is their own and to acknowledge sources used and assistance provided. The extent to which such a declaration can be trusted is open to question. Every year in New South Wales, Australia, cases are identified in which teachers cannot agree that the submitted material is the original work of a student (Bennett 2009). If in doubt about a student's contribution, it is incumbent on teachers and examination authorities to try to ascertain if a student actually carried out a project on his or her own, the extent to which help was obtained from others, if material was downloaded from a website, or if results were falsified (such as in a science project).

Moderation by Inspection

Before submitting the marks awarded candidates on the school-based element of an examination, schools should carry out an internal standardization exercise to establish the appropriateness of the marks awarded by different teachers. Consistency across schools is also required and moderation will be employed to achieve this. If moderation by inspection is in operation, details of the procedure should be provided to schools. They should be told the aspects of students' work that will be examined, the procedures to be used in adjusting scores awarded by schools, and how decisions by moderators can be challenged.

Moderation may be by visit, in which case external moderators visit each school to inspect the work of students. Another form of moderation involves the school sending relevant work for all candidates to the examining authority. In South Africa, grade 12 SBAs were moderated by the Department of Basic Education, the accredited assessment body, and a council (Umalusi) that set and monitored standards for general education (South Africa,

Department of Basic Education 2012). These forms of moderation can be expensive and time consuming. A less expensive alternative involves moderation of samples of students' work: the school sends the examination authority samples (10–20 percent) of student work and their associated marks from an assessment group chosen to represent the range of achievement levels.

Moderators are likely to be expert practitioners external to the school. They will look for evidence to support the results provided by teachers and relate evidence of student learning to the performance standards specified for the subject. They will not give advice about, or comment on, the work of individual students. Having marked the students' work, moderators will compare the marks they awarded to the marks awarded in the school. There are three possible outcomes of this exercise:

1. The school-based mark is accepted if the difference between it and the moderator's mark is within an acceptable level (for example, within 6 percent of the total mark for a unit).
2. Marks are adjusted if differences between the two exceed the allowed tolerance, but there is general agreement on the rank order of candidates.
3. Marks awarded by the moderator replace the school-based marks if the moderator's ranking disagrees significantly with the school's and marking is too inconsistent to allow rational adjustment.

In the United Kingdom, a large majority of awarded marks have been accepted unchanged (Johnson 2011).

Statistical Adjustment

The moderation of teachers' assessment of their students' work may use statistical procedures to adjust the outcomes following completion of the assessments. This may be done by having all students take a "reference test" (for example, a verbal reasoning test or other general test of skills or abilities). This procedure strives to address the issue of systematic variation between teachers in the criteria that they use for the award of grades or marks, and puts all students on a common scale.

It may be used to generate scores for individuals or to determine group parameters for the statistical moderation of school assessments (McCurry 1995). However, it is based on the questionable assumption that the characteristic measured by the reference test is an appropriate one on which to order student performance in other areas of achievement.

An alternative statistical procedure is to adjust teachers' assessment marks in light of students' performance on the external examination, which becomes the "reference test." This involves adjusting the average and spread of SBA marks to reflect the performance of students in the school on the external examination. Student marks will probably change, but their rank order, though not necessarily relative differences between students, will be maintained (see Bennett 2009). This kind of moderation can involve simple² or complex procedures (see Stanley et al. 2009). An obvious disadvantage of this approach is the assumption that performance on the external (probably written) examination is an appropriate criterion for scaling teachers' assessments of their students. A statistical adjustment would seem more appropriate when SBA and the external examinations share a substantial portion of assessment objectives. By contrast, expert judgment moderation would be more appropriate when the outcomes of the SBA and the external examination are very different. It would also be used when the number of candidates in a school is small (15 or fewer).

It should be noted that expert judgment and statistical moderation or adjustment are not mutually exclusive. When statistical moderation is the main method, it will most likely be supplemented by expert judgment of a sample of candidates' work. When expert judgment of a sample of candidates' work is the main method of moderation, it is not unusual to support it with an examination of statistical data on candidate performance on the external component (see, for example, HKEAA 2011).

The use of expert moderators is generally preferred to statistical moderation, which may be considered an interim measure until there is agreement that marking standards are being applied consistently.

CONCLUSION

Many positive features have been associated with the idea of teachers assessing their own students, not least the potential to improve the validity of public examinations. However, many examination agencies are faced with the challenge of gaining public support for SBA, in establishing that students' work can be fairly assessed and marked and that internal and external assessors are not subjected to pressure. Although SBA is well developed in some countries, efforts to have it contribute to grades in countries in which the external element in public examinations has been very strong have run into difficulty. A variety of reasons have been proposed to explain this. It is expected that marking standards will vary between teachers and schools, that malpractice will occur (individual students obtaining external assistance), that pressure will be placed on teachers and students, that judgmental moderation procedures will be expensive, that advantages may be conferred on students from high-status backgrounds, that not all teachers will have the assessment skills required for the task, and that a variety of administrative problems may be anticipated (for example, dealing with student absences and transfers).

Not least among the problems to be anticipated in the incorporation of a school-based element in a public examination is the need for a Ministry of Education or the examination authority to provide resources and adequate infrastructure. Box 13.4 summarizes key steps that should be considered in establishing this infrastructure.

Teachers, for their part, will be required to acquaint themselves with the requirements of SBA and to describe them to students. They will also be required to administer SBA tasks as an integral part of their teaching, to assess students' work or performance according to prescribed procedures, to authenticate students' work, to participate in school and moderation panels, and to submit students' marks and samples of their work to the examining authority (see HKEAA 2013).

BOX 13.4**Steps for Carrying Out School-Based Assessment as Part of a Public Examination System**

1. Ensure that curricula or syllabi for all subjects set out the key skills, understanding, and knowledge that students are expected to have acquired by the end of their course.
2. Determine the key decisions that have to be made and inform schools about the dimensions of school-based assessment tasks: the skills to be assessed, methods of assessment, criteria for grading (how achievement is to be judged and recorded; weighting for various skills), and exemplars of student work.
3. Decide which tasks are prescribed and which left to the discretion of schools.
4. Agree on the ratio of school-based to external assessment in determining a grade.
5. Establish procedures to ensure that all teachers are working to an agreed standard.
6. Design assessment and recording materials and provide them to schools.
7. Provide teachers with relevant professional development (assessment tasks they can use; consensus moderation procedures).
8. Identify expert practitioners who are prepared to act as moderators.
9. Develop a schedule to help ensure that moderation procedures are carried out in time in accordance with the examination timetable.

Sources: Adapted from Queensland Curriculum and Assessment Authority 2014 and Reyneke, Meyer, and Nel 2009.

NOTES

1. In Uganda in 2012, although science subjects were compulsory at O level, only 50 percent of secondary schools had functional laboratories, and 79 percent of schools did not conduct practical lessons in science (Tukacungurwa, Ssebbale, and Basiime 2012). In 2014 in Kenya, basic science equipment was found to be lacking, especially in rural schools, and some candidates saw science equipment for the first time hours before the examination (Mabatuk 2014).

2. The following linear equation can be employed to adjust school-based marks:

$$T_{ij} = X_i + r \frac{S_{xi}}{S_{yi}} (Y_{ij} - Y_i),$$

in which T_{ij} is the adjusted score of student j , Y_{ij} is the SBA score of student j , Y_i is the group's mean SBA score, X_i is the group's mean score on the external exam, S_{yi} is the standard deviation of SBA scores, S_{xi} is the standard deviation of external exam scores, and r is the correlation between SBA scores and external exam scores.

REFERENCES

- Airasian, P. W., and L. M. Abrams. 2003. "Classroom Student Evaluation." In *International Handbook of Educational Evaluation*, edited by T. Kellaghan and D. L. Stufflebeam, 533–48. Dordrecht, Netherlands: Kluwer Academic.
- Bennett, J. 2009. "Country Case Study: Australia (New South Wales)." In *Secondary School External Examination Systems*, edited by B. Vlaardingerbroek and N. Taylor, 55–78. Amherst, NY: Cambria.
- Bishop, J. H. 1998. "Do Curriculum-Based Exit Exam Systems Enhance Student Achievement?" Research Report Series RR-40, Consortium for Policy Research in Education, Graduate School of Education, University of Pennsylvania, Philadelphia.
- Black, P., C. Harrison, C. Lee, B. Marshall, and D. Wiliam. 2003. *Assessment for Learning: Putting It into Practice*. Maidenhead, Berkshire, UK: Open University Press.
- Campbell, M. 2018. "Standard Setting in Queensland: The Queensland Certificate of Education." In *Examination Standards: How Measures and Meanings Differ around the World*, edited by J.-A. Baird, T. Isaacs, D. Opposs, and L. Gray, 182–205. London: University College London Institute of Education Press.
- Carless, D. 2011. *From Testing to Productive Student Learning: Implementing Formative Assessment in Confucian-Heritage Settings*. London: Taylor and Francis.
- Clarke, E. 1987. "Assessment in Queensland Secondary Schools: Two Decades of Change 1964–1983." Queensland Department of Education, Brisbane.
- Crisp, V., and S. Green. 2013. "Teacher Views on the Effects of the Change from Coursework to Controlled Assessment in GCSEs." *Educational*

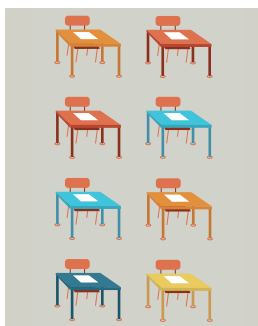
- Research and Evaluation: An International Journal on Theory and Practice* 19: 680–99.
- Crooks, T. J. 1988. “The Impact of Classroom Evaluation Procedures on Students.” *Review of Educational Research* 58: 438–81.
- Cumming, J. J., and G. S. Maxwell. 2004. “Assessment in Australian Schools: Current Practice and Trends.” *Assessment in Education* 11: 89–108.
- Darling-Hammond, L., and L. McCloskey. 2008. “Assessment for Learning around the World: What Would It Mean to Be Internationally Competitive?” *Phi Delta Kappan* 90: 263–72.
- Daugherty, R. 2011. “Assessment Paradigms: Designing and Implementing a Teacher-Based Assessment System—Where Is the Infrastructure?” Paper presented at the Seminar “Teachers’ Judgments within Systems of Summative Assessment: Strategies for Enhancing Consistency,” St. Anne’s College, Oxford, UK, June 20–22.
- Gipps, C., and G. Stobart. 2003. “Alternative Assessment.” In *International Handbook of Educational Evaluation—Part Two: Practice*, edited by T. Kellaghan and D. L. Stufflebeam, 549–75. Dordrecht, Netherlands: Kluwer Academic.
- Greaney, V., and T. Kellaghan. 2008. *Assessing National Achievement Levels in Education*. Washington, DC: World Bank.
- Green, S. 2014. “Models of Internal, School-Based Assessment: Challenges and Possibilities.” Paper presented at the 40th Annual Conference of the International Association for Educational Assessment, Singapore, August 25–30. http://www.iaea.info/documents/paper_371f236e1.pdf.
- Griffith, S. A. 2015. *School-Based Assessment in a Caribbean Public Examination*. Kingston, Jamaica: University of the West Indies Press.
- Grima, G. 2003. “School-Based Assessment: Implementation Issues and Practices.” Paper presented at the 21st Annual Association for Educational Assessment in Africa (AEAA) Conference, Cape Town, South Africa, August 25–30.
- Harlen, W. ed. 1994. *Enhancing Quality in Assessment*. London: Chapman.
- HKEAA (Hong Kong Examinations and Assessment Authority). 2011. “Hong Kong Diploma for Secondary School Examination: Moderation of School-Based Assessment Scores in the HKDSE.” HKEAA, Hong Kong SAR, China.
- . 2013. “Hong Kong Diploma for Secondary Education Examinations: Information on School-Based Assessment.” HKEAA, Hong Kong SAR, China.
- Johnson, S. 2011. “A Focus on Teacher Assessment Reliability in GCSE and GCE.” Office of Qualifications and Examinations Regulation (Ofqual) 11/4807, Ofqual, Coventry, UK.

- Kane, M. T. 2006. "Validation." In *Educational Measurement*, 4th ed., edited by R. L. Brennan, 17–64. Westport, CT: American Council on Education and Praeger.
- Keeves, J. P. 1994. *National Examinations: Design, Procedures and Reporting*. Paris: International Institute for Educational Planning.
- Kellaghan, T., and V. Greaney. 1992. *Using Examinations to Improve Education: A Study in Fourteen African Countries*. Washington, DC: World Bank.
- Kerr-Phillips, F. 2007. "School Based Assessment: The Need, the Reality, the Future—A Perspective from the Independent Examinations Board of South Africa." Paper presented at the 33rd Annual Conference of the International Association for Educational Assessment, Baku, Azerbaijan. <https://iaea.info/documents/school-based-assessment-the-need-the-reality-the-future/>.
- Mabatuk, V. 2014. "KCSE Candidates See Science Practical Apparatus for the First Time Hours to Exams." *Standard Digital*, October 19. <https://www.standardmedia.co.ke/article/2000138779/kcse-candidates-see-science-practical-apparatus-hours-to-exams>.
- Masters, G., and B. McBryde. 1994. "An Investigation of the Comparability of Teachers' Assessments of Student Folios." Tertiary Entrance Procedures Authority, Brisbane.
- Maxwell, G. S. 2002. "Moderation of Teacher Judgments in Student Assessment: Discussion Paper on Assessment and Reporting." Queensland School Curriculum Council, Brisbane.
- McCurry, D. 1995. "Common Curriculum Elements within a Scaling Test for Tertiary Entrance in Australia." In *Admission to Higher Education: Issues and Practice*, edited by T. Kellaghan, 253–60. Princeton, NJ: International Association for Educational Assessment and Dublin: Educational Research Centre.
- Meadows, M., and L. Billington. 2005. "A Review of the Literature on Marking Reliability." Manchester, UK: Assessment and Qualifications Alliance.
- Moss, P. A. 2003. "Reconceptualizing Validity for Classroom Assessment." *Educational Measurement: Issues and Practice* 22: 13–25.
- Ofqual (Office of Qualifications and Examinations Regulation). 2013. "Review of Controlled Assessment in GCSEs." Ofqual 13/5291, Ofqual, Coventry, UK. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/377903/2013-06-11-review-of-controlled-assessment-in-GCSEs.pdf.
- Opposs, D. 2016. "Whatever Happened to School-Based Assessment in England's GCSEs and A Levels?" *Perspectives in Education* 34: 52–61.

