THINKING VALUE-
MODELS IN EDUCATION

l Perspectives on Tests and Assessment-Based Accountability

Audrey Amrein-Beardsley

Foreword by: Diane Ravitch

ROUTLEDGE

RETHINKING VALUE-ADDED MODELS IN EDUCATION

Since passage of the No Child Left Behind Act in 2001, academic researchers, econometricians, and statisticians have been exploring various analytical methods of documenting students' academic progress over time. Known as value-added models (VAMs), these methods are meant to measure the value a teacher or school adds to student learning from one year to the next. To date, however, there is very little evidence to support the trustworthiness of these models. What is becoming increasingly evident, yet often ignored mainly by policymakers, is that VAMs are (1) unreliable, (2) invalid, (3) nontransparent, (4) unfair, (5) fraught with measurement errors, and (6) being inappropriately used to make consequential decisions regarding such things as teacher pay, retention, and termination. Unfortunately, their unintended consequences are not fully recognized at this point either. Given such, the timeliness of this well-researched and thoughtful book cannot be overstated. This book sheds important light on the debate surrounding VAMs and thereby offers states and practitioners a highly important resource from which they can move forward in more research-based ways.

Audrey Amrein-Beardsley, Ph.D., is an Associate Professor in Mary Lou Fulton Teachers College at Arizona State University. She is one of the top Edu-Scholars in the nation, honored for being an academic who is contributing substantially to public debates about the nation's educational system. She is also creator and host of the blog: VAMboozled! (vamboozled.com).

RETHINKING VALUE-ADDED MODELS IN EDUCATION

Critical Perspectives on Tests and Assessment-Based Accountability

Audrey Amrein-Beardsley

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

First published 2014
by Routledge
711 Third Avenue, New York, NY 10017

and by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

Library of Congress Cataloging in Publication Data
Amrein-Beardsley, Audrey, author.
Rethinking value-added models in education : critical perspectives on tests and assessment-based accountability / Audrey Amrein-Beardsley. – First published 2014.
p. cm.
Includes bibliographical references and index.
1. Educational accountability–United States. 2. Educational indicators–United States. 3. Educational tests and measurements–United States. 4. Teachers–Rating of–United States–Mathematical models. I. Title.
LB2806.22.A46 2014
379.1'58–dc23 2013037411

ISBN: 978-0-415-82011-0 (hbk)
ISBN: 978-0-415-82012-7 (pbk)
ISBN: 978-0-203-40990-9 (ebk)

Typeset in Bembo
by Wearset Ltd, Boldon, Tyne and Wear

Printed and bound in the United States of America by Publishers Graphics, LLC on sustainably sourced paper.

This book is dedicated to the orphaned children at the Assisting Cambodian Orphans and the Disabled Organization (ACODO) Orphanage of the Kingdom of Cambodia. All of my personal royalties from this book will go to this orphanage.

My husband, children, and I believe that the royalties from this book will add more "value" to these children's lives at ACODO than they ever would ours.

For more information about the ACODO Orphanage, or to sponsor a child, donate money, or even donate your time should you visit Siem Reap, Cambodia, please visit: www.acodo.org

Figure 0.1 My daughter Blythe, son Aidan, and some of their ACODO friends in Siem Reap, Cambodia.

CONTENTS

FIGURES

TABLES

FOREWORD

Diane Ravitch

Audrey Amrein-Beardsley has written a timely, useful, and important book. She describes and analyzes in great detail the imposition of test-based evaluation of teachers, the theory behind it, the real-life consequences of this strategy, and its fundamental flaws.

This is not just another academic book by a scholar. It is a book written by a highly qualified scholar who worked as a math teacher for several years and understands the connections between theory and practice. More than that, it is a book informed by research and fieldwork. More than that, it is a book in which the author's outrage simmers slightly beneath the surface and occasionally comes into full view, as when she equates the purposeful demoralization of classroom teachers as "intellectual genocide." These are strong words, but they are not used lightly.

Amrein-Beardsley documents the insanity of test-based evaluation, or value-added modeling (VAM). It has been in use in Tennessee for more than 20 years, and it has now been imposed on the nation's public schools by the Obama administration's *Race to the Top* program. Amrein-Beardsley demonstrates that research does not support the use of VAM. The public and educators, she says at one point, have been VAMboozled by the advocates of this method of evaluating teacher quality.

What is known about VAM, she writes, is that it lacks validity and reliability. Some educational researchers support its use, but about 95% do not. Their voices, like the voices of teachers, are ignored by policymakers who think they have found an easy way to judge teacher quality, to hand out bonuses to those with high rankings, and to fire those who have low rankings.

As Amrein-Beardsley shows, the rankings are inaccurate, unstable, and unfair. Those who teach special education or gifted students or English language learners are at high risk of "failing." Only approximately 30% of the teachers are subject to the rankings based on VAM, and many districts give those teachers ratings based on the performance of the whole school, which is unfair and irrational.

No one understands the rankings, least of all the teachers. They get a high score and a bonus one year, a low score and a warning the next year. From their perspective, they have changed nothing in their teaching, and they have no idea why they were better or worse.

As Amrein-Beardsley shows in her fieldwork, Houston relied on VAM and fired teachers who had been recognized as "Teacher of the Month" and "Teacher of the Year" in their school.

If anyone went in search of a way to demoralize teachers and principals and to harm American public education, they could not have found a more valuable tool than VAM. Where sensible nations seek ways to strengthen collaboration among teachers and to improve their practice, we have adopted a strategy that encourages teaching to the test, gaming the system, and setting teacher against teacher.

Audrey Amrein-Beardsley's book could not have arrived at a more propitious moment. It is time for researchers, educators, and the public to recognize the damage now inflicted by federal policy on our nation's teachers and public schools. It is time to end it.

Diane Ravitch

PREFACE

The important thing is not to stop questioning. Curiosity has its own reason for existing.
Albert Einstein

Within the pages of this book is a series of research-based accounts about how the American public education system has, for the past 30 years, been derailed by the same socially engineered theory of educational reform and change; a Measure and Punish (M&P) Theory of Change that has been successively reinvented over the past three decades as politicians and policymakers have tweaked this or that to reform America's purportedly failing public schools. The M&P Theory of Change suggests that by holding districts, schools, teachers, and students accountable for meeting higher standards, as measured by student performance on high-stakes tests, administrators will supervise America's public schools better, teachers will teach better, and as a result students will learn more, particularly in America's lowest performing schools. In this interpretation of utopia, students' test scores will increase, America's global prominence will be reclaimed, and the nation will achieve the rehabilitated educational system and global superiority the public, now, so deeply desires.

The M&P Theory of Change is based on a paucity of empirical research evidence, however. Very few scientific studies have evidenced that this theory of change works. Rather, countless scientific studies have evidenced that the numerous educational policies based on this theory of change have caused unintended consequences. In addition, numerous academic researchers suggest that the negative side effects altogether outweigh the few positive benefits that may have been realized, if at all, post policy.

As well, teachers and principals working in practice under this M&P Theory of Change have much to say about the theory's critical shortcomings, but their voices have often been marginalized, dismissed, or even rejected, particularly when they have been positioned as unprofessional or even recalcitrant when charged with protecting their own self interests or the status quo. While educators working day-to-day in America's public schools would certainly be the best to speak truth to power, particularly about such policies that in theory make sense but in practice make nonsense, their voices too often fall on deaf ears.

Correspondingly, policy attention has now turned to a new set of improved educational measurement systems that can, again in line with the same M&P Theory of Change, improve upon the precision with which we might evaluate teacher (and school/district) effectiveness and more accurately hold educators accountable for doing their jobs and meeting even higher standards (e.g., the new Common Core State Standards currently being implemented across the nation). The M&P Theory of Change continues to hold, but atypically this time as value-added models (VAMs) are simultaneously being positioned at the center of the operations more "scientifically" supporting this same theory of change. Put differently, VAMs are now to help us better navigate the familiar road to utopia, by more appropriately and accurately helping us hold educators responsible for supposedly veering too far off of the same route to reform.

