

GLOBAL PERSPECTIVES ON EDUCATIONAL TESTING

Examining Fairness, High-Stakes
and Policy Reform

By Keena Arbuthnot

ADVANCES IN EDUCATION
IN DIVERSE COMMUNITIES:
RESEARCH, POLICY AND PRAXIS

VOLUME 13

**GLOBAL PERSPECTIVES ON
EDUCATIONAL TESTING:
EXAMINING FAIRNESS,
HIGH-STAKES AND POLICY
REFORM**

ADVANCES IN EDUCATION IN DIVERSE COMMUNITIES: RESEARCH, POLICY AND PRAXIS

Series Editor: Carol Camp Yeakey

Recent Volumes:

- Volume 3: Teacher Unions and Education Policy: Retrenchment or Reform? – Edited by Ronald D. Henderson, Wayne J. Urban and Paul Wolman
- Volume 4: Suffer the Little Children: National and International Dimensions of Child Poverty and Public Policy – Edited by Carol Camp Yeakey, Jeanita W. Richardson and Judith Brooks Buck
- Volume 5: Higher Education in a Global Society: Achieving Diversity, Equity and Excellence – Edited by Walter R. Allen, Marguerite Bonous-Hammarth and Robert Teranishi
- Volume 6: Power, Voice and the Public Good: Schooling and Education in Global Societies – Edited by Rodney K. Hopson, Carol Camp Yeakey and Francis Musa Boakari
- Volume 7: As the World Turns: Implications of Global Shifts in Higher Education for Theory, Research and Practice – Edited by Walter R. Allen, Robert T. Teranishi and Marguerite Bonous-Hammarth
- Volume 8: Living on the Boundaries: Urban Marginality in National and International Contexts – Edited by Carol Camp Yeakey
- Volume 9: Health Disparities Among Under-Served Populations: Implications for Research, Policy and Praxis – Edited by Sheri R. Notaro
- Volume 10: The Obama Administration and Educational Reform – Edited by Eboni M. Zamani-Gallaher
- Volume 11: Mitigating Inequality: Higher Education Research, Policy, and Practice in an Era of Massification and Stratification – Edited by Robert T. Teranishi, Loni Bordoloi Pazich, Marcelo Knobel and Walter R. Allen
- Volume 12: The Power of Resistance: Culture, Ideology and Social Reproduction in Global Contexts – Edited by Rowhea M. Elmesky, Carol Camp Yeakey and Olivia Marcucci

ADVANCES IN EDUCATION IN DIVERSE
COMMUNITIES: RESEARCH, POLICY AND
PRAXIS VOLUME 13

**GLOBAL PERSPECTIVES
ON EDUCATIONAL
TESTING: EXAMINING
FAIRNESS, HIGH-STAKES
AND POLICY REFORM**

BY

KEENA ARBUTHNOT
Louisiana State University, USA



United Kingdom – North America – Japan
India – Malaysia – China

Emerald Publishing Limited
Howard House, Wagon Lane, Bingley BD16 1WA, UK

First edition 2017

Copyright © 2017 Emerald Publishing Limited

Reprints and permissions service

Contact: permissions@emeraldinsight.com

No part of this book may be reproduced, stored in a retrieval system, transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without either the prior written permission of the publisher or a licence permitting restricted copying issued in the UK by The Copyright Licensing Agency and in the USA by The Copyright Clearance Center. Any opinions expressed in the chapters are those of the authors. Whilst Emerald makes every effort to ensure the quality and accuracy of its content, Emerald makes no representation implied or otherwise, as to the chapters' suitability and application and disclaims any warranties, express or implied, to their use.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN: 978-1-78635-434-1 (Print)

ISBN: 978-1-78635-433-4 (Online)

ISBN: 978-1-78714-896-3 (Epub)

ISSN: 1479-358X (Series)



ISOQAR certified
Management System,
awarded to Emerald
for adherence to
Environmental
standard
ISO 14001:2004.