- Queensland Curriculum and Assessment Authority. 2014. "School-Based Assessment: The Queensland System." Queensland Curriculum and Assessment Authority, Brisbane.
- . 2017. "Why Senior Schooling in Queensland Is Changing." Queensland Curriculum and Assessment Authority, Brisbane. https://www.qcaa.qld.edu.au/downloads/senior/snr_new_assess_te_senior_schooling_changing_tscript.pdf.
- . 2018. "Externally Moderated School-Based Assessment: For Students Completing Year 12 in 2018 and 2019." Queensland Curriculum and Assessment Authority, Brisbane. https://www.qcaa.qld.edu.au/downloads/publications/school-based_assess_auth_subj.pdf.
- Reyneke, M., L. Meyer, and C. Nel. 2009. "School-Based Assessment: The Leash Needed to Keep the Poetic 'Unruly Pack of Hounds' Effectively in the Hunt for Learning Outcomes." *South African Journal of Education* 30: 277–92.
- Satterly, D. 1981. *Assessment in Schools*. Oxford, UK: Blackwell.
- Shepard, L. 2001. "The Role of Classroom Assessment in Teaching and Learning." In *Handbook of Research on Teaching*, 4th ed., edited by V. Richardson, 1066–1101. Washington, DC: American Educational Research Association.
- South Africa, Department of Basic Education. 2012. "National Protocol for Assessment Grades R–12." South Africa, Department of Basic Education, Pretoria. <https://www.education.gov.za/Portals/0/Documents/Policies/NatProtAssess.pdf>.
- South Africa, Umalusi, Council for Quality Assurance in General and Further Education and Training. 2015. "Quality Assurance of the 2015 National Senior Certificate (NSC) Examinations and Assessment of the Independent Examinations Board (IEB)." South Africa, Umalusi, Pretoria.
- Stanley, G., R. MacCann, J. Gardner, L. Reynolds, and I. Wild. 2009. "Review of Teacher Assessment: Evidence of What Works Best and Issues for Development." Oxford University Centre for Educational Assessment, Oxford, UK.
- Tukacungurwa, C. M., S. Ssebbale, and N. K. Basiime. 2012. *The Quality of Science Education in Uganda*. Kampala: Uganda National Council for Science and Technology.
- United Kingdom, Joint Council for Qualifications. 2015. "Instructions for Conducting Controlled Assessments (Legacy GCSE Qualifications): 1 September 2015 to 31 August 2016." Joint Council for Qualifications, London. <https://www.fazakerleyhigh.org/docs/Exam/ControlledAssessments.pdf>.

- Wall, D. 1997. "Impact and Washback in Language Testing." In *Encyclopedia of Language and Education—Vol. 7: Language Testing and Assessment*, edited by C. Clapham and D. Corson, 291–302. Dordrecht, Netherlands: Kluwer Academic.
- Western Australia School Curriculum and Standards Authority. 2014. *Western Australian Certificate of Education (WACE) Manual: General Information for Senior Secondary Schooling 2015–2016*. Rev. ed. Cannington: Western Australia School Curriculum and Standards Authority.
- . 2016. "About Moderation." Western Australia School Curriculum and Standards Authority, Cannington, Australia. <http://wace1516.scsa.wa.edu.au/moderation/about-moderation>.
- Wolf, D., J. Bixby, J. Glen, and H. Gardner. 1991. "To Use Their Minds Well: Investigating New Forms of Student Assessment." *Review of Research in Education* 17: 31–74.
- World Bank. 2001. "Toolkit on Public Examination Systems in Developing Countries." World Bank, Washington, DC.

CHAPTER 14



CONSIDERATIONS FOR EXAMINATION REFORM

INTRODUCTION

Public curriculum-based examinations are a key component of many education systems throughout the world. They not only provide the basis for important decisions about individual students, they can affect the educational experiences of all students who come into their orbit, determining to a significant extent what they learn and how they learn. Student performance on a public examination tends to be recognized and rewarded more than any other form of learning achievement. The content and form of examinations dominate instructional and classroom assessments in schools, especially in the grades that are close to the terminal exams, such as at the end of lower-secondary and upper-secondary school.

This book's review of examination practices in many educational systems, including in low- and middle-income countries, underscores the fact that public examination approaches vary considerably throughout the world (Baird et al. 2018), that no assessment system is without its defects, and that the reform of a system will inevitably involve a series of compromises and trade-offs that will affect the extent, nature, and timing of changes.

This final chapter describes issues to be considered in bringing about reform. The next section focuses on the range of options that examination authorities might consider in determining the nature of the assessment task to be administered in the final examination. The subsequent section lists some options for improving administrative aspects of the examination. The final section describes some characteristics of a good examination system that policy makers might consider when reviewing reform options.

EXAMINATION REFORM CONSIDERATIONS

The reasons for changes in examination systems differ substantially from one situation to another. It would seem that what is common to change is that it is initiated when one or more aspects of an examination system are perceived by influential stakeholders to be dysfunctional or unsatisfactory in some respect.

Inevitably, efforts to bring about change will depend on the state of development of the public examination system. In some educational systems, progress would be evident if the public examination were administered on the specified dates, or if examination papers were not leaked prior to the date of administration, or if the number of court appeals related to examination results was reduced. More developed examination systems might equate progress with an increase in the percentage of higher-order cognitive skills assessed, or with the successful implementation of a program of oral assessment as part of language testing, or with the introduction of a form of school-based assessment that contributed to the overall validity of the examination.

Given the level of diversity among examination systems, it is not surprising to find that some have strengths in certain areas (such as examination security) but may be relatively weak in others (such as attention to validity). Irrespective of the stage of development of individual examination systems, it is hoped that some of the issues addressed in previous chapters will help identify areas for improvement, in particular, better aligning of examinations and curriculum,

improving scoring, reducing malpractice, enhancing equity, and using results to improve teaching. Some of these may not be applicable to individual systems, others may have been addressed, while some may be too ambitious to be considered given the pressures and challenges confronting overworked and underresourced examination systems.

In general, it has proven difficult to bring about major changes to examination systems (Isaacs and Gorgen 2018). In some jurisdictions, change has been incremental, involving, for example, altering the format of questions or the categories for reporting results. In a few others, change has been pronounced, creating upheavals in education systems by replacing school-based certification with external examinations (in the countries of the former Soviet Union) or by abolishing external examinations and introducing school-based certification (as in Queensland, Australia, and in Sweden). In China, the *gaokao* reform pilot program fell behind schedule because of the many demands it placed on students, teachers, schools, and examination authorities (China Policy 2019).

Reforms can have associated downsides or opportunity costs that can provoke opposition from some quarters. For example, opposition to reforms such as abolishing a public examination (for example, a primary school leaving or mid-secondary-level examination) may come from examination officials, teachers who derive income from supervising or scoring papers, individuals and businesses that offer grinds or supplementary tuition, the media, advocates of using examination results as measures of teacher and school accountability, and schools that rely on examinations for selection purposes.

Before embarking on an examination reform program, policy makers might well reflect on background factors that might influence the outcome of reform initiatives (see box 14.1). The significance of each factor in a reform program will depend on current examination practices and on the precise nature of the proposed reform program.

Despite the fact that education systems tend to be slow to embrace changes in their examination systems, efforts to improve are ongoing and further changes can be expected. As a consequence, some of the details in this volume about examination systems in some countries may be out of date before it reaches the reader.

BOX 14.1**Contextual Factors to Be Considered When Planning Examination Reform**

- The willingness of the political and educational authorities to support change
- The likely impact on public confidence in the examination system
- The availability of financial resources, and the policy on examination fees
- The likely impact on students from low-income families
- Possible opposition from those likely to suffer financial loss or loss of status
- Possible opposition from students and parents
- Possible opposition from influential stakeholders, such as religious and ethnic representatives, teacher unions, and providers of supplementary tuition
- The available level of technical competence in the examination agency
- The implementation capacity of the examination agency
- The effectiveness of the supportive infrastructure, such as computers and the internet
- The time and resources needed for students and teachers to become familiar with the proposed reforms, including pilot testing of new examination formats
- The time needed to modify curricula, textbooks, and other instructional materials
- The likely impact of examination results on users, such as employers and tertiary-level selection procedures

OPTIONS FOR IMPROVING EXAMINATION ITEMS AND QUESTIONS

In reviewing current practice or in contemplating the introduction of new forms of assessment, policy makers and examination agencies have a wide range of options, including placing more emphasis on assessing higher-order cognitive skills. Their task is to assess carefully the advantages and disadvantages associated with each before deciding on the form of assessment that best fits their educational traditions and expectations. Initial discussions may reveal areas of conflict

such as disagreements between advocates for examinations that promote broad learning with less emphasis on memorization and those who are concerned that such an approach might undermine the integrity of the examination and promote forms of malpractice, as was the case in the China pilot program (China Policy 2019). In some instances, problems such as the negative impact of grind schools or shadow education may be outside the control of examination agencies; many of these can be attributed in great part to the inadequacies of mainstream education systems and the growth of unregulated private sectors (Bhorkar and Bray 2018).

Multiple-Choice Tests in Place of Essay-Type Examinations

In recent decades many examination bodies have increased the emphasis on multiple-choice testing (see chapter 4) at the expense of essay-type examinations. Advantages of multiple-choice testing include more extensive curriculum content coverage, enhanced reliability, reduction in time required to score tests, lower overall costs, and ability to provide feedback on common student errors. Multiple-choice tests, however, have a number of limitations, including possible reduced validity due to the tendency to focus on factual information; limited development of practical, oral, and writing skills; difficulty assessing depth of knowledge, skill in organizing and synthesizing information, and solving complex problems; and encouraging guessing behavior. A strong case can be made for using both multiple-choice and essay questions in the same examination subject in order to capitalize on their respective strengths.

Reflecting “Real-Life” Situations

Examination agencies may wish to consider introducing assessments that reflect “real-life” situations. Authentic assessment requires candidates to show mastery of important knowledge and skills by applying them in authentic settings to the types of problems or situations faced by people operating in the “real world.” Students might be expected to access books and other printed material as well as websites and develop skills other than those normally associated with more traditional examination formats, such as the ability to locate, organize, and

analyze information. Examination agencies should appreciate that authentic or real-life types of assessments may not adequately cover various curriculum areas, that developing items and scoring rubrics is challenging, and that the dependency on nonschool resources may bias the process in favor of some candidates from more affluent backgrounds. This form of assessment requires relatively high-level teaching skills and may prove particularly challenging for examination systems in countries that lack basic educational resources.

Including Achievements in the Examination That Cannot Be Assessed in Written Responses

Examination authorities can augment their written or multiple-choice assessments with external or internal school-based assessment of students' oral or aural language skills and with practical assessments in science or vocational skills. These augmentations can lead to improved validity due to enhanced curriculum coverage, more authentic forms of assessment, higher-quality student learning experiences, an increase in student motivation, and an increased probability that skills will be taught. However, it can increase examination administration time as well as costs associated with travel and related fees of external assessors or moderators (if used) and yield relatively low reliability indexes. Resource constraints might make this form of assessment difficult to implement in some situations.

Designing Assessment Practices That Contribute to Meeting Challenges Facing Education

Both curriculum developers and examination agencies are likely to come under pressure from employers and government bodies to ensure that examination content reflects the need to prepare students for the world of work. While there is considerable debate surrounding the nature of the knowledge, skills, and habits required in the workplace of the future, they are generally meant to include such "twenty-first century skills" as problem solving, critical thinking, synthesis, oral and written communication, innovation, teamwork, and use of information and communications technology. Options for assessing these skills might encompass different forms of written examinations, more

one-to-one assessments, individual and team projects, and continuous assessments, many of which will have cost implications.

OPTIONS FOR IMPROVING ADMINISTRATION AND REPORTING

Abolishing Examinations That Are No Longer Needed

Reform-minded Ministers of Education might learn from the experiences of the many countries that abolished the primary school leaving certificate examination in situations where it was no longer needed, as all or virtually all students transferred to the next phase of the educational system.¹ Abolishing the primary school leaving examination can help reduce overall examination costs and spare students and teachers some of the negative consequences of these exams, such as grade repetition, narrowing of the effective curriculum, “teaching to the test,” overemphasis on rote learning, and malpractice (see chapter 5). In this regard, it is worth noting that a 2018 review of education in Sub-Saharan Africa recommended the elimination of primary school leaving examinations on the grounds that they impeded access to lower-secondary schools (Bashir et al. 2018). Although some may argue that removing these examinations eliminates an objective way of monitoring achievement trends, it should be appreciated that traditional primary school leaving examinations are not appropriate instruments for measuring achievement trends.²

One Final Exit Examination or Separate Certification and Selection Examinations

Using one public examination to serve both certification and selection functions has the advantage of more efficient use of examination agency resources, lower costs, less disruption of the school timetable and teaching time, and higher motivational value. However, because of the primacy of the selection function in many systems, a single examination designed to serve both functions may be based on a course of study that is too challenging for students not seeking admission to higher levels of education, who are often in the majority, and

may result in low-scoring candidates unnecessarily and unfairly being considered nonachievers and failures. A separate end-of-school certification examination can offer students who achieve the required minimum standard a formal certificate of achievement to be used when seeking employment. It may lack, however, the motivational incentive of the more high-stakes selection examination and may have lower status among employers. Computer-adaptive testing (CAT) would allow accurate assessment of both high- and low-achieving students (Bakker 2014); however, CAT tests tend to be expensive to construct, require pilot testing of items, allow candidates different amounts of time, and require that examination agencies have high-level psychometric skills (Thissen and Mislevy 2000).

Increasing the Number of Options for Taking Examinations

Having an opportunity to repeat an examination during the academic year can be considered equitable when issues such as illness and family bereavement arise. Offering the examination twice in one year can increase the time devoted to examination preparation for students and teachers. It can also restrict the amount of instructional time, not only for examination candidates but also in some instances for the large number of students in lower grades who do not have to attend school during examination periods. Allowing repeats during the school year requires examination agencies to concentrate almost exclusively on administrative tasks, leaving them with little or no time to upgrade overall examination quality or staff skills. Marking standards may differ depending on the time of year the examination was administered.

Modifying Scoring and Reporting Systems

Adjusting or inflating examination marks to ensure candidates reach key cutoff points (pass/fail) is not uncommon; one board, for instance, inflated mathematics scores by as much as 16 marks in 2016 (Bhattacharji and Kingdon 2017). Such a practice has little to recommend it and increases the problem of discriminating among higher-scoring students seeking admission to competitive tertiary-level programs. Second, examination results may be reported in the form of

raw or standardized scores, percentage scores, levels such as honors, pass or fail, or letter grades A, B, C. Each type of score has limitations, including being subject to measurement error.³ Options for modifying scoring are presented in box 14.2.