VAMs by definition are designed to isolate and measure teachers' (or schools'/districts') contributions to student learning and achievement on large-scale standardized achievement tests as groups of students move through school from one grade level to the next. VAM statisticians attempt to measure value-added by mathematically calculating the "value" a teacher (or school/district) "adds" to (or detracts from) student achievement scores from the point at which students enter a classroom (or a school/district) to the point they leave. In more precise terms, VAM statisticians attempt to calculate value added by computing the difference between students' composite test scores at these two points in time, after which they compare the added/detracted value (or growth/decline) coefficients to what they predicted beforehand and to the coefficients of other "similar" teachers (or schools/districts) who posted "similar" value-added estimates at the same time. VAM statisticians then position teachers (or schools/districts) accordingly, and typically hierarchically along a categorical yet arbitrary continuum, assigning teachers (or schools/districts) high to low value-added categorizations with negative and positive differences yielding negative and positive value-added classifications respectively. From here, high-stakes decisions (e.g., merit pay, teacher tenure, teacher terminations) can more appropriately and accurately be made, again, so it is assumed.

Yet, here, there is also a paucity of research illustrating that VAMs work in the ways theorized. Rather, a plethora of studies exist in which researchers have evidenced that VAMs cause unintended, perverse consequences instead. Again, evidence indicates that this same M&P Theory of Change, even with its advanced VAM metrics, is still grossly flawed, misguided, and potentially harmful.

As social engineering theory would have it, the new and improved VAM metric systems are continuously bamboozling, or in this case VAMboozling us as a collective society keenly interested in the fate of the American public school system. At the same time, those marketing and selling their VAMs, often for profit, are raking in millions in federal, state, and district monies. That is, not only are we being VAMboozled, we are paying to be VAMboozled as our political leaders continue to use taxpayer revenues to further advance a tested, tried, and untrue, and highly deceptive socially engineered theory of change.

Accordingly, within this book I present everything I think everybody who cares about America's public education system, and in particular this educational accountability trend, needs to know about VAMs. In Part I, I introduce a historical frame and some foundational knowledge, as well as a poignant case in which this VAM-based theory was implemented in practice. More specifically, in Chapter 1, I provide the background to the aforementioned socially engineered theory of change and how this history has brought us to a newly found

love affair with VAMs. In Chapter 2, I deliver a series of research-based tales about how VAMs have been adopted and used in practice. Also included is the series of tales about how this theory has thus far fallen short in practice, namely in the Houston Independent School District (HISD). In Chapter 3, I present a national overview capturing what states are doing in terms of adopting VAM-based policies and in terms of the uses and consequences being adopted or mandated as per such policies.

In Part II of the book, I describe a series of "heroic" assumptions (Rubin, Stuart, & Zanutto, 2004) that I believe all involved must understand, and with which all must agree should they buy into VAMs and the VAM-based estimates from which VAM-based inferences are being drawn. Specifically, in Chapter 4, I detail the assumptions that are typically used to rationalize and justify the adoption of VAMs, and in Chapter 5, I dig deeper into the particular test-based, statistical, and methodological assumptions that go along with VAM adoption and use. It should be noted here that these assumptions are widely used not only to promote VAM adoption and use, but also to sell for-profits' and sometimes non-profits' VAM-based systems to states, districts, and local schools. Whether the assumptions discussed are indeed research-based facts and truths is another topic of interest that I explore within both of these chapters as well.

In Part III of this book, I examine the non-traditional issues with the traditional notions of research in education and educational measurement. The notions of interest here include reliability, validity, bias, fairness, and use. More explicitly, in Chapter 6, I present what are, in my opinion, the most major issues that are still impeding the practicality and usability of VAMs. These issues are being caused, primarily, by the instabilities being observed across models and over time when examining VAM-based estimate consistency (i.e., reliability). In addition, I examine all types of validity evidence (e.g., content-, criterion-, construct-, and consequential-related evidence of validity) that also continue to be the source of concern, specifically in terms of how accurate and truthful the inferences being made using VAM estimates might in reality be. In Chapter 7, I present how issues with bias and the non-random assignment of students to classrooms (and teachers to classrooms) are both compounding and complicating these methodological and pragmatic issues further.

In Part IV and in the final chapter of this book, Chapter 8, I explore a series of conventional alternatives and solutions – conventional in the sense that these alternatives and solutions are often and customarily offered by other researchers conducting and disseminating research in this area. Counter-intuitively, I end with an even-more conventional solution that I believe, with some good research evidence in support, might be the best solution available for evaluating and holding teachers accountable for their effectiveness as the professionals they likely are, and we as taxpayers pay them to be.

In short, within this book is everything I have ever read about VAMs in education from both academic and popular outlets, provided here in context and depth for reader consumption.

Fittingly, because this is probably the hottest and most controversial topic surrounding current debates about educational reform, there is a lot to be said. Conducting a Google search for "value-added models," for example, yields approximately 40 million results, while searching for "value-added models in education" yields approximately 16 million results. While I have certainly not read everything available on the topic, as this would be an impossible task, I have read probably more than most educational researchers conducting research in this area (see, for example, the over 600 references in the reference list). In addition, not

one day has gone by while I was writing this book during which two or three new research manuscripts, newspaper articles, blog entries, webinars, or other informational pieces about VAMs have been made public online (or on paper), on which I have also kept up to keep this book as current as humanly possible. While there will certainly be pieces of information that I have missed by the time this book is published, keeping up with the literature in this area has been a continual task, as has been condensing everything "of value" in terms of the literature on this topic.

Ultimately, these efforts should help allow consumers, including readers from the schoolhouse to the White House, become better educated and empowered to make more informed decisions about the VAMs being adopted with rapidity across the country. It is imperative that the major and minor issues that pertain to VAMs, and the issues that continue to inhibit VAM practicality be understood before more educators and policymakers blindly buy into VAM-based systems. Without accurate, research-based, and unbiased information about both the intended and the unintended consequences of the VAMs ruling supreme over all other contemporary measurement and test-based accountability systems, Americans will continue to be VAMboozled by such VAM-based policies. That said, I hope to afford readers a user-friendly academic guide to help them navigate and better understand VAMs.

However, I do reserve the right to speak truth to power as I aptly, genuinely, justly, and *not* dispassionately see it. For in writing this book, I embrace my professional obligation and moral imperative as one of a handful of academic researchers analyzing the intended, and one of very few academic researchers analyzing the unintended, effects of VAMs. I also embrace this imperative as a former public school teacher, noting that not many people (if any) who are advocating for, advancing, or even working on VAM methods and the statistics used to capture educator/educational effectiveness have ever set foot in a classroom, much less attempted to understand the realities and complexities that come along with teaching, especially in some of the nation's highest needs schools. Just because many of the people advancing and promoting VAM-based policies for America's public schools have likely been students themselves in America's private or perhaps public schools, this does not an educational expert make! Not to mention a successful educational researcher, and certainly not a successful educational reformer.

I am proud and honored to say that I am a former mathematics teacher with seven years of grade 7–12 teaching experience, teaching mathematics in both public middle and high schools to students who were (and still are) highly at-risk. I also have 15 years of educational research experience, through which, given my background in mathematics, I also bring what I believe is an atypical scientific perspective, particularly in terms of using statistics and educational measurement techniques to understand and explain phenomena, especially in the social sciences of which education is a part. I believe that numbers are often socially constructed, but they are not nearly often enough understood or interpreted as such. This perspective also resonates throughout this book, as it should, because this is so often misunderstood. As Albert Einstein so brilliantly put it, "As far as the laws of mathematics refer to reality, they are not certain, and as far as they are certain, they do not refer to reality." With this assertion, I certainly align.