Certificate Number 1985
ISO 14001



INVESTOR IN PEOPLE

This book is dedicated to William, Alfreda & Stephanie

This page intentionally left blank

CONTENTS

LIST OF TABLES	<i>ix</i>
ACKNOWLEDGMENTS	<i>xiii</i>
INTRODUCTION	<i>xv</i>
CHAPTER 1: INTERNATIONAL ASSESSMENTS AND FAIRNESS ISSUES	<i>1</i>
CHAPTER 2: THE ARBUTHNOT ASSESSMENT FAIRNESS FRAMEWORK	<i>23</i>
CHAPTER 3: TAAF FRAMEWORK PHASE I: COUNTRY SELECTION	<i>37</i>
CHAPTER 4: TAAF FRAMEWORK PHASE II: TEST AND SUBTEST PERFORMANCE	<i>79</i>
CHAPTER 5: TAAF FRAMEWORK PHASE III: ITEM-LEVEL PERFORMANCE	<i>127</i>
CHAPTER 6: TAAF FRAMEWORK PHASE IV: CULTURAL IMPACT	<i>147</i>
CHAPTER 7: RETHINKING FAIRNESS AND INTERNATIONAL ASSESSMENTS	<i>159</i>

REFERENCES	<i>177</i>
ABOUT THE AUTHOR	<i>183</i>
INDEX	<i>185</i>

LIST OF TABLES

Table 1	Comparison of Major International Assessment Programs	4
Table 2	Percentage of Students with Various Home Resources.....	47
Table 3	Percentage of Students Whose Parents Have a College Degree or Work Professionally	48
Table 4	Percentage of Students Who Spoke the Language of the Test	49
Table 5	Percentage of Students Expecting to Earn a College Degree or Better	50
Table 6	Percentage of Students Whose Teachers Reported Instruction was Limited by Lack of Nutrition	51
Table 7	Percentage of Teachers with a Bachelor’s or Postgraduate Degree.....	52
Table 8	Teachers’ Years of Experience	52
Table 9	Percentage of Students Whose Teachers Feel “Very Well” Prepared to Teach Math Topics.....	53
Table 10	Percentage of Students Whose Teachers Feel “Very Confident” Teaching Mathematics	53
Table 11	Percentage of Students Whose Teachers Feel “Very Confident” in Teaching Activities.....	54
Table 12	Percentage of Students Whose Teachers Feel “Satisfied” with Their Career.....	55
Table 13	Percentage of Students Whose Teachers Report Instruction is Limited by Disruptive Students “a Lot”	56
Table 14	Percentage of Instructional Time Spent on Mathematics	56
Table 15	Percentage of Students Whose Teachers Used Textbooks or Computer Software for Instruction	57
Table 16	Percentage of Students “Memorizing Rules, Procedures, and Facts”, “Explaining Their Answers” or “Applying Facts, Concepts and Procedures” on Every or Almost Every Lesson...	58
Table 17	Percentage of Students Who Liked Learning, Valued, Felt Confident and Were Engaged in Mathematics.....	59

Table 18	Percentage of 8th Grade Students Expecting to Earn at Least a College Degree	61
Table 19	Percentage of the Frequency at which 8th Grade Students are Given Tests or Exams	61
Table 20	Percentage of 8th Grade Students Whose Teachers “Always or Almost Always” Require Students to Solve Certain Types of Questions	62
Table 21	Comparison of Countries’ Responses to Questionnaire Factors	63
Table 22	Definitions of Content Domains on the TIMSS Mathematics Test	81
Table 23	Definitions of Cognitive Domains on the TIMSS Mathematics Test	81
Table 24	TIMSS Math Achievement Scores for Chinese Taipei	85
Table 25	TIMSS Math Achievement Scores for Finland.....	85
Table 26	TIMSS Math Achievement Scores for the United States	86
Table 27	TIMSS Math Achievement Scores for Qatar	86
Table 28	TIMSS 2011 Math Test Average Scale Scores and Rank	87
Table 29	Country Comparison of Math Achievement by Content Domains.....	88
Table 30	Country Comparison of Math Achievement by Cognitive Domains.....	88
Table 31	Average Point Difference and Range of Point Differences for Each Country Comparison for 4th Grade.....	90
Table 32	Percentage of Items with Large Item-Level Differences for Each Country Comparison for 4th Grade.....	91
Table 33	Average Point Difference and Range of Point Differences for Each Country Comparison for 8th Grade.....	98
Table 34	Percentage of Items with Large Item-Level Differences for Each Country Comparison for 8th Grade.....	98
Table 35	Average Number of Item Omissions by Country.....	110
Table 36	Average Number of Omissions by Country for 4th Grade	112
Table 37	Average Number of Omissions by Country for 8th Grade	112
Table 38	Percentage of Large Omissions by Country for 4th Grade	114
Table 39	Percentage of Large Omissions by Country for 8th Grade	114
Table 40	Average Omissions by Item Type and Rate of Omission by Country for 4th Grade	115
Table 41	Average Omissions by Item Type and Rate of Omission by Country for 8th Grade	115
Table 42	Percentage of Items Not Reached by Country	116