Reducing examination marks to grade bands, while appearing to simplify the reporting process, increases the likelihood of students with different achievement levels (such as 60 percent and 69 percent) getting identical grades. Agencies may wish to reflect on the current widely accepted practice of adding student raw scores across a range of different subject areas. This procedure has limitations as it does not take into account the difficulty levels of different examination subjects or the variation in marks awarded for individual subjects (see chapter 7).

Labeling Students as Failures

Media reports often highlight examination failure rates; to cite one example, in one instance media reported that 64 percent of 2017 final year students had failed the board examination (India Today 2017). The common practice in many countries of assigning a failing grade

BOX 14.2

Options for Scoring Examinations

- Replacing raw or percentage scores with letter grades
- Preserving percentages attaining different cutoff points such as “pass” or “proficient,” to help ensure that standards appear to be kept constant over time
- Artificially adjusting scores through changing the passing criterion or awarding “grace marks” to achieve political objectives, for example, ensuring a minimum failing rate in a particular subject, such as the national language
- Increasing the weights assigned to different subjects such as mathematics and science to reflect perceived changing national economic priorities
- Varying the weights assigned to each question or item

for performance on a secondary-level exit examination seems to ignore the fact that the student, in all likelihood, has a record of academic achievement, having completed at least primary school and, in most instances, four to five years of secondary level, and may have passed a formal state examination at the end of primary or during secondary level. While recognizing that many examination boards and systems have long-established traditions of reporting failure rates, consideration might be given to using carefully designed examinations and scoring systems to describe low achievement levels without attaching a failing label to a student's record.

Reviewing Cost Options

Overall running costs of public examination systems are likely to escalate with increasing candidate numbers and efforts to broaden the scope of assessment in the interests of enhancing validity. Options for cutting costs, such as reducing the number of questions, increasing the proportion of easily scored questions, or lowering payments for examination supervisors and correctors, should take into consideration their likely impact on the quality of the examination papers, on teaching and learning, and on levels of confidence in the integrity of the results. The expense associated with labor-intensive performance examinations, such as orals, art, music, and aspects of science, needs to be weighed against the advantages and disadvantages of not using this type of examination format, especially bearing in mind the maxim that “what is examined tends to be taught, and what is not examined tends to be ignored.” Many systems, even if favorably disposed to offering performance examinations, may be unable to afford the additional administrative costs. Opportunity costs in the form of negative impacts related to high-stakes examinations should also be considered. These include grade repetition or early dropout prompted by fear of examination failure, and the negative impact on nonexamination students who cannot attend schools while they are being used as examination centers. As economies improve and the demand for more educated workforces increases, competition for scarce places and the pressure on students to perform well on exit examinations are likely to intensify, bringing with them serious problems of

examination-related stress (Pells 2017) and, in some instances, student suicide (see chapter 5 and, for example, Kale 2018; Lindsey 2003; and Straits Times 2013).

Addressing Examination Security

No board or agency can afford to ignore the issue of examination security. Breaches in examination security, some of which were highly sophisticated, have been reported in low-, middle-, and high-income countries. Perpetrators of various forms of malpractice have included examination and government officials, admissions officers, examination administrators, commercial and government printers, school principals and teachers, as well as parents and students. Failure to improve examination-related security could undermine the integrity of the entire examination process. Appropriate legally backed sanctions should be put in place and resulting laws and regulations should be both enforceable and enforced. Some threats to examination security can be addressed by measures outlined earlier (see chapter 11). Security-related solutions should be culturally acceptable. To familiarize themselves with the ever-changing challenges to examination security, many of which involve sophisticated technology, agencies should share information and draw on the experiences of other similar bodies.⁴

Introducing Modern Technologies

Modern technologies such as computers, scanners, iPhones, video cameras, and the internet can have both positive and negative impacts on many facets of the work of public examination systems. Advantages include increased efficiency in examination paper construction, candidate registration, printing and packaging examination papers, scanning and scoring answer sheets, and processing, analyzing, and reporting of results; on-screen marking; and enhanced security. Disadvantages of using information technology (IT) at the school level for registration and administration of public examinations include the expense of purchasing, installing, and storing equipment; dependency on high-level technical support, internet connectivity, appropriate software, and a constant power supply; and reluctance of some teachers and students

to use computers. Online examinations may face the problem of system overload when thousands of candidates log on to dedicated websites at the same time.⁵ Examination boards should give students adequate time to become familiar with the new technologies and examination formats through an extensive program of pilot testing. Increased efficiencies attributable to IT will inevitably lead to a loss of full- or part-time employment within examination agencies and may pose political challenges. Agencies will have to consider how to respond to the ongoing threat posed by computer hackers, including their ability to break into agency files, access examination papers, and alter results. Some examination boards and ministries of education may regard modern technologies as unaffordable at this time.

Using Results

Student performance information should be used to provide feedback to teachers and educational policy makers (UNESCO 2013). Common mistakes made by candidates can provide information of use to teacher educators and providers of teacher professional development courses, and for curriculum developers (see chapter 6). Policy makers can use data from examinations to monitor changes in candidate subject selection, compare results from different regions and subgroups of interest (such as rural or socially disadvantaged students or private schools), and explore the reasons for poor performance and propose remedial strategies. Poor examination results can prompt policy makers to initiate reform programs.⁶ Analysts can advise such users of results as policy makers and admissions officers on the reliability of current marking systems, on the extent of equivalence of grades across subject areas, on the impacts of current subject-weighting systems, and on the likely effects of modifying subject weights on tertiary-level intake.

Publishing School League Tables

Some countries publish league tables that rank schools based on examination results while others have barred their publication. Proponents of publishing league tables argue that they promote school and teacher accountability, are valued by parents, enhance

student learning, and are of interest to the general public. Arguments against the use of examination results as measures of school or teacher effectiveness often highlight the lack of consideration of the particular context (such as location, teaching quality, and home background factors) in which the examinations were conducted (Kellaghan and Greaney 2001; Kellaghan, Greaney, and Murray 2009). Examination agencies considering whether or not to publish results might reflect on the direct and indirect outcomes of supporting a policy of publishing school league tables, including reasons why some educational systems forbid publication (see chapter 6).

CHARACTERISTICS OF A GOOD EXAMINATION SYSTEM

Examination agencies should guard against rigid adherence to the same procedures without considering opportunities to improve. The following list of characteristics provides some criteria an examination agency might use to reflect on or evaluate aspects of its current system. Progress in achieving the performance levels implied by these characteristics may be quite difficult to attain, especially in the short term. It will depend on the stage of development of the system; levels of political, administrative, financial, and technical support; examination agency leadership and implementation capacity; and the willingness of teachers and other stakeholders to embrace change.

Fitness for Purpose

An examination should produce scores that serve its intended purposes or functions (such as certification and selection). It should accurately reflect examinees' achievements, distinguish between levels of achievement, and in the case of selection, help identify students most likely to benefit from further education. Curriculum authorities should be confident that the examination questions are aligned with the curriculum and adequately reflect important curriculum objectives. Examination procedures should be sufficiently transparent to assure users that threats to validity (see chapter 7) are addressed. The examination's scoring and security procedures should give the general public and other stakeholders confidence in its results (see chapter 11).

Assessment Capacity

Examination agencies should have policies and procedures that foster capacity building to help ensure that various staff members have statistical and psychometric experience, are aware of major regional and international developments in technical and operational aspects of public examinations, techniques used by candidates and others to engage in examination malpractice, and relevant curriculum developments. Some staff members will require high-level technical competence in using modern technology; IT systems, including computers, software, scanners, and internet facilities, should be capable of meeting demands for secure and timely handling of large amounts of data. For systems contemplating school-based marks contributing to a final examination score, examination agencies should work closely with ministries of education and institutes for teacher development to help ensure that teachers become proficient in marking their own students.

Equity and Integrity

Ministries of Education should ensure that formal mechanisms exist to monitor the conduct of the examination system and the way in which results are used. Examinations should be deemed fair and impartial. No particular examinee or group of examinees should have an unfair advantage over others; neither should any particular group be unfairly disadvantaged. Indicators of equity and trust include a high level of public confidence in the integrity and competence of examination agency staff, including those who supervise the administration of examination papers and mark or grade students' work. Enhanced equity can be promoted by taking steps such as ensuring that examination fees, if required, do not place an excessive burden on parents or guardians; making appropriate provision for candidates with special needs; and not including culturally or linguistically inappropriate questions, items, or tasks. Individual public examination results should be treated as confidential and not disclosed to others who do not have a clear right to know the results; if results are made available for research purposes, names of individual candidates and

schools should be removed. Examination boards and agencies should constantly review threats to the integrity of examinations, including those posed by modern technology.

Efficiency and Cost-Effectiveness

The examination agency should conduct its work in an effective and efficient manner and deliver its services making the best possible use of physical, financial, and human resources. The examination authority should account for all expended resources and comply with sound fiscal procedures and have its records audited. Examinations should be administered according to agreed-upon schedules and results should be issued on time and in an appropriate form for decision making, for example, for selection for the next highest level of the educational system. Ministries of education should consider eliminating formal transition examinations in cases where virtually all students move to the next level. In a similar vein, the practice in some countries of administering the examination in two parts at the end of the first and the second semesters should be reviewed, especially in situations in which schools are closed during both examination periods.

Transparency

Approaches to transparency will vary according to prevailing cultural norms and administrative practices. These approaches can vary from no communication to providing press releases and holding press conferences. To promote transparency, agencies can use websites to make public information such as the examination schedule, registration and identification requirements including those for private or external candidates, arrangements for candidates with disabilities or diverse educational needs, fees, types of external and internal assessments used, regulations on items allowed at the examination, rules for those wishing to leave the room during the examination, penalties for malpractice, methods of obtaining results, and policies for repeats and appeals. Efforts to make the examination process transparent should not violate legal and propriety obligations and should recognize existing data protection safeguards (Connolly and Cullen 2017).

Impact on Teaching and Learning

The examination authority or agency should ensure that, where possible, the preponderance of questions or items is not limited to recall of factual information and that the breadth of the intended curriculum is not unduly narrowed by the content of the examination. It should provide timely, relevant feedback for teachers, curriculum personnel, and other interested parties and for those responsible for teacher professional development, and for subject matter associations where they exist. It should conduct periodic studies to monitor the extent of examination grade inflation. It should support or carry out periodic evaluations to monitor the consequences and impacts of examinations on teaching, learning, and education quality.

CONCLUSION

This book concurs with Roach's observation that "after many vicissitudes, examinations had won their place, not because they were popular, but because they could not be dispensed with. The question . . . was . . . not whether they should be retained or abandoned, but how their efficiency could be improved" (Roach 1971, 286). Since this comment, there has been considerable progress in our understanding of key aspects of examinations, which have been reflected to some extent in various examination boards and agencies modifying their procedures for developing questions, scoring, and using results. Attention has been focused on the technical areas of validity and marker reliability and on developing performance criteria, standards, and moderation procedures. Other changes include the introduction of test formats such as aural, oral, performance, and portfolio assessment (in addition to written examinations); more secure printing of examination papers; and streamlining of selection procedures based on results.

Serious challenges remain such as accommodating much greater numbers of candidates with greater levels of diversity, competing selection and certification needs covering important but often untested areas of the official curriculum, increasing transparency,

setting standards, scoring and grade inflation, and the ever-present threats to examination security. Efforts to lessen examination-related pressure are likely to be less than successful as long as the number of applicants for certain courses or jobs exceeds the number of available places.

A consideration of the large variation in public examination systems should serve to show that “one’s own national or regional system does not have the inevitability that custom and practice seem to imply” (Black 1996, 19). Because of their relative isolation, rigid traditions, and embedded administrative practices, examination officials may need time to move away from a “We’ve always done it this way” mindset. They should respect but not be bound by the established patterns of power and authority and the procedures and support systems that have evolved over many decades. Examination agencies may expand their professional horizons through formal training and by learning from the experiences of other examination agencies through exchange programs and by adopting pragmatic practices that offer reasonable possibilities of improvement.² They should also consider taking the lead in supporting public awareness programs about upcoming changes to important aspects of examinations, and in developing procedures to counteract uninformed comments from the media and politicians (Spöttl et al. 2016). In some educational systems, teacher representatives and the broad academic community may have to be actively involved in the reform process. In most situations, examination reform efforts will need support from those who hold major positions of power or influence within a country’s system of government. The political leadership may have to support legislative and budgetary changes, confront opposition from vested interests, and use the media to make the case that proposed reforms are in the best interests of the student, the education system, and the national economy.

“Quick fixes” are unlikely to bring about effective improvements to existing systems. Time will be needed for examination agencies to become familiar with new technologies and administrative processes. In addition, considerable time will be needed to ensure that students and teachers are accustomed to new formats and procedures before taking

the public examinations. Consideration might be given to introducing changes in a phased manner over a period of years (Hill 2013).

Modifying existing examination systems will often require a compromise between the ideal and what is possible and will involve working within constraints of budget, time, and national politics, as well as ironing out conflicts among stakeholders. While working through these constraints, ministries of education and examination systems should keep in mind that good quality examinations can play important constructive roles in educational systems by certifying student achievements, by selecting students, and by providing feedback to the educational system. In particular, well-designed and administered examinations can contribute to the quality of students' educational experiences by dictating what teachers teach and, more important, what students learn.

NOTES

1. Countries that administer the primary school leaving examination include most countries in the Caribbean and some in Africa, as well as Poland and Singapore.
2. As the questions on annual examinations change, it is not possible to claim that the examination tasks are of equal difficulty from one year to the next or that they measure the same content or set of skills. Norm-referenced approaches, if used, mask changes that might have occurred in student achievement levels. Sample-based national assessments (and in some instances international assessments) can provide less expensive and more appropriate methods for monitoring national trends in student achievement.
3. Measurement errors should not be regarded as mistakes as commonly understood; they refer to random variations in student examination scores.
4. They could, for instance, organize regional meetings or attend examination security-related sessions at conferences organized by international bodies such as the International Association for Educational Assessment, the Association for Educational Assessment in Africa, and the Comparative and International Education Society.
5. Media reports indicate that the Arab Republic of Egypt's pilot testing of its new examination system, which involves the use of tablets, encountered problems related to internet connection and computer overload (Al-Youm 2019).