In the case of VAMs, it seems that our "excessive attention to quantitative data *impedes* – rather than aids – in-depth understanding of [these] social phenomena" (Quintero, 2012). When researchers look for answers where they believe the data are better, "rather than where the truth is most likely to lie," this has been otherwise called the "streetlight effect."

According to Freedman (2010):

> The fundamental error here is summed up in an old joke scientists love to tell. Late at night, a police officer finds a drunk man crawling around on his hands and knees under a streetlight. The drunk man tells the officer he's looking for his wallet. When the officer asks if he's sure this is where he dropped the wallet, the man replies that he thinks he more likely dropped it across the street. Then why are you looking over here? the befuddled officer asks. Because the light's better here, explains the drunk man.

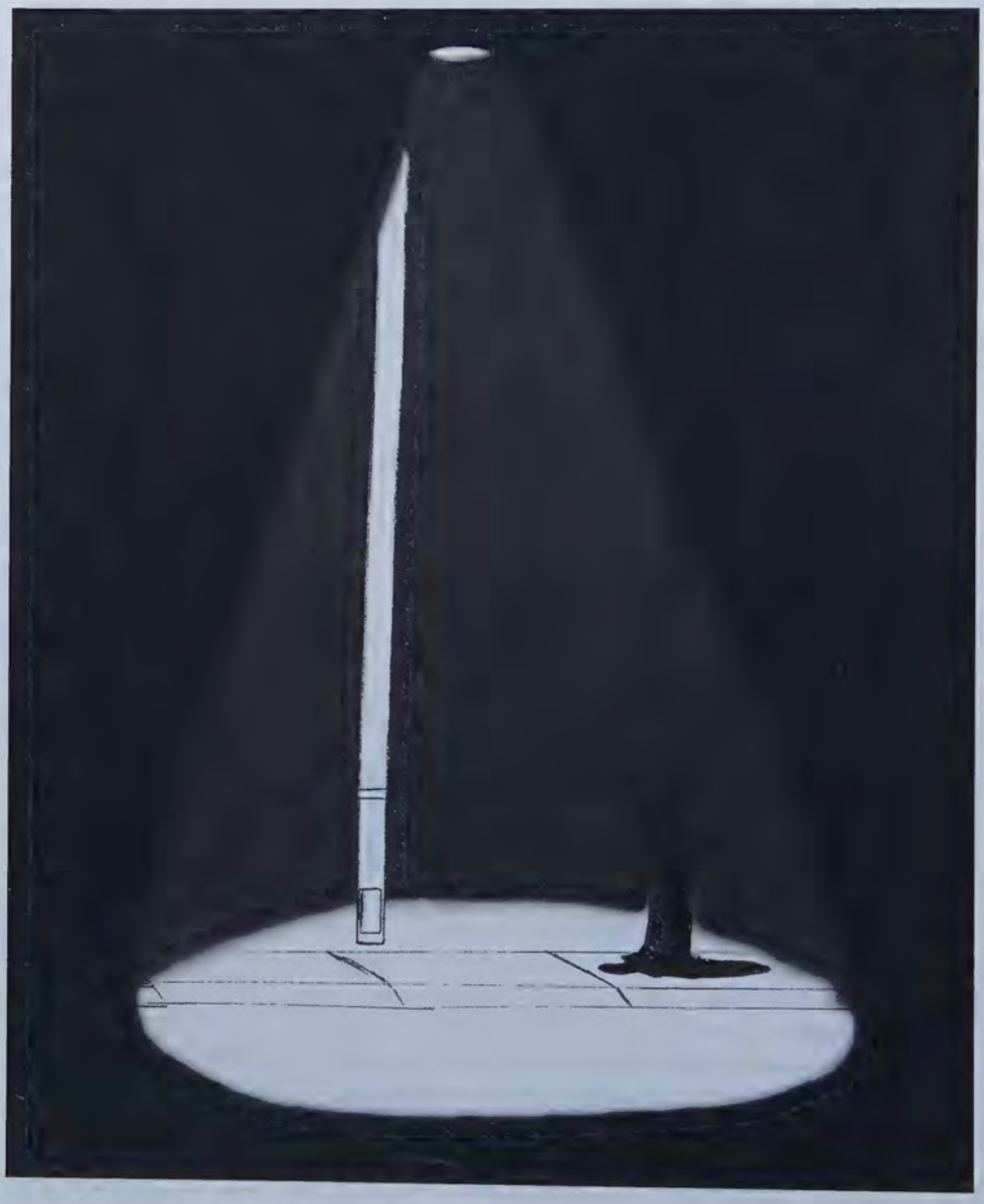

Figure 0.2 A drunk man under a streetlight.

The drunk is in good company here. Researchers often do not question whether they are looking for the right answers in the right places, they just defer to where they might find what they perceive as the best, and in this case most objective answers available, regardless of where else perhaps better and more accurate answers might exist. "Many, and possibly most, scientists spend their careers looking for answers where the light is better rather than where the truth is more likely to lie ... [all the while] ... hoping it turns out to be relevant" (Freedman, 2010). With this contention I also wholeheartedly agree, particularly in the case of VAMs. Correspondingly, I continue to position my research given my beliefs about what mathematics and statistics can and cannot do, particularly when used in isolation of other related, valid, and supportive data.

That said, I write here, definitely with passion and I hope some lucidity, but with the honest goal to let the readers be their own judges on the usability of VAMs and what they promise and contend for America's public schools and public school students. Take in, understand, think critically, and (hopefully) engage!

ACKNOWLEDGMENTS

When thinking throughout the year while writing this book about whom I would like to acknowledge for helping me, the person in the first position never changed. While no longer with us in the physical world, I will forever acknowledge my father, Dr. Roy Amrein, as the person who has had the greatest influence on me as a person and as a scholar. I know he is still with me, watching over me, and providing me with his guidance, encouragement, support, and undying love. I know I have made him proud!

I would also like to acknowledge my mom, Sharon Amrein, for she too, with her kind heart, patience, sense of caring, cute sense of humor, unconditional love, and calm demeanor, has influenced me greatly; and my brother, Dr. Derek Amrein, who is so similar to my mom, and who likewise has positively influenced me throughout our years as siblings. I would like to acknowledge him specifically for his kindness, thoughtfulness, enthusiasm, sense of humor, and his enduring dedication to our family, especially since the loss of our dad.

I also want to acknowledge my husband of now 10 years, Richard Beardsley, who is not only an amazing husband, but a wonderful father to our two children, Aidan and Blythe, and my two stepdaughters, Jessica and Alicia. If it was not for him, and the inordinate amounts of time and energy he took to protect me and my writing throughout the year, in many instances taking on all parenting roles and responsibilities, I might have lost my sense of self, not to mention my sense of mind. Related, I would like to acknowledge Aidan and Blythe for being such wonderful children and always making me laugh and smile, sharing with me their warmth and love, and helping me also keep perspective of the things that really matter most in my life. The two of them are and forever will be my best and most favorite accomplishments! I would also like to acknowledge Jessica for her generosity and her spirit and flare, and Alicia for her selflessness, helpfulness, and genuine kindness. They have also helped me in their own unique ways, often keeping me stable and sane as well.

As for those who helped me more directly with the book, I would like to thank Dr. Tom Haladyna for his editorial assistance, and general advice, guidance, and prudence, especially in terms of his knowledge and expertise in educational measurement, testing and assessment, and statistics. In addition, I will forever acknowledge my main and everlasting mentor, Dr. David Berliner. He and I have now worked closely together for 15 years. I am sure those who

know, or know of, David understand that I could not have asked for a better, wiser, more insightful, and more ardent mentor and friend. Similarly, I will forever acknowledge one of my other eternal mentors, Dr. Gene Glass. His brilliance I will continuously find inspiring. Thanks are also in order for Dr. Chris Clark who helped me find the strength and inspiration to get this book off of the ground.