Table 43	Number of TIMSS Mathematics Topic Areas Intended to be Taught by the End of 4th Grade.....	120
Table 44	Percentage of Students Taught the TIMSS Mathematics Topic Areas in 4th Grade	120
Table 45	Number of TIMSS Mathematics Topic Areas Intended to be Taught by the End of 8th Grade.....	121
Table 46	Percentage of Students Taught the TIMSS Mathematics Topic Areas in 8th Grade	122
Table 47	Sources of Item Bias Results	138
Table 48	Description of Items that Favor One Cultural Context Country over Another	154

This page intentionally left blank

ACKNOWLEDGMENTS

I have many mentors and colleagues whom I must thank for their selflessness in helping me to achieve my academic and career goals, including, but not limited to, Drs. Edmund W. Gordon, M. Christopher Brown II, William Trent, Sara Lawrence-Lightfoot, Eugene Kennedy, Stafford Hood, Carol Camp Yeakey, Katherine E. Ryan, and Bridget Terry Long. Each of you have been instrumental in shaping my career, and I am truly grateful for your leadership and vision to light the way for those of us who come after you. Additionally, I must acknowledge those whom I consider my contemporaries and who work alongside me to leave our mark in our respective fields. We support and inspire one another and provide much needed shoulders on which to lean in times of trial, and are true encouragers in times of triumph. These individuals include, but are not limited to, Dawn Williams Salter, Maurice Samuels, Darrell Ray, and Sassy Wheeler. Last, I must acknowledge those scholars whom I have had the distinct honor of mentoring and advising during their academic pursuits, including but not limited to, Tireka Cobb, Jared Avery, Adam Elder, Erin Scott-Stewart, and Guadalupe Lamadrid.

Personally, I want to thank my parents William and Alfreda Arbuthnot and my sister Stephanie Arbuthnot whose love and support have always been the wind beneath my wings. I cannot begin to tell each of you how much I appreciate the unconditional love and unwavering support that you have shown me throughout the years. You continue to encourage me to meet my highest potential and to always put God first. As a child growing up, each one of you taught me in your own special way how to face adversities with dignity and fortitude. Those lessons taught me how to survive in the face of opposition and how to thrive in the midst of it. In addition to my immediate family, I have been blessed with an incredibly supportive extended family, including but not limited to, Barbara Whitener Gardner, Kelly Gardner, Kimberly Gardner, Michael Bates, the Murry family and other aunts, uncles, and cousins. I also want to thank my friends, many of whom have become family, for always making the choice to love and support me regardless of the circumstances. There are too many to name but they include Jas and Samaah Sullivan, Vashti Person, Franklin Mosley III, Kaisha Mozee, Valerie Byers,

Tyrone Perkins, LaShonda Harvey, Eric Cook, Bridget Perkins, Kimberly Cole, and the Seven Sisters. Last, to all my godchildren, mentees and former students, including but not limited to, Tierra Webb, Aja and Mariah Brooks, Dallas Hawkins, Jennifer Cook, Aisha Camphor and Marcus Jones, you are all an inspiration to me.

I want to also thank all of the friends and colleagues I have made across the world. The opportunity to meet wonderful people from various places around the globe has reshaped my research focus not only on issues plaguing testing and measurement in the United States, but also on expanding this line of research to issues pertaining to education worldwide. Improving education and understanding the way in which all students learn, regardless of race, ethnicity, or country of origin, have become central to my research agenda. My experiences abroad have shown me that all parents, educators, and policy-makers intuitively want the same thing: *to make education accessible, fair and engaging to all children* regardless of who they are or from where they come.