6. For example, following declines in primary and secondary examination pass rates, the Tanzanian government introduced a major reform program in 2013 (World Bank 2018).
7. Examination agencies in Kenya, Tanzania, and Uganda supported short-term staff exchanges and training programs (Kellaghan and Greaney 1992). Currently, the Association for Educational Assessment in Africa promotes cooperation among examination agencies through its conferences and its publication, *Journal of Educational Assessment in Africa*.

REFERENCES

- Al-Youm, A.-M. 2019. "Egypt's New Electronic High School System Fails at First Exam." *Egypt Independent*, March 25. <https://www.egyptindependent.com/egypts-new-electronic-high-school-system-fails-at-first-exam>.
- Baird, J.-A., T. Isaacs, D. Opposs, and L. Gray. 2018. *Examination Standards: How Measures and Meanings Differ around the World*. London: University College London Institute of Education Press.
- Bakker, S. 2014. *The Introduction of Large-Scale Computer Adaptive Testing in Georgia Political Context: Capacity Building, Implementation, and Lessons Learned*. Washington, DC: World Bank.
- Bashir, S., M. Lockheed, E. Ninan, and J.-P. Tan. 2018. *Facing Forward: Schooling for Learning in Africa*. Africa Development Forum Series. Washington, DC: World Bank.
- Bhattacharji, P., and G. Kingdon. 2017. "The Great Indian Exam Debacle." *ORF Issue Brief* 191, July 1. Observer Research Foundation, New Delhi.
- Bhorkar, S., and M. Bray. 2018. "The Expansion and Roles of Private Tutoring in India: From Supplementation to Supplantation." *International Journal of Educational Development* 62: 148–56.
- Black, P. 1996. "Commentary." In *Examining the Examinations*, edited by E. D. Britton and S. A. Raizen, 19–21. Dordrecht, Netherlands: Kluwer Academic.
- China Policy. 2019. "Gaokao: Reform and Reality Clash." *CP Signal*, January 25. <https://policyn.com/wp-content/uploads/2019/01/190125-China-Policy-gaokao-reform-and-reality-clash.pdf>.
- Connolly, B., and D. Cullen. 2017. "European Court Rules That Exam Scripts and Comments Constitute Personal Data." William Fry, December 21. <https://williamfry.com/newsandinsights/news-article/2017/12/21>

/european-court-rules-that-exam-scripts-and-comments-constitute-personal-data.

Hill, P. 2013. *Examination Systems*. Paris: United Nations Educational, Scientific and Cultural Organization.

India Today. 2017. "Shocking! Around 8 Lakh Students Failed in Bihar Board Class 12 Exams." India Today Web Desk, May 31. <https://www.indiatoday.in/education-today/news/story/bihar-board-class-12-980203-2017-05-31>.

Isaacs, T., and K. Gorgen. 2018. "Culture, Context and Controversy in Setting National Examination Standards." In *Examination Standards: How Measures and Meanings Differ around the World*, edited by J.-A. Baird, T. Isaacs, D. Opposs, and L. Gray, 307–30. London: University College of London Institute of Education Press.

Kale, S. 2018. "In India, High-Pressure Exams Are Creating a Student Suicide Crisis." *Wired*, April 20. <https://www.wired.co.uk/article/india-kota-student-suicide-exams-institutes-of-technology>.

Kellaghan, T., and V. Greaney. 1992. *Using Examinations to Improve Education: A Study in Fourteen African Countries*. Washington, DC: World Bank.

———. 2001. *Using Assessment to Improve the Quality of Education*. Paris: United Nations Educational, Scientific and Cultural Organization, International Institute for Educational Planning.

Kellaghan, T., V. Greaney, and T. S. Murray. 2009. *Using the Results of a National Assessment of Educational Achievement*. Washington, DC: World Bank.

Lindsey U. 2003. "Academic, Family Pressure Blamed for Egyptian Teen Suicides." *Daily Star*, August 28. <http://www.dailystar.com.lb/Culture/Art/2003/Aug-28/106212-academic-family-pressure-blamed-for-egyptian-teen-suicides.ashx>.

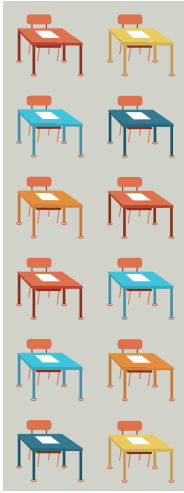
Pells, R. 2017. "Greater Numbers of Teenagers Are Seeking Help Due to Exam Results Stress." *Independent*, August 16. <https://www.independent.co.uk/news/education/education-news/teenagers-exam-stress-gcse-a-levels-numbers-rise-mental-health-a7896011.html>.

Roach, J. 1971. *Public Examinations in England, 1850–1900*. Cambridge, UK: Cambridge University Press.

Spöttl, C., B. Kremmel, F. Holzknacht, and C. Alderson. 2016. "Evaluating the Achievements and Challenges in Reforming a National Language Exam: The Reform Team's Perspective." *Papers in Language Testing and Assessment* 5: 1–22.

Straits Times. 2013. "South Korean School Suicides Total 139 Last Year." August 20. <https://www.straitstimes.com/asia/south-korean-school-suicides-total-139-last-year>.

- Thissen, D., and R. J. Mislevy. 2000. "Testing Algorithms." In *Computerized Adaptive Testing: A Primer*, 2d ed., edited by H. Wainer, N. Dorans, D. Eignor, R. Flaugher, B. Green, R. Mislevy, L. Steinberg, and D. Thissen, 101–33. Hillsdale, NJ: Erlbaum.
- UNESCO (United Nations Educational, Scientific and Cultural Organization). 2013. *The Use of Student Assessment for Policy and Learning Improvement*. Bangkok: UNESCO Asia and Pacific Regional Bureau for Education.
- World Bank. 2018. *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC: World Bank.



GLOSSARY OF ACRONYMS AND TECHNICAL TERMS

Abitur	An academic qualification conferred on German students at the end of secondary school based on student grades and on the written and oral examinations.
ACARA	Australian Curriculum, Assessment and Reporting Authority.
achievement	Describes candidates' results based on examination performance standards.
ACT	American College Testing. A standardized test of English, reading, mathematics, and science that measures high school students' readiness for tertiary-level education in the United States.
Advanced Placement	A College Board program that offers college-level curricula and examinations to high school students in the United States and Canada that can assist high-scoring students in gaining admission to tertiary-level institutions.
AEAA	Association for Educational Assessment in Africa.
AERA	American Educational Research Association.

analysis of variance	A statistical procedure used to determine whether the difference between two or more groups on a single variable (such as mathematics examination marks) is significant or whether the groups can be considered to come from the same population and the difference can be attributed to chance.
analytic scoring	A scoring procedure in which markers or raters evaluate candidates' responses according to a number of features such as content, organization, focus, mechanics, and ideas and assign a score indicating a level of quality to each one.
authentic assessment	A form of assessment that requires candidates to demonstrate mastery of important knowledge and skills by applying them in realistic settings to the types of problems or situations faced by people operating in the "real world."
Baccalauréat	Academic qualification awarded to students at the end of secondary school in France and elsewhere, mainly in French-speaking countries in North and Sub-Saharan Africa.
Bagrut	Matriculation examination, Israel.
Baguwen	Eight-legged essay in Chinese imperial examinations.
CAT	Computer-adaptive testing. On-line testing that adjusts the levels of difficulty of test items presented to students to match their ability levels based on their responses to previous items.
CBSE	Central Board of Secondary Education, India.
chi square	A statistical test of association between two categorical variables, such as gender (males and females) and overall examination results (pass and fail).
CIE	Cambridge International Examinations. More recently known as Cambridge Assessment International Education.

construct-irrelevant variance	Occurs when one or more irrelevant constructs are being assessed in addition to the one that the examination is supposed to be measuring.
construct validity	In the context of examinations, refers to how well a particular examination measures what it claims to measure.
content standards	Broad statements that describe the breadth and level of knowledge, skills, and understanding represented in an examination.
content validity	In the context of a public examination, refers to the extent to which the examination adequately represents the content, knowledge, and skills specified in the syllabus.
correlation	A measure of the degree of relationship or the extent to which two sets of measures, such as examination scores in mathematics and science, co-vary for the same individuals.
criterion-related validity	In the context of examinations, refers to the degree of empirical relationship between examination scores and criterion scores (such as independent standardized test scores or ratings).
CSE	Certificate of Secondary Education. Awarded in England, Wales, and Northern Ireland based on performance at regional examinations. Replaced by the GCSE.
CXC	Caribbean Examinations Council.
DIF	Differential item functioning. A form of statistical analysis used to determine if individuals from different major subgroups (such as boys and girls) with the same latent trait (ability), as measured by an examination or test, have a different probability of giving a certain response to questions or items in an examination or test.
EACEA	Education, Audiovisual and Culture Executive Agency of the European Union.

ecological validity	The extent to which research findings can be generalized to the “real world” or everyday life settings.
Edexcel	Offers GCSE qualifications in England, Wales, and Northern Ireland. Title derives from the words Education and Excellence; privately owned (Pearson).
ENEM	Exame Nacional do Ensino Médio, Brazil.
ESSA	Every Student Succeeds Act. US Department of Education, “Reauthorization of the Elementary and Secondary Education Act,” 2015.
extended essay	Form of assessment in which candidates write extended essays to demonstrate mastery of curriculum content and writing skills.
<i>gaokao</i>	Selection examination for higher education administered at the end of secondary school, China.
GCE A level	General Certificate of Education Advanced Level. Qualification based on examination normally taken by students at the end of the final two years of secondary education in England, Wales, and Northern Ireland. Singapore A level examinations are somewhat similar to UK A levels.
GCE O level	General Certificate of Education Ordinary Level. Qualification based on examinations offered in some British Commonwealth territories. Replaced by the GCSE in England, Wales, and Northern Ireland.
GCSE	General Certificate of Secondary Education. Qualification offered in England, Wales, and Northern Ireland, normally taken by students ages 15–16; generally required before a student can proceed to A-level classes.
grace marks	Additional marks awarded by some examination boards to students scoring near key cutoff decision points to get them to the next scoring level.
<i>hagwons</i>	Private tutoring institutions or cram schools in the Republic of Korea.

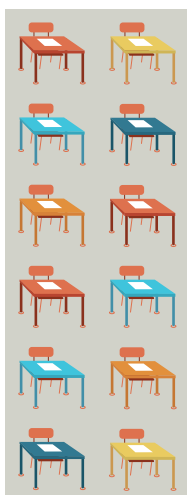
halo effect	The tendency to make specific inferences on the basis of an overall initial impression, such as when a response to the first question in an examination paper affects scores awarded on subsequent questions.
HKEAA	Hong Kong Examinations and Assessment Authority.
holistic assessment	A global approach to assessment that focuses on an entire unit of work, such as an examination essay, rather than on specific elements.
IAEP	International Assessment of Educational Progress
IB	International Baccalaureate examination. Examination based on an International Baccalaureate diploma program curriculum designed for students ages 16–19; uses both external and internal assessments.
interrater reliability	A measure of the extent of agreement between marks awarded by two or more markers for the same assessment.
intrarater reliability	A measure of the extent of agreement between marks awarded by the same marker on two or more occasions for the same assessment.
IRT	Item response theory. Statistical approach that seeks to establish an individual's proficiency level or position on an assumed underlying single trait or ability, such as mathematics, by using characteristics of test items.
IT	Information technology.
Keju	Chinese imperial examination.
levels-based marking scheme	A system of scoring in which a marker classifies a candidate's performance into a single defined level and assigns a grade or mark.
malpractice	General term used to describe actions that interfere with the integrity of an examination and that attempt to gain an unfair advantage for a candidate over others. Other terms include "misconduct," "dishonest conduct," "cheating," "unfair practice," "irregularity," "dishonesty," and "corruption."

Matura	Maturity diploma. School leaving examination offered in countries in Central and Eastern Europe as well as in Italy (Maturita); generally required for admission to tertiary-level education.
measurement error	The difference between a measured quantity and its true value. It includes random error (which is a naturally occurring error that is to be expected in any test or examination situation) and systematic error, which is not based on chance but caused by factors such as poorly designed or administered examinations.
NCLB	No Child Left Behind. US Department of Education, "Reauthorization of the Elementary and Secondary Education Act," 2001.
OECD	Organisation for Economic Co-operation and Development.
Ofqual	Office of Qualifications and Examinations Regulation. Responsible for regulating qualifications, examinations, and assessments in England.
opportunity cost	Economic term that refers to the value of a choice, relative to an alternative, for example, the cost incurred by missing out on benefits associated with work (such as income) when a young person decides to remain in school.
Overall Position	Ranking used to select students for university or colleges of technical and further education, Queensland.
paper-and-pencil test	A form of test or examination that requires students to respond on paper or on scannable answer sheets (as opposed to, for instance, responding on computers or tablets); includes multiple-choice, short answer, and essay-type tests and examinations.
performance assessments	Generally applies to assessments of learning outcomes that require examination candidates to perform a task such as conducting a conversation in a foreign language, constructing an object, or completing a project or portfolio. Markers normally rate the outcome against agreed-upon standards. Also known as authentic assessment.