Thanks also to former public school teacher and current value-added specialist, Dr. Sarah Polasky, who helped me edit every chapter herein, even one week after recovering from the birth of her second child. Thanks are also in order for former public school teacher (and soon to be Dr.) Jessica Holloway-Libell. As my graduate research assistant, she too offered me the prompt editorial assistance and intellectual insight that (hopefully) will also make this book what I have set it out to be. To my other graduate student, (soon to be Dr.) Noelle Paufler and her husband Tibor, thanks are in order for both of them and their creative assistance with many of the figures, images, and illustrations included herein. Thanks also go out to Dr. Kent Sabo and Dr. Shelby Maier for their similar contributions. And to my former graduate student, Dr. Clarin Collins, many thanks for sharing the same passion for this research, particularly in the Houston Independent School District (HISD) from which some of this research evolved.

Within the Mary Lou Fulton Teachers College at Arizona State University, I must acknowledge the colleagues whom I also admire and treasure. Most importantly, I would like to acknowledge Dean Mari Koerner. As the Dean she has supported me the most. Not only is she a great leader, she is also an amazing friend who always keeps me laughing. And thanks to my other good friends and colleagues at ASU: Drs. Anna Arici Barab, Wendy Barnard, David Carlson, Gustavo Fischman, David Garcia, Ida Malian, Suzanne Painter, Jeanne Powers, Joe Ryan, Frank Serafini, and Kate Weber. I honestly love what I do, in great part because I get to work with wonderfully smart, bright, enthusiastic, inspiring, and often-amusing colleagues like all of you!

Finally, to the three practitioners who have also helped me throughout this journey. To Robert Morse, also a local public school practitioner, who helped me make this book and its at times highly complex material as (hopefully) comprehensible as possible. To one of my best friends from the fifth grade, Andrea Stark, for keeping my academic perspective grounded in the current everyday situations of the American public school teacher every time we get together. And to my sister-in-law, Michele Amrein, a former public school teacher turned teacher educator who has also had a keen sense for "keeping it real."

My father would also have been so proud of the person she has become, as he was always so passionate about us contributing to America's public school system and, in particular, dedicating our lives to serve its purpose as per its Thomas Jeffersonian roots. As my father would have said, could he have said so himself, "An educated citizenry is a vital requisite for our survival as a free people." To acknowledge all public school teachers and what they continue to strive to do for us in service to, and in protection of, our democratic liberties, and despite the prevailing forces continuously working against them, thank you also for everything you do.

PART I

Introduction

1

SOCIALLY ENGINEERING THE ROAD TO UTOPIA

So long as there are men there will be wars.

Albert Einstein

The summer before I began writing this book I spent time with my family traveling in Asia, throughout the remarkable landmarks of China, the rainforests and rice fields of Vietnam, the pristine beaches and waters of Thailand, and the beautiful but war-ravaged and still war-recovering Cambodia. While the whole trip was indeed moving, what I found most inspiring was our time well spent in Cambodia. Walking the ancient temples of Angkor Wat, meeting some of the kindest people in the world, especially given their palpable Buddhist and Hindu principles and beliefs, and experiencing everything we possibly could of Cambodian culture.

To my chagrin, I did not know until my arrival (or recall from my courses in world history) that the country recently survived one of the bloodiest genocides in history, one that occurred as recently as 1975 at the end of the Vietnam War, following America's invasion of Cambodia and following Nixon's carpet bombing of the Viet Cong occupying the country's eastern borders (see Figure 1.1).

Soon after America withdrew its forces, the country's capital city, Phnom Penh, fell, and the communist regime of the Khmer Rouge ascended. If you recall the Academy Award winning movie, *The Killing Fields*, you might remember the story, and the true account of a Cambodian photojournalist, Dith Pran, who befriended the *New York Times* correspondent, Sydney Schanberg. Schanberg by happenstance abandoned Pran nearing tyrant Pol Pot's bloody "Year Zero" takeover and cleansing campaign (Putnam, Smith, & Joffé, 1984).

The idea behind the "Year Zero" crusade was that in order for the communist Khmer Rouge regime to rule supreme (1975–1979), everything including the country's cultures, traditions, and many of its people had to be purged. This led to "necessary violence" and mass genocide which claimed the lives of approximately two million objectionable civilians of the then nearly eight million who resided in the country. The "undesirables" who were targeted and ultimately executed (or died of starvation, disease, torture, or in labor camps in the killing fields) included the country's most educated, specifically Cambodia's professors and teachers,

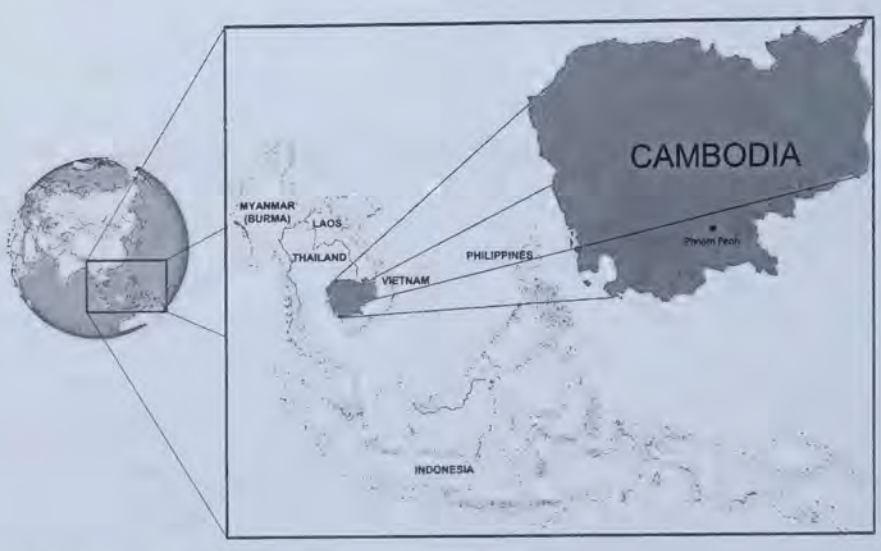

Figure 1.1 Map of Cambodia.

doctors, lawyers, police and businessmen and women, artists, musicians, writers, and the like. All who were deemed professionals, intellectuals, or even literate because they wore eyeglasses were pejoratively viewed as oppressive and subversive, and singled out for extermination. They were besieged for their alleged capacities to resist the Khmer Rouge's ultimate end – a utopian, self-sufficient, agriculturally focused, ethnically pure, and classless, communist-controlled state.

While the Cambodian photojournalist, Pran, was one of the "undesirables," he barely survived, undercover as his peasant alter ego – an uneducated taxi cab driver. He eventually escaped by walking his way out of the killing fields, only to discover that more than 50 of his family members had been killed. He lived the rest of his life working in the U.S. as a photographer for the *New York Times* and relentlessly promoting global awareness about the Cambodian Holocaust. He passed away in 2008 at the age of 65, with his horrific thoughts and memories "still alive to [him] day and night" (Dith Pran, 2004; see also Farrell & Rummel, 2008; Martin, 2008). In his "Last Words" interview with the *New York Times*, Pran noted that "in order to survive, you [had] to pretend to be stupid, because they [did not] want you to be smart. They [thought] that the smart people [would] destroy them ... you [had] to show that you [were] not a threat to [survive]" (Farrell & Rummel, 2008). The Khmer Rouge's attempts to contrive and influence attitudes and social behaviors by cleansing (i.e., social engineering by genocide) still haunt the Cambodian people today.