INTRODUCTION

Much of my career has focused on investigating group differences in standardized test performance and test fairness issues in the United States. My interest in this area was sparked by my experiences as a high school mathematics teacher. As a teacher, I realized the impact that high-stakes tests had on students, particularly Black students, who are considered to be the minority in the United States. Research has shown that on most standardized tests, White students outperformed their Black counterparts. I wanted to determine whether these high-stakes tests were fair to all test taker groups, and, consequently, spent years as a graduate student and professor examining fairness issues and investigating test- and item-level performance patterns. My research investigated Black and White test takers and highlighted the differences in the test-taking experiences between the two groups (Arbuthnot, 2009, 2011a, Arbuthnot & Ryan, 2005). The results undeniably showed that cultural differences between Blacks and Whites had a significant impact on their standardized test experience and performance (Arbuthnot, 2011a,-, 2015a-, 2015b; Arbuthnot & Lyons-Thomas, 2016). Moreover, the research identified different mathematics subtests and types of items that were differentially harder for one group in comparison to the other. Additionally, the research focused on how students from various groups process test items differently while taking a standardized test. The research provided a new way to understand and conceptualize the way that different race/ethnic groups, with contrasting cultural backgrounds, experienced and performed on tests. My book *Filling in the Blanks: Understanding the Black White Achievement Gap* (Arbuthnot, 2011) provided comprehensive details from over a decade of research regarding the variations between Black and White test-takers' experiences. One line of the book's research examined the similarities in the test-taking experience and performance of Black students and female students on standardized mathematics tests. The results indicated that Black and female students performed similarly on these examinations. To explain this finding, I examined the commonalities between Black students and female students in the United States. Consequently, the cultural similarities between the two groups helped to explain why the test-taking patterns of Black and female students were comparable.

Simultaneously, I dedicated my research to addressing issues related to test fairness as well. My writing focused on examining the high-stakes testing systems and providing empirical research to challenge test fairness issues. This research highlighted the roles and responsibilities of test developers in ensuring that tests were fair to all test takers, as well as challenging test users and consumers to critically examine the way in which they interpreted test results (Arbuthnot, 2011a, 2012a, 2015a, 2015b; Arbuthnot & Lyons-Thomas, 2016).

I continued my research and worked on issues related to domestic test fairness; concurrently, I was presented with opportunities to participate in consulting and grant opportunities abroad. Most of my opportunities abroad involved countries in the Middle East, mainly Qatar and the United Arab Emirates. From these experiences, I developed a better understanding of the educational challenges that Middle Eastern countries faced. I realized that some of the issues that students faced in that region were similar to the difficulties that I recognized in my research with Black and female students in the United States. With my background in standardized testing, I turned to international assessments to investigate the similarities between minority and female students in the United States and students from the Middle East, expanding my line of research on testing and fairness to a global scale.

This book highlights the basis and justification for the research and provides a detailed exploration of how to examine fairness issues on international assessments. It is my hope that audiences around the world will utilize this book in the quest for understanding and conceptualizing variations in the way in which test takers from different countries and cultures learn and perform on tests.

The purpose of this book is to investigate fairness issues on international assessments. The text begins with an overview of the current state of international assessments and reveals the ways in which many countries have utilized results from international assessment initiatives to inform educational policy and practice at the national level. The book then describes the various international assessment programs that have been implemented over the last several decades. The text then focuses on the Trends in International Mathematics and Science Study (TIMSS) assessment, one of the most popular and longstanding international assessment programs. The book utilizes TIMSS assessment data that includes the 4th and 8th grade mathematics test in conjunction with information obtained from a variety of stakeholder questionnaires (i.e., students, teachers, and administrators) from each of the participating countries. Additionally, the use of details concerning fairness issues is included as well as how research conducted on fairness issues in the

United States provides a framework for examining issues of fairness at the global level.

In order to investigate fairness issues on the global scale, the author introduces The Arbutnot Assessment Fairness (TAAF) Framework as a means to systematically examine test- and item-level performance patterns and fairness issues. The TAAF Framework has four phases: (a) country selection, (b) test and subtest performance and test-taking patterns, (c) item-level differences and patterns, and (d) cultural impact. Completing these four phases of the TAAF Framework provides a clearer understanding and interpretation of test- and item-level differences and provides an in-depth examination of fairness issues on international assessment. The author provides a detailed example of how to use the TAAF Framework to inform multiple stakeholder groups about differences in performance patterns among countries and test fairness issues. The countries of Chinese Taipei, Finland, the United States, and Qatar were chosen for the research based on three criteria: cultural context, performance patterns, and educational reform policies. The book provides a critical examination of the educational practices, assessments, accountability measures, and cultural norms for each of these countries.

The book utilizes the empirical data provided by the mathematical portion of the TIMSS international assessments to demonstrate and elucidate how to analyze the international assessment data and use multiple data sources to examine issues of fairness on international assessments. This book ends by challenging readers to deliberate more thoughtfully and to exercise caution with respect to the ways in which test- and item-level performance data is interpreted on international assessments.

To conclude, the author encourages readers to be mindful when interpreting test performance and insists that the various stakeholder groups, the test developers, educational theorists, policymakers, and practitioners, better understand the differences in performance patterns between countries and the issues surrounding test fairness on international assessments.