PIRLS	Progress in International Reading Literacy Study.
PISA	Programme for International Student Assessment.
predictive validity	In the context of examinations, refers to the ability of the scores on the examination to predict later performance (for instance, at university or in a training program).
private candidates	Usually refers to candidates who, at the time of the examination, are not enrolled in the school in which the examination is being administered.
PST	Private supplementary tutoring.
psychometrics	A branch of statistics that deals with the application of statistical and mathematical techniques to educational and psychological testing.
QCAA	Queensland Curriculum and Assessment Authority.
reliability	Refers to consistency of measurement or scoring. See <i>internal consistency</i> , <i>interrater reliability</i> , and <i>intrarater reliability</i> .
SABER	Systems Approach for Better Education Results. World Bank source of comparative data and knowledge on educational policies and institutions.
sampling error	A measure of the error or the difference due to taking a sample from a population rather than using the whole population.
SAT	A standardized test of literacy, writing, and mathematics used for tertiary-level admission in the United States and accepted in some other countries. Previously known as the Scholastic Aptitude Test, subsequently as the Scholastic Achievement Test, and most recently as the SAT.
SBA	School-based assessment. A form of assessment in which teachers' ratings of their students' work contribute to the final examination mark. Frequently subject to external moderation.
school league table	Generally, a ranking of schools based on public examination results.
SES	Socioeconomic status.

shadow education	Instruction in academic subjects provided by tutors for a fee to enhance a student's school progress, including examination performance.
short written supply-type item	Item that requires the candidate to supply a response in the form of a word, numbers, sentence, or short paragraph.
SPBEA	South Pacific Board for Educational Assessment.
SSC	Secondary School Certificate examination in Bangladesh, India, and Pakistan.
standard deviation	A measure of the extent to which numbers, such as examination marks, vary.
standard setting	The process of converting examination marks or scores into reported outcomes such as honors/pass/fail, 1/2/3/4/5, or A/B/C/D/E.
subject pairs analysis	A method of comparing grades of candidates who have taken a particular pair of examination subjects.
table of specifications	Chart that portrays subject content areas on a horizontal axis and cognitive or intellectual skills hierarchically on a vertical axis, designed to help ensure an adequate representation of the content and skills of the subject or domain being assessed.
Tawjihi	General Secondary Education Certificate Examination, Jordan.
<i>thanawiya amma</i>	Secondary school leaving examination in a number of countries in the Middle East and North Africa, including the Arab Republic of Egypt.
TIMSS	Trends in International Mathematics and Science Study.
twenty-first century skills	Various definitions, but generally includes problem solving, critical thinking, creativity, teamwork, and the use of information and communications technology.
UCLES	University of Cambridge Local Examinations Syndicate.

Umalusi	Quality assurance authority for general and advanced education and training, South Africa.
USE	Unified State Examination, Russian Federation.
WAEC	West African Examinations Council.
washback effect	In the context of public examinations, refers to the effects that a high-stakes examination may have on individuals, policies, or practices within a classroom, a school, an educational system, or society as a whole.
z-score	A measure of the number of standard deviations a particular score or data point is from the mean (average).



INDEX

Note: Page numbers followed by f refer to material in figures, t refers to tables, and n refers to notes.

- ability, 25, 81, 110, 111, 135, 153,
175–77, 179, 185, 192, 231,
302–03, 305
- ability tests, 176, 185n2
- Abitur (Germany), 9, 11, 13n2, 19,
22, 29, 32, 53, 174, 254, 301
- ACARA. *See* Australian Curriculum,
Assessment and Reporting
Authority (ACARA)
- accommodations
 - definition and purpose, 240–41
 - disabilities or diverse educational
needs and issues arising
from, 246–49, 247–48t
 - exemptions and reporting, 242–43
 - identifying students with
disabilities or diverse
educational needs for,
241–42
 - types of, for disabilities or diverse
educational needs, 243–46
- accountability, as a function of
examinations, 19, 27, 36, 104,
116n2, 131–38, 224, 281,
290–91
- achieved weighting of examination
components, 155, 159n6
- achievement. *See* student achievement
- achievement tests, 62, 102, 103,
158n1, 255, 308
- ACT. *See* American College
Testing (ACT)
- adaptive testing, 30, 93, 230, 286, 302
- administration
 - of examinations, 50, 75–90, 152,
164, 168, 208, 218–22,
225, 244, 264–65
 - options for improving, 285–91
 - variation in bodies involved in, 29
- advanced-level (A-level)
examinations, 11, 64, 176–77,
304
- advanced placement examinations, 11,
60, 185n1, 301
- advantages, of public examinations,
99–104
- AEAA. *See* Association for
Educational Assessment
(AEAA) (Africa)
- AERA. *See* American Educational
Research Association (AERA)
- affirmative action, 207, 209n5
- Africa
 - administration in, 29
 - AEAA, 296n4, 296n7, 297, 301
 - as a conflict area, 201–2

- balance between external and internal components in, 31
- elimination of primary leaving examination, 285
- emphasis on achievement in, 106
- examination stages in, 28
- history of public examination systems in, 54, 55, 56
- language group membership and, 205
- repetition rates in, 116–17n5
- aggregation of scores, 155–56
- Alaska, computer-based large-scale assessment programs in, 95
- Albania, history of public examination systems in, 58
- A-level examinations. *See* advanced-level (A-level) examinations
- Algeria
 - Baccalauréat administered in, 9
 - controlling and preventing malpractice in, 229
 - examination stages in, 28
- alignment, lack of, as a characteristic of examinations, 149–50
- American College Testing (ACT), 18, 301
- American Educational Research Association (AERA), 301
- analysis of variance, 164, 302
- analytic scoring, 88–89, 302
- ancillary competencies, as a characteristic of examinations, 152
- Anglophone West African countries, academic grades in, 181
- Angola, 13, 28
- answer scripts
 - availability of, 18–19
 - marking of, 88–93
- anxiety, impact of, 85, 108–9, 153
- aptitude tests, 22, 62, 148, 158n1, 176, 177, 185n2, 308
- Arab Republic of Egypt. *See* Egypt, Arab Republic of
- Armenia
 - examination predictability in, 151–52
 - fees in, 30, 37n5
 - number of subjects taken in, 32
- Asia. *See also* East Asia; South Asia
 - administration in, 29
 - examination stages in, 28–29
 - history of public examination systems in, 54–55
 - written examinations in, 43
- assessment capacity, 48, 115, 137n2, 282, 292
- assessment practices, improving, 279–91
- Association for Educational Assessment (AEAA) (Africa), 296n4, 297n7, 301
- Australia
 - aligning marks and standards in, 185n6
 - comparability issues in, 179, 180
 - comparability of grades in, 174
 - examination reform in, 182
 - GMSAT, 37–38n7
 - incentive feedback in, 131
 - place of residence as an equity issue in, 200
 - rules for high school graduation in, 255–57
 - SBA certification in, 255, 281
 - system assessment and public examinations systems in, 4
 - technology and assessment in, 94, 233n4
- Australian Curriculum, Assessment and Reporting Authority (ACARA), 301
- Austria, student responses in, 33
- authentic assessment, 262, 283, 284, 302
- Azerbaijan, selection and certification in, 23–24, 57
- Baccalauréat (France), 9–11, 19, 22, 29, 34, 53, 55, 68n9, 150, 164, 193, 302
- Bagrut (Israel), 25–26, 302
- Baguwen, 45, 302
- Baluchistan, malpractice in, 223
- Bangladesh
 - administration in, 29
 - balance between external and internal components in, 32
 - examination types administered in, 9

- focus of public examinations in, 107
 - history of public examination systems in, 54–55
 - malpractice in, 223
 - multiple boards and comparability of grades in, 174
- Beijing, malpractice in, 225
- Belgium
 - balance between external and internal components in, 31
 - lack of curriculum-based external examinations in, 3
- Bhojpuri, language group membership and, 205
- Bhutan
 - balance between external and internal components in, 32
 - controlling and preventing malpractice in, 229–30
- bimbel*, 112
- Bloom's Taxonomy, 77, 79, 80f, 106
- Boko Haram, impact on examinations of, 202
- Bombay, history of public examination systems in, 54
- bonus marks, as motivation, 25–26
- Book of Rites*, 44
- Booth, James, 52
- Boston Public Schools, 59
- Botswana
 - Cambridge Assessment International Education (CIE) and, 64
 - examination stages in, 28
 - primary school leaving examinations in, 12–13n1
- Brazil
 - history of public examination systems in, 58–59
 - selection in, 24
- Brethren of the Common Life, 49–50
- Britain. *See* England; United Kingdom
- Brunei Darussalam
 - examination stages in, 28–29
 - primary school leaving examinations in, 12–13n1
- budgets, for public examinations, 3, 296
- Burkina Faso, repetition rates in, 116–17n5
- Burundi, examination stages in, 28
- buxiban*, 112
- Calcutta, history of public examination systems in, 54
- California, computer-based large-scale assessment programs in, 95
- Cambodia
 - academic grades in, 181
 - tutoring in, 115
- Cambridge Assessment International Education (CEI), 64, 89, 144, 176, 177, 178, 227
- Cambridge International Examinations (CIE), 37n3, 64, 302
- Cambridge Tripos, 51–52, 67n4, 104
- Cameroon, examination stages in, 28
- Campbell's Law, 156
- Canada, SBA certification in, 255
- candidates
 - characteristics of, 153
 - document requirements for special needs, 242
 - special needs. *See* disabilities or diverse educational needs
- Caribbean
 - administration in, 29
 - history of public examination systems in, 54
 - primary school leaving examination in, 296n1
 - secondary school placement in, 7
 - selection in, 23
- Caribbean Examinations Council (CXC), 55–56, 265, 303
- caste-based reservation, 209n5
- CAT. *See* computer-adaptive testing (CAT)
- Cele, Johannes, 49
- Central Board of Secondary Education (India), 20, 76–78, 302
- Central Council of Education (Japan), 133
- centralized marking, 92
- Certificate of Secondary Education (CSE), 303
- certification
 - and validity, 149, 157–58
 - and reliability, 169

- as a function of examinations, 3, 7, 19–21
- characteristics
 - of candidates and validity, 153
 - of examinations and validity, 149–52
 - of public examinations, 17–19
- cheating. *See* malpractice
- Ch'eng-Chu Tao learning, 48
- chief examiner recommendations, 125–27
- Chile
 - factors related to change, 183
 - history of public examination systems in, 59
- chi square statistic, 192, 302
- China
 - additional tuition or cram schools in, 112, 115, 198
 - civil service selection in, 44, 53
 - disabilities or diverse educational needs in, 239
 - equity issues in, 191, 193, 200
 - examination-related stress in, 108–9
 - gaokao* examination, 26–27, 49, 198, 199, 229, 281, 304
 - Higher Education Entrance Examination, 49
 - history of examination systems in, 3, 44–49
 - malpractice in, 225, 229
 - number of subjects taken in, 32
 - pilot program in, 283
 - reform in, 18, 281
 - selection for university in, 23
 - written examinations in, 43, 44–49
- Ch'ing dynasty, 46
- choice
 - as a characteristic of examinations, 150–51
 - differences in, 34–35, 38n8
- City and Guilds of London Institute, 67n6
- civil service assessment, 44, 53, 55, 68n8, 117n6
- classroom assessment, 4–6, 259, 279
- cognitive levels, 26, 77–81, 110, 146, 168
- College of Preceptors, 67n6
- Collegium Romanum, 50
- collusion, 109, 219, 221, 226, 229
- Colombia, history of public examination systems in, 58
- comparability issues, 29, 35, 56, 151, 169, 175–80, 183, 184, 253
- Comparative and International Education Society, 296n4
- computer-adaptive testing (CAT), 30, 93–94, 230, 286, 302
- concentration, spread of examinations versus, 33
- confidence, public, 100, 183–84, 282, 288, 291, 292
- conflict areas, examinations in, 201–3
- Confucianism, 45, 48, 67n3
- construct-irrelevant variance, 95, 150, 151, 152, 159, 303
- construct validity, 147–48, 303
- constructed response items, 33, 84–85, 195
- construction
 - examination syllabus, 75–77
 - of examinations, 77–81
 - question paper production and printing, 86–88
- content standards, 7, 173, 184, 303
- content underrepresentation, as a characteristic of examinations, 149–50
- content validity, 145–47, 157, 303
- contextualized achievement approach, to estimating effectiveness, 135
- contextualized value-added approach, to estimating effectiveness, 136
- controlling activities, as a function of examinations, 26–27
- copying, 50, 109, 220–21, 226, 234n9. *See also* plagiarism
- correlation, 303
- corruption. *See* malpractice
- cost-effectiveness, 60, 66, 293
- costs, 30–31, 87, 92, 94, 167, 265, 285, 288–89
- Côte d'Ivoire, repetition rates in, 116–17n5
- coursework, 83, 84. *See also* school-based assessment (SBA)
- cram-type schools, 112, 197, 198, 304

- Creole, language group membership and, 205
- criterion-related validity, 146–47, 303
- Croatia, examination levels in, 33
- CSE. *See* Certificate of Secondary Education (CSE)
- cultural revolution, in China, 49
- curriculum-based assessment, 3, 7, 8–9, 58, 102, 103, 279. *See also* content validity
- CXC. *See* Caribbean Examinations Council (CXC)
- cybersecurity, malpractice and, 221, 228–29, 232
- Cyprus, statistical adjusting in, 178, 180
- Czech Republic
centralized system in, 26–27
examination formats and levels in, 33
- Democratic Republic of Congo,
repetition rates in, 116–17n5
- Denmark
academic grades in, 181
compulsory examinations in, 18
student responses in, 33
- dershane*, 112
- developing economies, history of
public examinations in, 54–56
- differential item functioning (DIF)
analysis, 192, 303
- disabilities or diverse educational needs
document requirements for
candidates with, 242
exemptions, 242–43
identification of students with,
241–42
issues in accommodation for,
246–48
overview of, 239–41
types of accommodation for,
243–46
- disadvantages, of public examinations,
2, 99–100, 104–15
- dishonesty. *See* malpractice
- dispersion of marks, 148, 155
- district-based merit, 209n5
- Djibouti, abolishing basic level
examinations in, 29
- domain underrepresentation, 104–5
- Dominican Republic, history of public
examination systems in, 59
- Durham University, 53
- EACEA. *See* Education, Audiovisual
and Culture Executive Agency
of the European Union
(EACEA)
- East Asia. *See also* specific countries
selection for university in, 22, 23
- Edexcel (Education and Excellence),
64, 304
- Edinburgh School of Arts, 37n2
- Education, Audiovisual and Culture
Executive Agency of the
European Union (EACEA),
303
- Educational Testing Service, 158n1
- efficiency, as a characteristic of good
examination systems, 293
- ego (performance) goals, 110
- Egypt, Arab Republic of
controlling and preventing
malpractice in, 230
examination-related stress in,
108–9
examination stages in, 28–29
history of public examination
systems in, 55
information technology in, 93
number of subjects taken in, 32
pilot testing of examination system
in, 296n5
tutoring in, 112
- electronic devices and malpractice,
219–20, 221, 227–28, 230,
231, 232, 289
- encrypted question papers, 227
- Endorsed National Senior Certificate
(South Africa), 243
- England. *See also* United Kingdom
comparability of grades in,
174–80
examination boards in, 11, 37n4,
64, 174, 183
examination levels in, 33
General Certificate of Education
(GCE), 11, 29, 33, 90t,
164, 304