What is resounding is that mass murder and methodical cruelty as preludes to utopia have occurred in many ways during many times and in many parts of the world in the past. Social engineering by genocide, whether via religious, racial, ethnic, or national group cleansing, is not a new historical phenomenon. In terms of religious cleansing, during the Middle Ages, a period generally fraught with intellectual obscurity and atrocity, inquisitional courts implicated, indicted, and burned at the stake perceived heretics who spoke out about or against, or thought differently from the politically dominant Catholic Church. With regards to racial cleansing, within the last century the Nazi-sponsored Holocaust aimed to completely exterminate the Jewish race, the race that according to Hitler was not a religion. This ultimately

resulted in the execution of six million Jews and the slaying of five million other "undesirables" including homosexuals, gypsies, people with disabilities, and other political and religious adversaries. Regarding ethnic cleansing, just 20 years ago, nearly one million people in Rwanda were murdered due to ethnic tensions between the Tutsi and Hutu peoples. Concerning national groups, America's actions in the name of Manifest Destiny are also considered by many to qualify as social engineering by genocide. The numbers of Native Americans and Mexican nationals who perished during times of rationalized expansion and redemption are comparable to the fatalities recorded during the Nazi-sponsored Holocaust (Cesarani, 2004; see also Smith, 2013).

While all these persecutors used various forms of genocide to socially engineer the roads to their interpretations of utopia, however, none of these instances matched what occurred in Cambodia, not in terms of the sheer numbers of deaths or types of terror, but in terms of goals and objectives. The Khmer Rouge's deliberate and systematic efforts to destroy a group of perceived intellectuals (not to say that these groups were not intellectuals themselves) and the regime's efforts to extinguish their allegedly threatening and abhorred ideas, add much to our thinking about intellectual genocide and social engineering. In Cambodia, intellectual genocide was used to promote communism, thwart any hopes or prospects of a democratic state, and turn the country over to a state of the people, where the people were defined as the country's workers, farmers, laborers, and peasants. In the end, the Khmer Rouge's aims and actions to exterminate all intellectuals and socially engineer a totalitarian and utopian dictatorship devastated the country instead. The country is still in the process of healing and recovery (Jenkins, 2012).

Social engineering in the political sciences

However, while social engineering is often viewed negatively, particularly, for obvious reasons, in cases of social engineering by intellectual, religious, race, ethnic, or national group genocide), social engineering by other means is quite common in the political sciences. Most governmental and private groups, in their efforts to promote or protect private or the public's perceived interests, attempt to methodically sway or change public attitudes and behaviors in one way or another. While the goal remains the same – to control or influence society, the polis, or city-state by eliminating or reducing human agency or agents (Marx, 1995) – in the political sciences, social engineering relies on more subtle means. Here, social engineers use other powerful instruments to influence attitudes and social behaviors, ultimately to promote and promulgate a societal, oft-utopian ideal. Social engineering tools include scare tactics; propaganda and rhetoric often vetted through mass media outlets; generalizations, assumptions, and rationalizations; and incentives and disincentives to ultimately engineer the social behaviors desired.

It should be mentioned that social engineering is not always adversarial, however. For example, at the end of the Civil War, President Abraham Lincoln used highly controversial tactics to socially engineer the country's path to free the slaves of the Confederate South. His notorious actions, along with his remarkable leadership, led to his Emancipation Proclamation of 1863, which ultimately led to the adoption of the Thirteenth Amendment to the U.S. Constitution in 1865. This made slavery and involuntary servitude illegal throughout the U.S. This serves as a prime example of how governmental policies also constitute a powerful social engineering tool, again when control is the means and a utopian ideal the end.

The emancipation of the slaves in the South was certainly at the time utopian, while more often than not perceived as impractical and unreasonable.

By definition, a public policy is in itself a tool used by governments to define a course of action that will ultimately lead to a high-level, supreme, and desirable end. Sometimes public policies can be unwise, however, when seemingly principled means produce unintended, unanticipated, and perverse consequences instead. Often, the inadvertent effects outweigh the intended consequences for which the policy was engineered in the first place, as social engineers' attempts to control and alter their environments are imperfect and often contradict or distort what is virtuous, pure, and good (Campbell, 1976; Wheatley, 1992).

In 1962, for example, Rachel Carson published her acclaimed book *Silent Spring* – the book now widely credited with initiating the environmental movement and the founding of the Environmental Protection Agency. In it, Carson wrote lucidly and passionately about the U.S. Department of Agriculture's assault on the environment. She criticized the department for turning a blind eye, and disregarding the many empirical research studies from which she drew her conclusions but that were not new or inaccessible, just conveniently ignored (Griswold, 2012). She also criticized the department's ignorant and purchased acceptance of the chemical industry's oratorical, "research-based," and lobbied claims that their synthetic pesticides, from which they were profiting, were the best to control and kill agricultural pests. In fact, the government's excessive spraying of DDT (dichlorodiphenyltrichloroethane) was at one time so universally accepted that the public forgot to question its use. That is, until Carson wrote *Silent Spring*, which ultimately led to the banning of DDT in 1972, at least in the domestic U.S.A.[1]

Here, the public, having been socially conditioned, naïvely trusted that simply because the pesticide was familiar, it was safe and effective. The public's opinion had been socially engineered into believing DDT and other harsh pesticides were necessary and harmless. Unfortunately, however, what ended up being destroyed were the ecosystems into which the pesticides were introduced.

The culprits included those who socially engineered the severity of the problem, those who amplified public alarm about the problem, those who chemically manufactured the solution to the problem, those who marketed false promises about the solution to the problem, and those who subsequently profited by having provided the best, if not only solution to the problem turned severe. The unintended effect was that what the government and industrialists provided to control the problem caused effects markedly worse than the problem needing to be controlled in the first place (agricultural pests). Metaphorically speaking, the spring had gone silent because of the absence of the non-pest wildlife that the synthetic pesticides also killed (Carson, 1962; Griswold, 2012; see also Amrein-Beardsley, 2009b).

Social engineering via the use of federal and state tax policies to encourage or discourage societal behaviors has also become commonplace. This is especially true with the U.S. Department of the Treasury and its use of tax credits, breaks, deductions, and exemptions since the passage of the Tax Reform Act of 1986. The rationale is that via the tax system and its provisions, many societal behaviors and activities can be encouraged using incentives for things like charitable giving, homeownership, and the purchase of health insurance. Similarly, many societal behaviors and activities can be discouraged through disincentives like "sin" taxes to reduce alcohol and tobacco use or gambling. More recently "fat" taxes have been designed and introduced to curb the sedentary activities seen as contributing to America's obesity epidemic. Empirical evidence substantiating that the use of federal and state tax policies to

encourage or discourage, or socially engineer these or other societal behaviors in fact works is scarce, however.

The most obvious of unintended consequences that come along with such (dis)incentive systems are tax distortions. Such distortions often come about as a result of system gaming techniques, whereby for those who are best equipped to understand and manipulate the incentives and disincentives put into place can exploit them to their advantage. Take, for example, the recent scandal with Apple Inc., whereby Apple CEO Tim Cook and associates have continued to use opposing taxation loopholes in Ireland and the U.S. to avoid paying billions in taxes. The U.S. Senate subcommittee investigating Apple also investigated Microsoft, Hewlett-Packard, and other multinational companies, charging that they too have exploited loopholes in the U.S. federal tax code avoiding billions in U.S. taxes as well (Associated Press, 2013).

The unintended effects that come along with such (dis)incentive systems then become cyclical. When the distortions are recognized and new tax policies are devised and implemented to offset and disincentivize the distortions discovered, the perceptibly more reasonable and sound tax remedies cause yet another series of unintended effects, and the cycle continues. This is precisely why many of the country's finest tax experts strongly discourage social engineering via federal and state tax policies (Lazear & Poterba, 2006; see also Chamberlain, 2005). The U.S. Department of the Treasury's attempts to control or influence society, the polis, or the city-state by using social engineering tools has not, for the most part, worked in the ways intended and has rather caused loopholes that simply cannot be closed.

City planning, housing, and transportation policies have also been devised to engineer the social world, namely to protect certain social environments and to reinforce the artificial borders within which various social classes reside. For example, the U.S. Department of Housing and Urban Development's strategies of the not-so-distant past were put into place to keep lower-income people out of certain areas via residential segregation. This has been supported by the strategic planning of urban and suburban areas, the U.S. Federal Housing Administration's former redlining policies, and the clever planning of public transportation systems, or more specifically city bus and subway routes and stops between, for example, Washington DC and Georgetown, Boston and Lexington, and New York City and Newark, New Jersey. Inter-area transports were once calculatedly devised to deter inter-neighborhood migration and to further segregate groups of (poor) people, and to keep them from other (wealthy) groups of people. This stimulated and resulted in residential segregation, which understandably led to other forms of segregation, racism, and classism given the other demographic variables that are still highly correlated with neighborhood and community characteristics (i.e., racial, ethnic, nationality, language, and social class characteristics). This also led to increased discrimination in that these groups became more susceptible to prejudicial treatment, having been increasingly defined by geographic borders and having become more concentrated and isolated than before.

Collectively, these interrelated conditions perpetuated even greater levels of un- and under-employment and, accordingly, higher levels of per capita poverty. This led to increased rates of violence, crime, and other anti-social behaviors over time (e.g., illegal drug use, noncompliance and disobedience, defiant behaviors toward authority, etc.). Likewise, those geographically marginalized unendingly had more difficulties gaining entry into decent jobs that would better help them "pull themselves up by their bootstraps," and increase their earning power and cultural and human capital, and improve upon their subsequent chances to succeed

(see also Berliner, 2012). These conditions inhibited access to quality healthcare, typically as a result of a lack of or no medical or dental insurance, to affordable housing and other basic services and resources, and to early childhood and higher educational opportunities. All of these and other interrelated conditions contributed to the increased ghettoization of urban neighborhoods throughout the U.S. (Anyon, 1997; Kozol, 2005; Marx, 1995; see also Caro, 1975 for an account illustrating how Robert Moses used his power to (mis)shape the city and state of New York).

Combined with "white flight" (i.e., the movement of typically white, European American people from highly concentrated urban to suburban areas), this also contributed to the increased segregation of America's public schools in these neighborhoods. This has been most aptly illustrated in Jonathon Kozol's books (see, for example, Kozol, 1991, 1995, 2000, 2005).

Nonetheless, while policies and laws have been enacted since to counter the negative effects of these dystopian policies (e.g., fair housing laws, revoking "separate but equal" strategies, educational desegregation laws following *Brown v. Board of Education* in 1954), it is quite clear that these compensatory and restorative policies have had little consequence. In New York City, for example, half of the city's 1,600 schools currently have enrollments that are over 90% Hispanic/Latino(a) and black (Fessenden, 2012). It is certainly clear that racial segregation still exists and that current federal and local policies continue to perpetuate such problems (see also Berliner, 2012; Orfield, 2009).

Social engineering in American educational policy

It is also certainly clear that federal and local educational policies (e.g., district and school zoning, charter and magnet schools, open enrollment, vouchers and other choice policies, policies promoting the privatization of America's public schools) continue to perpetuate these societal problems (Boustan, 2011; Eaton & Rivkin, 2010; Kozol, 2005; Street, 2005). Added to the burden of such educational policies, it is also dauntingly difficult in poor communities to recruit and retain highly effective teachers, especially given inequitable teacher salaries and benefits. It is also difficult to secure, construct, and maintain sufficient educational facilities; to obtain and sustain quality resources and technologies and the highly effective teachers who know how to use such resources in instructionally relevant and efficacious ways; to hire and retain highly effective school administrators in progressive leadership positions; to promote decent levels of parent, family, and community support and involvement, and the like. The absence of these conditions clearly thwarts school achievement and inhibits the opportunities for students to learn in schools situated in poor, segregated communities (Anyon, 1997; Berliner, 2010, 2012; Biddle, 2001; Eaton, 2007; hooks, 2000; Katz, 1989; Kozol, 1991, 1995, 2000, 2005; Nieto & Bode, 2008). This is yet another tale of policies that have been socially engineered and implemented because they make sense, ideologically and theoretically, but that generate unintended, unanticipated, and perverse consequences instead.

Within education, there have also been federal and local educational policies many attribute to politically conservative governmental and private attempts at social engineering. The U.S. continues to reintroduce the teaching of phonics in classrooms, not based on the research evidence but driven by a back-to-basics ethos, eternally taking the country back to the Victorian era when phonics was first introduced. This concept has socially engineered itself so deeply into our thinking about reading/language arts instruction that the U.S. can hardly

break free from its primitive logic. As well, politically conservative policies supporting the use of English-only in schools, and preventing various forms of bilingual education, have been attributed to governmental attempts to socially engineer an English-only, English-first society. Even school uniforms have been mandated at schools to deter or control materialism and gang affiliations or activities, and to increase student achievement. Yet, as Goodman (2006) notes, there is no scientific evidence whatsoever supporting the claim that dress codes increase student achievement or deter much of anything (see also Simonson, 1998). Yet the public still sees this as a rational approach to reach a commonsense end. As well, conservative public and private pundits immortalize these views using the same social engineering tools of the past, available at their discretion (e.g., scare tactics; propaganda and rhetoric; generalizations, assumptions, and rationalizations).

A paradigm case

The federal report *A Nation at Risk* serves as the paradigm case of social engineering by rationalistic means in American education. In 1983, the National Commission on Education released *A Nation at Risk* (U.S. Department of Education, 1983) in which the commission argued that schools in the U.S. were performing poorly in comparison to other industrialized nations and that our national security was therefore in jeopardy. This fear had been lingering since the Russians beat the Americans to space with their successful launch of Sputnik in 1957. However, in *A Nation at Risk*, the perpetrators of our national frailties – America's public schools – were at last named. Citing lower than expected national and international student test scores, deterioration in school quality, a diluted and diffused curriculum (particularly in mathematics and the sciences), and setbacks on other indicators of America's preeminence, the commission triggered a nationwide panic. They prompted anxiety and alarm about the weakening condition of the American public school system and, consequently, the nation's global supremacy and economic dominance.

Soon thereafter, Berliner and Biddle (1995) scientifically and effectively demonstrated that the claims made in *A Nation at Risk* were erroneous and grossly exaggerated (see also Berliner, 2011). They demonstrated that America's low but still above average test scores were more related to the demographics of the nation, as compared to the other industrialized nations that participated in the international tests. They exposed instead the very real problems impeding greater success in America's public schools and demonstrated that, despite the evidence, the federal government set out to "manufacture" or socially engineer a "crisis" about the American public school system.

Nonetheless, the commission's *A Nation at Risk* still had a massive impact. The report undeniably changed the general public's opinions and attitudes about the American public school system, for the first time in history, in line with the alleged desires of its social engineers. Its effects still echo today.

Increased accountability for results

Out of *A Nation at Risk* the stronger accountability movement was born. It resulted in a spirited move towards more rigorous standards and increased accountability mechanisms, both of which were to help bring the nation out of its perceived (and, as argued, socially engineered) educational crisis. The commission recommended that states institute higher standards to

homogenize and improve curricula and administer large-scale standardized achievement tests more frequently to hold students and educators accountable for meeting those standards. It was at this time that Koretz (1996) noted that the nation began relying almost solely on measurement-driven educational policies to turn around what was perceived, and now tangibly feared, as America's failing public school system and slipping supremacy.

Thereafter, nearly every state developed its own enhanced standards, testing, and accountability policies, all of which focused primarily on increasing student achievement on states' large-scale standardized achievement tests, as well as requiring the kind of instruction (i.e., focused on basic skills and facts) that was thought necessary to increase student achievement on these tests (Center on Organization and Restructuring of Schools, 1995; Meyer, 1997). The Measure and Punish (M&P) Theory of Change was that by holding districts, schools, teachers, and students accountable for performance on the states' large-scale standardized achievement tests, administrators would supervise the schools better, teachers would teach better, and students would learn more, particularly in the nation's lowest performing schools. Soon thereafter, students' test scores would increase, the nation's prominence would be reclaimed, and the nation would achieve the utopian society that was now so deeply desired (see Figure 1.2).