- General Certificate of Secondary Education (GCSE), 84, 132–33, 134f, 176–77, 304
- history of public examination systems in, 51–53, 64, 66, 67n4–8, 195
- incentive feedback in, 131–38
- Joint Matriculation Board public examinations, 195
- Office of Qualifications and Examinations Regulation (Ofqual), 11, 266, 306
- performance assessment in, 83, 84, 258–59, 280, 283–84, 306
- publication of results in, 116n2, 133, 134f, 135, 290–91
- retrenchment to traditional assessment forms in, 84
- school league tables in, 134f, 135
- selection for university in, 24
- English, language group membership and, 205
- equity issues
 - integrity, 310–12
 - characteristic of good examination systems, 292–93
 - ethnic group membership, 204–7
 - gender, 193–96
 - language group membership, 204–7
 - overview of, 191–93
 - place of residence, 199–203
 - socioeconomic status, 30, 115, 196–98
 - types of schools, 203–4
- escuela de cursos intensivos*, 112
- ESSA. *See* Every Student Succeeds Act (ESSA) (2015)
- essay-type questions and examinations, 10, 30, 32–33, 59, 65, 85, 157, 164, 167, 195, 225, 283
- essays, extended, 81–82, 304
- Estonia, feedback provided in, 124
- Ethiopia
 - controlling and preventing malpractice in, 229
 - examination stages in, 28
 - examinations as a political tool in, 224
 - fees in, 30
 - repetition rates in, 116–17n5
- ethnic group membership, 204–7, 208
- Europe. *See also* specific countries
 - balance between external and internal components in, 31
 - examination stages in, 28–29
 - history of public examination systems in, 3, 8, 9, 49–54, 56–57, 67n5
- Every Student Succeeds Act (ESSA) (2015), 61–62, 304
- Exame Nacional do Ensino Médio (Brazil), 58–59, 304
- examination boards, 11, 18, 29, 30, 37n4, 53, 55–56, 64, 107, 138, 174, 180, 182, 217, 228, 288, 290, 293, 294
- examination items and questions
 - constructing tasks, 77–84
 - differences between item types, 84–85
 - for accommodating disabilities or diverse educational needs, 244
 - options for improving, 282–85
 - predictability in, 151–52
- Examination Malpractice Act (Nigeria), 229
- examination-related costs. *See* costs
- Examination Statute of the University of Oxford, 52
- examination syllabus, 11, 56, 75–77, 80, 94, 145–46, 151, 180, 256, 257f
- examination systems
 - characteristics of good, 291–94
 - differences among, 28–35
 - variety in, 6–8
- examination-taking skills, focus on, 107, 138
- examinee-related factors, effect on reliability of, 164–65
- examinee responses, forms of, 81–85
- examiner-related factors, effect on reliability of, 165
- exemptions, for disabilities or diverse educational needs, 242–43
- exit examinations. *See* public examinations
- exit tests (United States), 62

- extended essays, 81–82, 304
- external assistance, 219–21
- external components, balance
 - between internal components and, 31–32
- external examinations. *See* public examinations
- extrinsic motivation, 25, 111

- Farish, William, 51
- feedback
 - guidance, 124–30, 283, 290, 294, 296
 - incentive, 131–38, 139n6
 - markers, 92
 - using results of performance to provide, 290
- fees
 - differences in, 30–31, 37n6, 63–64
 - socioeconomic status and, 197, 208, 292
- Fiji, comparability problems in, 180
- fingerprint-recognition, 225, 228
- Finland
 - accountability in, 104
 - balance between external and internal components in, 31
- fitness of purpose, as a characteristic of good examination systems, 291
- Five Classics, 45, 67n2
- foreign materials, smuggling of, 220
- Four Books, 45, 67n2
- France
 - academic grades in, 19, 181
 - administration in, 29
 - Baccalauréat, 9–11, 19, 22, 29, 34, 53, 55, 68n9, 150, 164, 193, 302
 - balance between external and internal components in, 31
 - certification in, 19, 36
 - choices in, 35
 - civil service selection in, 53
 - feedback provided in, 124
 - gender equity issues in, 193
 - history of public examination systems in, 3, 9, 53, 55, 60, 191, 193
 - number of subjects taken in, 32
 - options in, 34, 150
 - reliability of marking in, 164
 - selection for university in, 22
- French, language group membership and, 205
- frontistirio*, 112
- functions, of public examinations, 19–28, 35–36. *See also* feedback

- The Gambia, abolition of public examination in, 255
- gaokao* examination (China), 26–27, 49, 198, 199, 229, 281, 304
- Gaza, as a conflict area, 202–3
- GCE. *See* General Certificate of Education (United Kingdom)
- GCSE. *See* General Certificate of Secondary Education (United Kingdom)
- gender equity issues, 192, 193–96, 208, 224, 225, 264
- general bias, effect on marking of, 166
- General Certificate of Education (Sri Lanka), 19, 20
- General Certificate of Education (United Kingdom), 11, 29, 33, 89, 90t, 134, 137, 164, 176–77, 304
- General Certificate of Secondary Education (United Kingdom), 84, 132, 137, 139n7, 164, 176, 304
- General Medical Council, 67n6
- general scholastic ability, 175. *See also* ability
- General Secondary Education Certificate Examination, Jordan (Tawjihi), 108, 308
- Georgia
 - history of public examination systems in, 57
 - information technology in, 93
- Germany. *See also* Abitur
 - academic grades in, 19
 - administration in, 29
 - balance between external and internal components in, 32
 - certification in, 19
 - choices in, 35

- comparability of grades in, 174, 176
 - compulsory examinations in, 18
 - history of public examination systems in, 3, 66
 - number of subjects taken in, 32
 - publication of answer scripts in, 18–19
 - selection for university in, 22
 - student responses in, 33
- Ghana
 - history of public examination systems in, 68
 - incentive feedback in, 131–32
 - repetition rates in, 116–17n5
- ghost centers, 222
- Gladstone, William, 52
- Global Disability Summit, 239, 240
- Global Monitoring Report*, 193
- GMSAT. *See* Graduate Medical School Admission Test (GMSAT)
- government printers, 87t, 289
- grace marks, 154, 287, 304
- grade thresholds, setting of, 89
- grading, comparability of, 174–83.
 - See also* scoring; marking
- Graduate Medical School Admission Test (GMSAT), 37–38n7
- Great Britain. *See* England; United Kingdom
- Greece
 - additional tuition or cram schools in, 112
 - balance between external and internal components in, 31
 - disabilities or diverse educational needs in, 243
 - selection and certification in, 23
- Guangxu reform (1898), 45
- guidance feedback, 124–30, 156, 267
- hagwons*, 115, 304
- halo effect, 166, 305
- Higher Education Admissions Center (Oman), 101
- Higher Education Entrance Examination (China), 49
- Higher School Certificate Examination (New South Wales, Australia), 182
- Higher Secondary School Certificate, 29
- Hindi, language group membership and, 205
- history, of written examinations, 43–67
- holistic assessment, 88–89, 305
- Hong Kong SAR, China
 - comparability problems in, 179
 - disabilities or diverse educational needs in, 242, 243, 245
 - Examinations and Assessment Authority (HKEAA), 127, 128, 305
 - feedback provided in, 127, 128
 - reforms in, 18
 - score adjustment, 179
 - selection for second level in, 23
 - tutoring in, 112, 114
- Hungary, selection and certification in, 23–24
- IAEP. *See* International Assessment of Educational Progress (IAEP)
- IB. *See* International Baccalaureate (IB)
- Iceland
 - balance between external and internal components in, 31
 - feedback provided in, 124
- identification
 - cards, 219, 228
 - numbers, 207, 228, 230
 - preventing impersonation in regard to, 228
 - students with disabilities or special needs, 241–42, 245
- impact, on educational processes and outcomes, 156
- impersonation, 109, 208, 219, 223, 228
- improper assignment, 222
- incentive feedback, 131–38, 139n6
- income generation, as a function of examinations, 27, 31. *See also* private supplementary tutoring (PST); shadow education
- inconsistent marking, 163, 165–68, 269
- India. *See also* specific cities
 - administration in, 29, 37n4
 - balance between external and internal components in, 32

- caste-based reservation in, 209n4
- Central Board of Secondary Education, 302
- certification in, 20–21
- comparability of grading in, 174, 181
- controlling and preventing malpractice in, 227, 228, 233n4
- equity issues in, 191, 209n5
- examination-related stress in, 108–9
- focus of public examinations in, 107
- grace marks in, 154
- history of public examination systems in, 54–55
- language group membership in, 205
- malpractice in, 220, 223
- number of subjects taken in, 32
- publication of marking/scoring keys in, 18
- question paper design in, 78
- reducing amount of assessment in, 32
- syllabus in, 75–77
- tutoring in, 112, 113
- Indian Central Board of Secondary Education, 20–21, 76–78
- Indonesia
 - additional tuition or cram schools in, 112
 - controlling and preventing malpractice in, 227
 - examination stages in, 28
- information technology (IT)
 - challenges of, 219–20, 221, 223, 248, 232, 293, 305
 - using in examination systems, 30, 64, 93–95, 230, 232, 289–90, 292
- in-house printing facilities, 86, 87t
- integrity
 - as a characteristic of good examination systems, 292–93
 - controlling and preventing malpractice, 227–31
 - forms of malpractice, 218–22
 - frequency of malpractice, 222–23
 - overview, 217–18
 - reasons for malpractice, 223–24, 225
 - strategies to discourage malpractice, 224–26
- intended weightings of examination components, 155, 156, 157
- internal components, balance between external components and, 31–32
- international assessment. *See*
 - Programme for International Student Assessment (PISA); Progress in International Reading Literacy Study (PIRLS); Trends in International Mathematics and Science Study (TIMSS)
- International Assessment of Educational Progress (IAEP), 102, 296n4, 305
- International Association for the Evaluation of Educational Achievement Reading Literacy Study, 102
- International Baccalaureate (IB), 11, 29, 59, 63, 64, 83–84, 255, 305
- international examinations, 63–64
- International Examinations Inquiry, 60
- international specialist security printers, 87t
- interrater reliability, 153–54, 164, 167, 305
- intimidation, 221, 208, 222, 230
- intrarater reliability, 153–54, 305
- intrinsic motivation, 25, 103, 111
- invigilators, 31, 109, 219, 225, 226, 228
- Iraq, controlling and preventing malpractice in, 229
- Ireland
 - accountability in, 104
 - balance between external and internal components in, 31
 - choices in, 35
 - comparability of grades in, 176
 - construct validity, 147–48
 - controlling and preventing malpractice in, 229, 233n5

- disabilities or diverse educational needs in, 241, 244, 245
- examination levels in, 33
- feedback provided in, 124, 125, 131
- Graduate Medical School
 - Admission Test (GMSAT), 37–38ⁿ⁷
- Leaving Certificate Examination, 25–26, 35, 106, 125, 155, 164, 175, 179, 185ⁿ⁵
- motivation in, 25–26
- number of subjects taken in, 32
- publication of answer scripts in, 18–19
- selection in, 101
- standards, 176
- table of specifications, 145
- test taking strategies, 153–54
- timing accommodations for
 - disabilities or diverse educational needs in, 245
- weights intended and achieved, 155–56
- iris-recognition data, 228
- Irish Central Applications Office, 101
- irregularity. *See* malpractice
- IRT. *See* item response theory (IRT)
- Islamic Republic of Iran, language
 - group membership and, 206
- isolation, of printers, 88^t
- Israel
 - as a conflict area, 202–3
 - Bagrut, 25, 302
 - choices in, 35
 - motivation in, 25
 - number of subjects taken in, 32
- IT. *See* information technology (IT)
- Italy
 - feedback provided in, 124
 - incentive feedback in, 131
 - student responses in, 33
- item bias, 192, 196, 207
- item response modeling, 185ⁿ⁴
- item response theory (IRT), 62, 305
- equity issues in, 191
- incentive feedback in, 133
- Jesuits, 50
- Joint Matriculation Board public examinations (England), 195
- Jordan
 - abolishing basic level examinations in, 29
 - impact of large-scale reform program in, 107–8
- juku*, 112
- Junior Certificate Examination (Ireland), 179, 185ⁿ⁵
- “just-in-time” printing, 87^t
- Karnataka, 113
- Kazakhstan
 - fees in, 30
 - number of subjects taken in, 32
- Keju, 44, 48, 305
- Kelly, Frederick J., 59–60
- Kenya
 - as a conflict area, 202
 - controlling and preventing malpractice in, 229
 - equity in, 272ⁿ¹
 - examination stages in, 28
 - incentive feedback in, 131, 137–38
 - information technology in, 93
 - primary school leaving examination in, 296ⁿ¹
 - repetition rates in, 116–17ⁿ⁵
 - secondary school placement in, 7
 - short-term staff exchanges and training programs in, 297ⁿ⁷
 - tutoring in, 112, 115
- Kenya Certificate of Primary Education, 128–30
- Kenya National Examinations Council, 127, 139ⁿ⁵
- Kenyan National Examination Commission, 93
- keylogging, 233ⁿ⁶
- knowledge underrepresentation of domain, 149–50
- Korea, Republic of
 - additional tuition or cram schools in, 112
 - examination-related stress in, 108–9
 - selection for second level in, 23

- tutoring in, 112, 115
- World Education Forum, 8
- Kuwait
 - abolishing basic level examinations in, 29
 - fees in, 30
- language group membership, 192, 204–7, 208
- Latin America, examinations in, 58–59, 66, 116
- Latin Schools, 50
- leakage, 3, 87, 88, 218, 227, 230
- learning
 - as a characteristic of good examination systems, 294
 - enhancement of, 27, 99, 115, 133, 185, 284, 291
 - influence on, 55, 107, 204
 - information on, 57, 127, 269
 - role of motivation in, 103
 - standards, 124, 206
- learning (mastery) goals, 110
- Leaving Certificate Examination (Ireland). *See* Ireland, Leaving Certificate Examination
- Lebanon, fees in, 30
- legitimizing membership in global society, as a function of examinations, 28
- Lesotho, examination stages in, 28
- levels-based marking schemes, 89–91, 305
- Liberation Tigers of Tamil Eelam, 203
- Liberia, examination stages in, 28
- Lisbon Recognition Convention, 11
- Lithuania
 - controlling and preventing malpractice in, 230
 - feedback, 126
 - National Examination Center, 126
 - reliability in marking in, 167
- Liverpool Collegiate Institute, 52
- local commercial printers, 87t
- London University Matriculation examinations, 67–68n7
- low-income countries, 2, 193, 199, 204, 282
- Lowe, Robert, 52, 139
- Loyola, Ignatius, 50
- Madras, history of public examination systems in, 54
- Malawi
 - controlling and preventing malpractice in, 229
 - district-based merit in, 209n4
 - examination stages in, 28
 - repetition rates in, 116–17n5
- Malaysia
 - allocation in, 23
 - ethnic group membership and, 204
 - examination stages in, 28–29
 - gender equity issues in, 193–94
- Mali
 - Baccalauréat administered in, 7
 - as a conflict area, 202
- malpractice
 - as a disadvantage of public examinations, 109
 - in awards, 222
 - in certificates, 222
 - controlling and preventing, 227–31
 - defined, 306
 - forms of, 218–22
 - frequency of, 222–23
 - in marking, 222
 - reasons for, 223–24, 225
 - socioeconomic status and, 197
 - strategies to discourage, 224–26
- Malta, publication of marking and scoring keys in, 18
- management, as a function of examinations, 27
- Mann, Horace, 59
- Maotanchang, China, 198
- Marathi, language group membership and, 205
- marketization, 112
- marking. *see also* scoring
 - additional marks as motivation, 25
 - centralized, 92, 230
 - equity, 207
 - holistic, 88–89
 - human factors affecting, 164, 166
 - income, 27
 - information technologies and, 93–95
 - item level, 168
 - malpractice in, 222
 - multiple, 167, 168