Ironically, thanks to those who socially engineered and successfully engrained this rational systems theory of educational improvement into the minds of so many, and most importantly into the consciousness of policymakers on both sides of the partisan divide, the same M&P

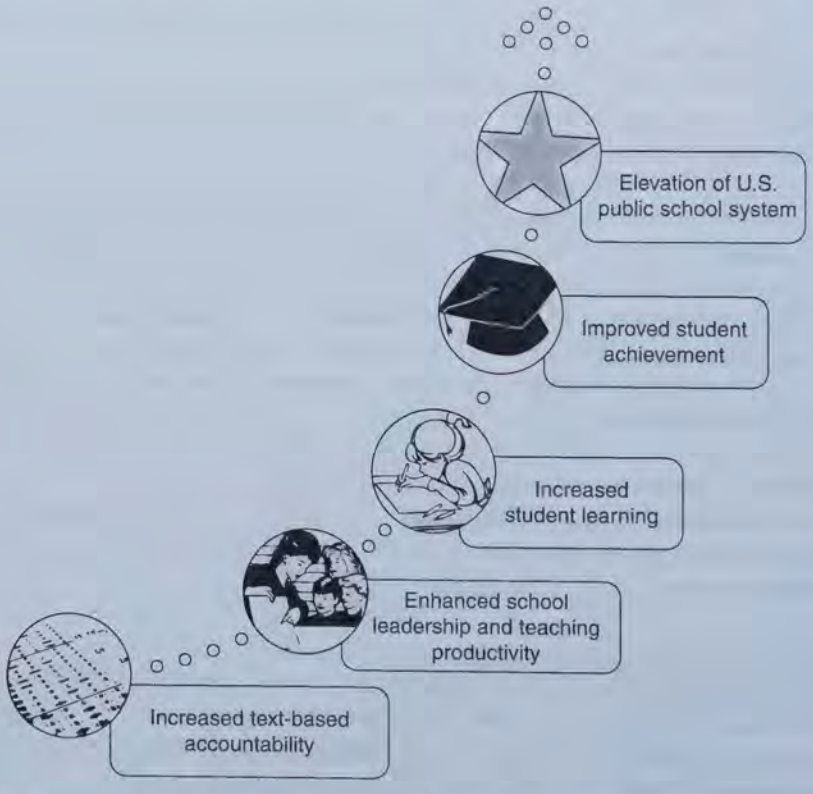

Figure 1.2 Measure and Punish Theory of Change logic map.

Theory of Change continues to hold. In fact, it continues to shape the thinking of America's educational policymakers today.

This is likely because it is such an overly simplistic causal model, while a better, more fully explicated causal model would be much more difficult to understand and, more importantly, difficult to defend. A more realistic interpretation of the M&P logic would focus not just on teachers and administrators but the other agents of change at play. These include but are not limited to: (a) students' levels of intelligence, prior academic achievements and experiences, and aptitudes; (b) students' levels of social capital, as first defined by Hanifan (1916) as "those tangible substances [that] count for most in the daily lives of people" (p. 130); and (c) students' levels of risk. This last I define, here and throughout this book, as the populations of students in schools who, disproportionately to their low-risk peers, have emotional/learning disabilities and/or come from high-needs, high-poverty, English-language deficient, culturally isolated (e.g., inner-cities, hoods, ghettoes, enclaves, and American Indian reservations), and often racial/ethnic minority backgrounds; although, it should be noted that a large number of students at risk are not students of color. Such major factors unquestionably complicate the overly simplistic M&P Theory of Change, but these factors must be overlooked should simplistic solutions continue to be sought, not to mention perpetually pursued.

Since the turn of the millennium, this has been most notably evidenced through the passage of the No Child Left Behind Act (NCLB, 2002), the Obama administration's proposal to revise the Elementary and Secondary Education Act (ESEA) and to "reform" NCLB (U.S. Department of Education, 2010a), and most recently via the federal *Race to the Top* (RttT) initiative (RttT, 2011). Billions of dollars in federal stimulus monies have already been and continue to be awarded to states – $4.35 billion being the initial sum budgeted for RttT (Duncan, 2009a). In line with the same M&P Theory of Change, the states receiving RttT awards, in exchange for these desperately needed federal funds had to promise to use students' test scores for even more consequential purposes, especially for teacher evaluation, termination, and compensation. The 40 states that applied for RttT funds in the first year of funding also had to agree to adopt even stronger accountability mechanisms if they were to secure waivers excusing them from not meeting NCLB's improbable goal that 100% of the students in their schools would be academically proficient by the year 2014. This too was a utopian, albeit still distant target, one to which every state in the nation[2] was predicted to fall far short (Dillon, 2010b; Duncan, 2011; Layton, 2012).

Notwithstanding, within any of these documents one can find the familiar legends and myths of a golden age of American education: that schools in the U.S. were once internationally superior but no longer are; that other industrialized nations are outperforming the country that was once the best in the world; that America must raise its expectations of students, teachers, and administrators and hold them more responsible for doing their jobs and meeting even higher expectations; and that the nation must do a better job at penalizing failure and rewarding success (just as the marketplace penalizes laziness and rewards hard work). Paradoxically, these are the reasons still being used to justify America's newly "reenvisioned federal role in education" (U.S. Department of Education, 2010b), re-envisioned again just three years ago, despite the fact that a nearly identical version and nearly indistinguishable vision was introduced 30 years ago in *A Nation at Risk* (U.S. Department of Education, 1983).

Not-so-temporary insanity

According to Albert Einstein, the definition of insanity is to repeat the same behaviors over and over again in the hope that different results will materialize the next time, perhaps after this is fiddled with or that is fine-tuned. Here, it seems, educational policymakers propagating and regenerating policies based on the same M&P Theory of Change are epitomizing insanity. For the past 30 years they have repeatedly justified nearly the same policy actions while using nearly the same social engineering instruments as the means to justify their anticipated ends.

These tools include the same scare tactics (e.g., about America's accelerating inferiority); rhetoric (e.g., the titles *A Nation at Risk*, *No Child Left Behind*, and *Race to the Top* in themselves); propaganda (e.g., the federal education administration paying a prominent African American authority $240,000 to promote NCLB on a nationally syndicated television show; Toppo, 2005); generalizations, assumptions, and rationalizations (e.g., what is needed is greater authoritative control, higher standards and accountability, and an expanded application of business model thinking in education); and incentives (e.g., tuition waivers for students, salary bonuses for teachers and administrators) and disincentives (e.g., prevention of student grade-to-grade promotion or high school graduation, teacher dismissal, administrator termination, school closure).

Yet an overwhelming preponderance of empirical evidence indicates that this M&P Theory of Change is grossly flawed and misguided, and has not yielded its intended effects, ever! This theory of change has never improved the nation's global standing, nor has it positively impacted the ever-persistent achievement gap (i.e., the observed disparities between students with relatively low or high levels of poverty or who are typically Asian American and white versus their African American, Hispanic/Latino(a), and American Indian peers).

Nor has this theory of change ever improved the test scores used across the nation to determine college admissions decisions (e.g., SAT or ACT scores), high school graduation rates, college going rates or levels of preparedness, and the like (Amrein & Berliner, 2002; Au, 2009; Haney, 2000; Heubert & Hauser, 1999; Hursh, 2008; Johnson, Johnson, Farenga, & Ness, 2005; Klein, Hamilton, McCaffrey, & Stecher, 2000; Mathis, 2011; Nichols & Berliner, 2007; Nichols, Glass, & Berliner, 2012; Orfield & Kornhaber, 2001; Timar & Maxwell-Jolly, 2012).