- of answer scripts, 88–93
 - on-screen, 90, 92–93, 168, 289
 - paper scripts at home, 90, 92
 - publishing of scoring keys for, 18–19
 - reliability of, 165, 167, 168. *See also* interrater reliability; intrarater reliability
 - schemes, 75, 77, 86, 89–91, 115, 126, 167, 256, 259
 - school-based assessment, 254, 264–65
 - standards, 168, 256, 267, 286
 - teacher proficiency in, 292
- mastery, grouping students based on, 49
- mastery (learning) goals, 110
- Matura, 306
- Mauritania, Baccalauréat administered in, 9
- Mauritius
 - Cambridge Assessment International Education (CIE) and, 64
 - examination stages in, 28
 - feedback provided in, 126
 - fees in, 30
 - language group membership and, 205
 - overseas examinations administered in, 37n3
 - tutoring in, 112, 115
- measurement errors, 264, 287, 296n3, 306
- measurement methods, as a characteristic of examinations, 150
- memorization, 2, 48, 55, 106, 107, 110, 283
- Mencius, 67n2, 67n3
- merit, 44, 48, 57, 131, 135, 155, 170, 200, 201, 209n3
- Mexico, history of public examination systems in, 58
- Middle East. *See also* specific countries
 - administration in, 29
 - examination spread in, 33
 - examination stages in, 28–29
 - selection in, 23
- Ming dynasty, 45
- Ministry of Education National Examinations Center (Georgia), 93–94
- misconduct. *See* malpractice
- moderation, of SBA, 268–69
- Mogadishu, as a conflict area, 203
- monitoring educational standards over time, as a function of examinations, 24–25
- Montaigne College, 49
- Morocco
 - abolishing basic level examinations in, 29
 - Baccalauréat administered in, 9
- motivation
 - as a function of examinations, 25–26, 36, 66
 - extrinsic, 103, 111
 - impact of on reliability, 165
 - impact of on school, 135
 - impact of on validity, 235
 - in public examinations, 111
 - intrinsic, 111
 - school-based assessment, 263
- Mozambique, repetition rates in, 116–17n5
- multiple-choice items
 - computerized testing of, 93–95
 - costs, 30
 - differences between item types, 84–85
 - discrimination issues, 143–44
 - gender differences and, 85, 195–96
 - history of, in Africa, 56
 - history of, in Latin America, 58–59
 - history of, in United States, 59, 60, 61, 66
 - international examinations in, 64
 - impact of on teaching, 107
 - limitations of, 85, 283
 - overview, 82, 83, 84–85
 - replacing essay-type examinations with, 283
 - special needs candidates and, 245
- Myanmar, 55, 115
- Namibia, Cambridge Assessment International Education (CIE) and, 64
- Napoleon Bonaparte, 10, 53, 66
- National Assessment of Educational Progress scores, 5, 81
- national assessments, 5, 59, 81, 116, 175–76, 254, 296n2

- National Certificate of Education (New Zealand), 33
- National Examination Center (Lithuania), 126
- national examinations. *See* public examinations
- NCLB. *See* No Child Left Behind (NCLB) Act (2001)
- Nepal
 - as a conflict area, 204
 - examination reform in, 55
 - fees in, 30
 - private and public schools in, 204
- Netherlands
 - accountability in, 104
 - student allocation in, 23
 - subject weighting in, 32
- New South Wales. *See also* Australia
 - aligning marks and standards in, 185n6
 - examination reform in, 182
 - plagiarism in, 264
 - school-based assessment in, 268
- New Zealand
 - balance between external and internal components in, 32
 - examination levels in, 33
 - examination spread in, 33
 - influence on SPBEA in, 56
 - publication of answer scripts in, 18–19
 - SBA certification in, 255
- Nigeria
 - as a conflict area, 202
 - controlling and preventing malpractice in, 229
 - focus of public examinations in, 107
 - language group membership and, 205
 - malpractice in, 223
 - place of residence as an equity issue in, 201
 - repetition rates in, 116–17n5
 - tutoring in, 115
- No Child Left Behind (NCLB) Act (2001), 61, 62, 306
- norm-referenced approaches, 147, 241, 296n2
- North Africa. *See also* specific countries
 - examination stages in, 28–29
- Northern Ireland
 - administration in, 11
 - examination levels in, 33
 - publication of results in, 116n2
 - school league tables in, 132–33
- Norway
 - balance between external and internal components in, 32
 - centralized system in, 6
 - “open-book” examinations, 37n1
 - reforms in, 18
- objectives
 - curriculum, 35, 80, 104, 169, 291
 - levels of, 90. *See also* Bloom’s taxonomy
 - political 57, 287
 - syllabus, 11, 75, 256
- OECD. *See* Organisation for Economic Co-operation and Development (OECD)
- Office of Qualifications and Examinations Regulation (Ofqual) (England), 11, 266, 306
- Oman
 - examination spread in, 33
 - gender equity issues in, 193, 194
 - Higher Education Admissions Center, 101
 - selection in, 101
- on-screen marking, 90, 92–93, 168, 289
- “open-book” examinations, 37n1
- opportunity costs, 197, 200, 209n1, 265, 281, 288, 306
- optical scanners, 230
- oral assessment,
 - administering, 84
 - candidates with diverse needs, 243, 246
 - emphasis on, 3, 33, 26, 104, 280, 283
 - history of, 9, 13, 18, 33, 50, 51, 53, 56, 59, 64, 67n4, 68n9, 131
 - improving, 284, 294
 - school-based assessment, 258, 261
 - scoring or marking, 31, 155, 265
- Organisation for Economic Co-operation and Development (OECD), 54, 124, 147–48, 204, 223, 306

- Ottoman Turks, 55
- outcomes-based paradigm, 8
- overall position, 257, 306
- Oxford and Cambridge School Examinations Boards (“Joint Boards”), 53
- Pacific region. *See also* specific countries
 - balance between external and internal components in, 31
 - selection for entrance in, 22
- packaging, 88t
- Pakistan
 - administration in, 29
 - comparability of grades in, 174
 - examination predictability in, 152
 - examination types administered in, 9
 - fees in, 30
 - focus of public examinations in, 107
 - grace marks, 154
 - history of public examination systems in, 54–55
 - malpractice in, 223
 - reliability in marking in, 167
- Palestine, as a conflict area, 202–3
- paper-and-pencil instruments, 149, 150, 157, 158n2, 306
- perception difference, effect on marking of, 166
- performance assessment, 33, 83, 84, 85, 180–83, 258, 259, 260, 307
- pilot testing, 94, 144, 282, 286, 290, 296
- PIRLS. *See* Progress in International Reading Literacy Study (PIRLS)
- PISA. *See* Programme for International Student Assessment (PISA)
- place of residence or location, as an equity issue, 199–203, 208, 224, 225
- plagiarism, 109, 220–21, 226, 264. *See also* copying
- “points” system, 35, 101
- Poland
 - examination levels in, 33
 - examination stages in, 28–29
 - history of public examination systems in, 57
 - political leadership, malpractice and, 224, 230, 282
 - primary school leaving examination in, 296n1
 - reforms in, 8
 - student achievement information on, 7
- Portugal
 - compulsory examinations in, 18
 - feedback provided in, 124
- predictive validity, 137, 148, 307
- preparatory courses, 24, 57
- primary school leaving examination, 12–13n1, 116–17n5, 137, 153, 195, 202, 281, 285, 296n1
- printer isolation, 88t
- printing, of papers, 55, 86–88, 87t, 88t
- private or external candidates, 219, 293, 307
- private schools, 26, 59, 204, 192, 290
- private supplementary tutoring (PST), 31, 57, 111–15, 307. *See also* shadow education
- production, of question papers, 86–88
- professional development
 - examination personnel of, 295, 297n7
 - teachers of, 56, 255, 262, 265, 266, 267, 272, 290, 294
- Programme for International Student Assessment (PISA), 102, 137, 147–48, 307
- Progress in International Reading Literacy Study (PIRLS), 206, 307
- proofreading, 86, 88t, 218
- Prussia
 - Abitur in, 9
 - civil service selection in, 53
 - impact on Mann, 59
 - university entrance qualifications, 53
- PST. *See* private supplementary tutoring (PST)
- psychometric approaches, 8, 61, 81, 143–44, 163, 178, 180, 185n4, 307

- public examinations
 - abolishing, 285
 - advantages of, 100–04
 - budgets for, 3
 - characteristics of, 17–19, 149–52
 - construction of, 77–81
 - differences between system
 - assessments and, 5
 - disadvantages of, 2, 104–15
 - functions of, 3, 19–28
 - levels in, 33
 - major traditions in, 8–11
 - options for taking, 34, 286
 - SBA compared to, 262
 - security of, 289
 - stages for, 28–29
- public schools, private schools versus, 204
- Punjab, fees in, 30
- QCAA. *See* Queensland Curriculum and Assessment Authority (QCAA)
- quality assurance, 257, 267
- Queensland, Australia
 - high school graduation process in, 255–57
 - reform in, 281
- Queensland Curriculum Assessment Authority (QCAA), 256, 257, 307
- quota systems, 111, 194, 200–201, 204, 207, 208, 224, 225
- Race to the Top Act (2009), 61
- random error, systematic error versus, 158n5
- random fluctuation, effect on marking of, 166
- rater drift, effect on marking of, 166
- Ratio Studiorum, 50
- real-life situations, reflecting, 283–84. *See also* authentic assessment
- Rede, C. A., 37n3
- reference test, 177, 266, 269, 270
- reform of examinations
 - characteristics of good examination systems, 291–94
 - considerations for, 280–82
 - impact of, 107–8
 - options for improving
 - administration and reporting, 285–91
 - options for improving examination items and questions, 282–85
 - overview, 279–80
- Regents Examination (New York), 60
- reliability. *See also* interrater reliability; intrarater reliability
 - defined, 307
 - effect of examinee-related factors
 - on, 164–65
 - effect of examiner-related factors
 - on, 165
 - effect of subject-related factors on, 165–67
 - improving, 149, 168
 - overview, 163–64
 - school-based assessment aspect of, 266
 - type of assessment, 84, 283, 284
- repeating grades, 2, 116–17n5, 285, 288
- reporting
 - modifying, 286–87
 - options for improving, 285–91
- Republic of Korea. *See* Korea, Republic of
- Republic of Yemen
 - abolishing basic level examinations in, 29
 - as a conflict area, 203
- Resources, devoted to examination preparation, 111–15
- response methods, for
 - accommodating special needs candidates, 244
- Ricci, Matteo, 50, 51
- Romania, history of public examination systems in, 57, 152
- Royal Institute of Chemistry, 67n6
- Royal Institute of Civil Engineers, 67n6
- Royal Society of Arts, 67n6
- rubrics, 82, 166, 180, 284. *See also* marking, schemes

- Russian Federation
 controlling and preventing malpractice in, 228
 examination-related stress in, 108–9
 feedback provided in, 124–25
 history of public examination systems in, 58
 incentive feedback in, 131
 language group membership and, 206
 malpractice in, 223, 228
 number of subjects taken in, 32
 place of residence as an equity issue in, 200
 revenue losses in, 24
 student allocation in, 23
 Unified State Examination (USE), 31, 58, 124, 131, 139n2, 200, 206, 223, 309
- Rwanda, repetition rates in, 116–17n5
- SABER. *See* Systems Approach for Better Education Results (SABER)
- sample-based national assessments, 296n2. *See also* national assessments
- sampling error, 192, 307
- Santayana, George, 43
- SAT. *See* Scholastic Achievement Test (SAT); Scholastic Aptitude Test (SAT)
- Saudi Arabia
 examination stages in, 28–29
 selection and certification in, 23–24
 tertiary-level admission in, 159n3
- Saxony, Mann on, 59
- SBA. *See* school-based assessment (SBA)
- scale of shrinkage, effect on marking of, 166
- scheduling students with disabilities or diverse educational needs, 245–46
- scholarship schemes, 101, 116n1
- Scholastic Achievement Test (SAT), 18, 62, 158n1, 245, 246, 307
- Scholastic Aptitude Test (SAT), 18, 62, 158n1, 307
- school-based assessment (SBA)
 about, 253–54
 arguments against, 263–66
 arguments in favor of, 261–63
 as part of a public examination system, 272
 defined, 307
 dimensions of, 258–61
 external examinations compared to, 262
 problems in, 266–70
 process assessment and, 260–61
 product assessment and, 260–61
 reducing examination burden of, 254–58
 role of, 253–70
- school league tables, 116n2, 131–35, 134f, 290–91, 307
- school types, 203–4
- scoring. *See also* marking
 about, 75
 aggregation, 5, 145, 155–56
 analytic, 88–89, 167, 196, 302
 approaches to, 88–89
 examination syllabus, 76–77
 holistic, 88–89
 modifying, 286–87
 procedures for, 153–55
- Scotland
 certification in, 37n2
 examination board, 11
 feedback provided in, 124
 school-based certification in, 255
 school league tables in, 132–33
- Secondary School Certificate
 examination (SSC), 113, 308
- security, of examinations, 86, 87t, 88t, 93, 202–3, 226, 227–31, 289, 291, 294–95, 296n4
- selection function of examinations, 1, 2, 8, 22–24, 58, 293.
 See also history, of written examinations
- Senegal, 7, 116–17n5
- Senior Certificate Examination (South Africa), 206
- Senior Secondary School Certificate (Ghana), 131–32
- Serbia, fees in, 30
- SES. *See* socioeconomic status (SES)

- settings, for accommodating
 - disabilities or diverse educational needs, 243, 244
- shadow education, 2, 111–15, 283, 308. *See also* private supplementary tutoring (PST)
- Shanghai, 48–49
- short written supply-type items, 33, 82–84, 308
- Sichuan Province, place of residence as an equity issue in, 200
- Sierra Leone
 - as a conflict area, 202
 - examination stages in, 28
- Singapore
 - Cambridge Assessment International Education (CIE) and, 64
 - comparability problems in, 179
 - examination stages in, 28–29
 - examination types administered in, 7, 9
 - incentive feedback in, 133
 - Primary School Leaving Examination, 137, 296n1
 - secondary school placement in, 7
 - tutoring in, 115
- situational factors, malpractice and, 224
- skills underrepresentation, as a
 - characteristic of examinations, 149
- Slovenia
 - examination levels in, 33
 - history of public examination systems in, 58
 - publication of answer scripts in, 18–19
 - reliability of marking in, 167
- smuggling of materials, 220, 226, 229
- socialist countries (former), written examinations in, 56–58
- Society of Jesus, 50
- socioeconomic status (SES)
 - as an equity issue, 191, 196–97
 - defined, 307
 - family, 135, 136, 137, 192, 196–97, 204
 - private supplementary tuition and, 113
 - selection based on, 23
- Song dynasty, 44
- South Africa
 - administration, 90
 - cheating in, 264
 - controlling and preventing malpractice in, 229
 - disabilities or diverse educational needs in, 243
 - inequity in, 206
 - language group membership and, 206–7
 - malpractice in, 229, 264
 - number of subjects taken in, 32
 - reform, 183
 - SBA in, 255, 261, 268–69
- South Asia. *See also* specific countries
 - balance between external and internal components in, 31
 - comparability of grades in, 174
 - focus of public examinations in, 107
 - history of public examination systems in, 54
 - place of residence or location, as an equity issue, 199
- South Pacific, administration in, 29
- South Pacific Board for Educational Assessment (SPBEA), 55–56, 308
- Soviet Union (former), reform in, 281
- Spain
 - balance between external and internal components in, 31
 - lack of curriculum-based external examinations in, 3
 - private supplementary tutoring or cram schools in, 112
- SPBEA. *See* South Pacific Board for Educational Assessment (SPBEA)
- special needs candidates. *See* disabilities or diverse educational needs
- Sri Lanka
 - as a conflict area, 203
 - certification in, 19
 - controlling and preventing malpractice in, 228
 - ethnic group membership and, 204

- examination types administered
 - in, 7
 - General Certificate of Education, 20
 - grading in, 19
 - place of residence as an equity
 - issue in, 200–201
 - tutoring in, 113, 116n1
- SSC. *See* Secondary School Certificate examination (SSC)
- standard deviation, 156, 159, 166, 308
- standardized tests, 61, 66, 143–44, 146, 163, 209n3, 241
- standard setting, 62, 177, 180, 182, 183, 308
- standards
 - changes over time in, 175–78
 - comparability of grades, 174–75
 - comparability problems, 178
 - overview, 173–74
 - performance, 180–83
 - public confidence, 183–84
- State Council (China), 49
- State Education Commission (China), 49
- statistical adjustment, 133, 179, 180, 266, 269–70
- stereotyping, effect on marking of, 166
- student achievement
 - assessment conducted by schools and, 253–73
 - assessment of students with disabilities or special needs and, 239–49
 - characteristics and purposes of examinations, 17–38
 - equity issues in, 191–208
 - historical and current approaches to assessment of, 43–68
 - measurement of, 75–90
 - validity of scores and, 143–58
 - reliability of scores and, 163–69
 - standards for, 173–85
- student empowerment, difference
 - between external examinations and SBA in, 262
- students
 - differences in responses from, 32–33
 - informing about SBA, 268
 - labeling as failures, 287–88
 - likely to succeed, focus on, 108
- Sturm, Johannes, 49–50
- subject pairs analysis, 175, 308
- subject-related factors, effect of on
 - reliability, 165–67
- subjects, differences in number of, 32
- Sub-Saharan Africa
 - as a conflict area, 201
 - abolishing examinations in, 285
 - language group membership and, 205
- West African Examinations Council, 126
- Sudan, as a conflict area, 201–2
- supervision, of print shops, 88t
- surveillance cameras, 229, 231
- Sweden
 - balance between external and internal components in, 31
 - ethnic group membership and, 209n5
 - lack of curriculum based external examinations in, 3
 - reform in, 281
- Switzerland
 - administration in, 29
 - student responses in, 33
- syllabus. *See* examination syllabus
- Syrian Arab Republic
 - as a conflict area, 203
 - examination stages in, 28–29
- system assessment. *See also* national assessments
 - about, 4–6
 - differences between public examinations and, 5
- systematic error, 146, 152, 159n5, 165
- Systems Approach for Better Education Results (SABER), 307
- table of specifications, 76, 145, 157, 308
- Taiwan, China
 - comparability problems in, 178, 179
 - selection for second level in, 23
 - tutoring in, 112, 115

- Tamil, language group membership and, 205
- T'ang dynasty, 44
- Tanzania
 examination stages in, 28
 reform in, 297n6
 short-term staff exchanges and training programs in, 297n7
- Tasmania, comparability problems in, 179
- Tawjihi, 108, 308
- Taxonomy of Educational Objectives.*
 See Bloom's Taxonomy
- teachers
 accountability of. *See*
 accountability, as a function of examinations
 assessments, 63. *See also* school-based assessment (SBA)
 impact of examinations on, 7, 103, 131, 132
 preparation of, 125, 133, 148, 290, 292, 294
 professional development of, 56, 92, 262, 267, 290
 provision of materials to guide, 267
 use of public examinations by, 2–3, 99
- teaching effect, as a characteristic of good examination systems, 294
- teaching to the test, 103, 138, 151
- Telugu, language group membership and, 205
- Tennessee, computer-based large-scale assessment programs in, 95
- test preparation of students, 218–19.
 See also private supplementary tutoring (PST); teaching to the test
- test-taking strategies, as a characteristic of candidates, 153
- testing accommodations, 240–48
- Texas, computer-based large-scale assessment programs in, 95
- Thailand, controlling and preventing malpractice in, 233n1
- thanawiya amma*, 112, 308
- theft, 221–22
- threats, to validity, 149–57, 293
- time zone issues, malpractice and, 228
- timing, for accommodating students with diverse needs, 245–46
- TIMSS. *See* Trends in International Mathematics and Science Study (TIMSS)
- topics examined, focus on, 105–6
- transparency, as a characteristic of good examination systems, 293
- Trends in International Mathematics and Science Study (TIMSS), 102, 103, 195, 308
- Treviso, incentive feedback in, 131
- Trinidad and Tobago
 gender equity issues in, 195
 overseas examinations administered in, 37n3
- Tripas. *See* Cambridge Tripos
- Tunisia, abolishing basic level examinations in, 29
- Turkey
 additional tuition or cram schools in, 112
 feedback provided in, 124
- Turnitin software, 225–26
- twenty-first century skills, 284, 308
- UCLES. *See* University of Cambridge Local Examinations Syndicate (UCLES)
- Uganda
 controlling and preventing malpractice in, 227
 examination stages in, 28
 fees in, 30
 focus of public examinations in, 107
 malpractice in, 224–25
 school-based assessment (SBA) in, 272n1
 short-term staff exchanges and training programs in, 297n7
- Umalusi, 268, 309
- underrepresentation of the domain, 104–5
- unfair practice. *See* malpractice
- Unified State Examination. *See* Russian Federation

- United Kingdom. *See also* specific countries
- administration in, 29, 64
 - choice in, 35, 150
 - civil service examination in, 117n6
 - comparability of grading in, 174, 182
 - controlling and preventing malpractice in, 229
 - disabilities or diverse educational needs in, 241
 - equity issues in, 191
 - ethnic group membership and, 207, 209n5
 - examination levels in, 33
 - examination-related stress in, 108–9
 - gender equity issues in, 193
 - General Certificate of Education (GCE), 11, 33
 - Graduate Medical School Admission Test (GMSAT), 37n7
 - history of public examination systems in, 3, 9, 52–53, 131
 - international examinations and, 63–64
 - malpractice in, 223
 - moderation of SBA in, 269
 - monitoring educational standards over time in, 24–25
 - number of subjects taken in, 32
 - publication of marking and scoring keys in, 18
 - scoring, 84
 - standard setting, 182
- United States
- ACT, 18
 - admission quotas, 209n5
 - affirmative action in, 209n5
 - computer-based large-scale assessment programs in, 95
 - decentralized system in, 6
 - decline in National Assessment of Educational Progress scores in, 81
 - disabilities or diverse educational needs in, 239, 243
 - ethnic group membership and, 209n5
 - history of public examination systems in, 3, 4, 59–63
 - international examinations and, 63, 64
 - lack of curriculum-based external examinations in, 3
 - malpractice in, 209n2, 223
 - National Assessment of Educational Progress, 5, 81, 176
 - rules for high school graduation in, 255
 - Scholastic Achievement Test (SAT), 18, 158n1, 308
 - Scholastic Aptitude Test (SAT), 18, 158n1, 308
 - standards-based reform efforts in, 110, 173
 - test preparation in, 218–19
 - timing accommodations for disabilities or diverse educational needs in, 245
- University of Bologna, 51
- University of Cambridge, 104, 193.
See also history, of written examinations
- University of Cambridge Local Examinations Syndicate (UCLES), 37n3, 64, 254, 308
- University of London, 37n3, 53
- University of Oxford Delegacy of Local Examinations, 52, 254
- University of Paris, 49
- Urdu, language group membership and, 205
- U.S. Americans with Disabilities Act, 241
- USE. *See* Russian Federation
- validity
- about, 143–44
 - aspects of, 146–48
 - choice and, 34–35
 - construct, 147–48, 303
 - content, 145–46
 - curriculum-based assessment and, 8–9
 - ecological, 163, 304
 - improving, 125, 254, 283, 284
 - malpractice and, 217
 - predictive, 148
 - SBA and. *See* school-based assessment (SBA)
 - threats to, 149–57

- value-added approaches, 135–37, 139n7
- Vietnam
 - language group membership and, 206
 - selection in, 24
- visually impaired, accommodations for the, 239, 245, 246
- WAEC. *See* West African Examinations Council (WAEC)
- Wales
 - examination boards in, 11, 37n4
 - examination levels in, 33
 - performance assessment in, 83
 - publication of results in, 116n2
 - retrenchment in, 84
 - school league tables in, 132–33, 134f
 - selection in, 24
 - value-added modeling in, 137
- washback effect, 12, 262, 309
- weightings, 19, 32, 155, 156, 159n6, 261
- Wendi, Emperor, 44
- West Africa
 - administration in, 29
 - fees in, 37n6
 - grades, 181
- West African Examinations Council (WAEC), 55–56, 126, 309
- Western Zhou dynasty, 44
- World Conference on Education for All, 7, 147
- World Education Forum, 7, 8
- written examinations
 - history of, 43–68
 - in China, 44–49
 - in developing economies, 54–56
 - in Europe, 49–54, 67n5
 - in former socialist countries, 56–58
 - in Latin America, 58–59
 - in United States, 59–63
 - international, 63–64
- Yemen. *See* Republic of Yemen
- Zambia, comparability of grading in, 181
- Zambian Junior Secondary School Leaving Certificate, 106
- z-scores, 178, 309

ECO-AUDIT

Environmental Benefits Statement

The World Bank Group is committed to reducing its environmental footprint. In support of this commitment, we leverage electronic publishing options and print-on-demand technology, which is located in regional hubs worldwide. Together, these initiatives enable print runs to be lowered and shipping distances decreased, resulting in reduced paper consumption, chemical use, greenhouse gas emissions, and waste.

We follow the recommended standards for paper use set by the Green Press Initiative. The majority of our books are printed on Forest Stewardship Council (FSC)–certified paper, with nearly all containing 50–100 percent recycled content. The recycled fiber in our book paper is either unbleached or bleached using totally chlorine-free (TCF), processed chlorine-free (PCF), or enhanced elemental chlorine-free (EECF) processes.

More information about the Bank’s environmental philosophy can be found at <http://www.worldbank.org/corporateresponsibility>.



High-stakes public examinations exert a dominant influence in most education systems. They affect both teacher and student behavior, especially at the middle and upper levels of secondary education. The content of past examinations tends to dictate what is taught and how it is taught and, more important, what is learned and how it is learned. By changing aspects of these examinations, especially their content and format, education systems can have a strong positive impact on teacher behavior and student learning, help raise student achievement levels, and better prepare students for tertiary-level education and for employment. Examination agencies, many of which have followed the same procedures over decades, can learn from the successes and failures of other systems.

This book addresses current issues related to the development, administration, scoring, and usage of these high-stakes public examinations, identifying key issues and problems related to examinations in many emerging market economies as well as in advanced economies. The book's primary audience consists of public examination officials on national, regional, and state examination boards, but the book should also be of interest to senior education policy makers concerned with certification and learning achievement standards, to academics and researchers interested in educational assessment, to governmental and education agencies responsible for student selection, and to professionals at development organizations.

"This extremely well-written and comprehensive book offers a timely review of the diversity of public examination practices worldwide; of the tensions between examinations and learning; and of the technical expertise involved in the creation of valid, reliable, and fair assessments. It reminds us that as "the diploma disease" takes hold with an ever-greater intensity at every stage of education worldwide, and the commercial business of testing flourishes, those concerned with educational quality and meaningful learning must be on guard to prevent the assessment tail wagging the educational dog."

Angela W. Little, *Professor Emerita, Institute of Education, University College London*

"This book is very well structured and written and draws on the authors' remarkable global knowledge across countries and histories. It will be a great asset both to administrators responsible for examinations and to academics and other professionals who seek to understand the nature and impact of examinations of different types and in different settings."

Mark Bray, *UNESCO Chair Professor of Comparative Education, University of Hong Kong; and former Director, UNESCO International Institute for Educational Planning*

"I am sure that *Public Examinations Examined*, which thoroughly analyzes the practice of public examinations in different countries and makes profound and well-grounded conclusions, will arouse very great interest and will serve to further improve public examinations."

Victor Bolotov, *Distinguished Professor, Higher School of Economics, National Research University, Moscow; member, Russian Academy of Education; and former Deputy Minister of Education, Russian Federation*

ISBN 978-1-4648-1418-1



SKU 211418