Relatedly, many early proponents of this M&P Theory of Change have since become its strongest opponents. This has happened specifically since the passage of NCLB as they themselves observed the absence of NCLB's intended effects and the presence of (too many) unintended consequences in their place. For example, more than 50 Republican House and Senate members took issue specifically with NCLB's testing mandates, noting the mandates' lack of intended but abundance of unintended, perverse consequences instead (Weisman & Paley, 2007). Republicans in Texas, the state that then Governor George W. Bush used as a petri dish to culture NCLB for the nation (following former Texas Governor Ross Perot's lead), have recently turned against its high-stakes testing components as well (Pauken, 2013). Diane Ravitch, one of the ardent crafters of NCLB, and for years a devoted advocate, has since made it her mission in life to undo its effects. Not a day passes during which she does not contribute to the evidence base confirming that NCLB (2002), the reauthorization of the Elementary and Secondary Education Act (U.S. Department of Education, 2010a), the *Race to the Top* initiative (RttT, 2011), and the whole M&P Theory of Change are doing more harm than good (Ravitch, 2010a; see also Denby, 2012; Mathews, 2013).

It should be noted, however, that a few academic researchers disagree with the research majority who have collectively demonstrated that this theory of change does not work. They have empirically argued their points with counterfactuals (Chetty, Friedman, & Rockoff, 2011; Hanushek, 2011; Hanushek & Raymond, 2005; Raymond & Hanushek, 2003; Stotsky, Bradley, & Warren, 2005; Winters, Trivitt, & Greene, 2010). As well, a handful of other highly politicized proponents have argued (and continue to argue) that the theory of change works, although they have argued their cases on less scientific and more dogmatic grounds (Coleman as featured in Lewin, 2012 and in Rotherham, 2011;[3] Kress, 2011;[4] Phelps, 2011;[5] Rhee, 2011;[6] Spellings, 2012[7]).

Altogether, they continue to advance the theory's hyper-reliance on large-scale standardized achievement tests, the tests that along with high-stakes consequences are (still) to be used as both the positive and negative motivators (still) needed to increase the (still) substandard levels of student performance (still) at issue. The irony here is that, again, research evidencing that these policies have ever borne their intended fruit is (still) very much absent. We as a collective nation still have grave concerns about our international presence and dominance. These are the same grave concerns we had after the commission released *A Nation at Risk* (U.S. Department of Education, 1983) 30 years ago, and despite the fact that federal and state educational policymakers have relied on this M&P Theory of Change to improve America's education condition since. Actually, state-level policies aligned with this theory of change have been around even longer than 30 years.[8]

Intellectual genocide

What is worse is that, akin to the intellectual genocide witnessed in Cambodia, as a result of this M&P Theory of Change and the federal and state policies implemented to uphold it, large groups of students have been categorically stripped of their inalienable (yet not U.S. constitutional) rights to an equal, fair, appropriate, and proper education. Policymakers in nearly every state have now adopted even higher standards and even stronger high-stakes testing policies to keep the higher standards in check, although the seriousness of the state policies and the severity of the consequences or stakes attached to test output still vary by state (Nichols et al., 2012).

Notwithstanding, since the passage of NCLB annual state spending on large-scale standardized achievement tests rose from $423 million in 2002 to almost $1.1 billion in 2008. This represents a 162% increase compared to a 20% increase in inflation over the same time period (Vu, 2008). Most recently it was estimated that we now spend well over $1.7 billion on tests, although this is likely a gross underestimate given that such estimates traditionally account only for state- not district- or school-level tests. This also excludes the costs that come along with tests' associated resources (e.g., test preparation packets, workbooks, and materials) and other related costs (e.g., administrative and instructional time costs for test preparation; Cody, 2012). Most importantly, this illustrates that we as a nation, and our politicians on behalf of us as a nation, are placing a good amount of our taxpayers' bets on a bankrupt theory of change.

Furthermore, with the wide-scale adoption of these policies, many students, especially those being educated in the country's lowest performing schools, have been unremittingly deprived of opportunities to learn in ways that are not driven by large-scale standardized achievement tests. Since the passage of NCLB, many students, especially those in the country's

lowest performing schools, have been increasingly susceptible to unprofessional test-based practices including teaching to the tests (not to be confused with teaching to the standards); teaching using scripted and prefabricated curricula to ensure that what is taught aligns with what is tested; teaching time being taken up by test preparation, test practice, and test rehearsals; hyper-emphasizing the rote memorization of facts and basic skills; narrowing the curriculum to match the content and concept areas tested; and, related, teaching the tested subject areas that "count" (i.e., mathematics and reading/language arts) while marginalizing or even eliminating other curricular areas and activities that do not "count" on high-stakes tests (i.e., social studies, sciences, art, music, physical education, library sciences, and recess).

As well, typically low-scoring students, including inordinate numbers of non-English proficient and special education students, have been purged (i.e., expelled, suspended, or simply excused) from school during test administrations to keep them, the "undesirables" when it comes to high-stakes tests, from participating and pulling test scores down. Students have also been counseled out of school, convinced to explore other options (e.g., alternative, "last chance," or adult education schools), or persuaded to strive for General Education Diplomas (GEDs) instead of traditional high school certificates. Eliminating undesirable students eliminates their scores: the scores that, if included or preserved, would pull composite test scores down.

Those students, mentioned above, whom educators have deemed the least likely to post high enough test scores, have also been academically shunned. This has occurred particularly during the weeks leading up to high-stakes tests as educators, who will be held accountable and potentially suffer punitive sanctions if test scores are low, may perceive certain students as hopeless, and hence, the most undesirable to take the tests. Undesirable students have been known to be retained in grade or credit hours to keep them from being eligible for high-stakes testing cycles (e.g., by thwarting progression in high school as sophomores/juniors might not be eligible for testing in their sophomore/junior year; see for example Haney, 2000). Some undesirable students have altogether disappeared from school rosters; that is, when administrators have created rosters and registered students for high-stakes testing purposes.

In other cases, underperforming students have been wrongly moved into exempt categories (e.g., special education and ELL categories), as misclassifying these students will prevent them from dragging down the performance of the teachers or the schools as a whole (Amrein & Berliner, 2002; Haney, 2000).[9] Recognizing this as an issue, the federal government started mandating minimum rates of test participation (NCLB, 2002), but it seems this still occurs. This was most recently evidenced in the state of Florida (Chakrabarti & Shwartz, 2013; see also Derby, 2013), the state often distinguished for the high-stakes testing policies put into place by former Florida governor Jeb Bush, brother of former President George W. Bush and potential contender for the 2016 presidential election. Ironically, like it was with the undesirables in Cambodia, undesirable students in America's public schools have had to demonstrate that they are not a threat in order to survive in high-stakes school systems, especially in the school systems most susceptible under America's increasingly distorted, stronger accountability policies.

Conversely, educators have focused inordinately on the students who are on the edge of passing high-stakes tests ("on the bubble"). The thinking here is that if educators teach to the test well enough these students just might clear the cut scores and pass, and help to bump composite test scores, even if ever so slightly, upwards. Educators have used "selective seating"

practices in which the students expected to post high scores are seated among the students expected to post low scores, covertly encouraging cheating.

Educators have also overtly cheated, for example, by erasing and changing students' incorrect answers to correct ones, explicitly giving students correct answers, persuading students to revisit incorrect answers, and the like. Such cheating instances have been widely publicized, most recently in the cases disclosed in the public schools of Atlanta and Washington DC (Perry & Vogell, 2009; Rhee, 2011) as well as in the public schools of Arizona (Amrein-Beardsley, Berliner, & Rideau, 2010; see also Toppo, Amos, Gillum, & Upton, 2011; Upton, 2011).

Collectively, these are just some of the unintended effects that have come about due to this distorted, yet socially engineered, M&P Theory of Change, and the educational policies constructed in its likeness (Amrein & Berliner, 2002; Grodsky, Warren, & Kalogrides, 2009; Hursh, 2008; Nichols & Berliner, 2007; Ravitch, 2010a; Reardon, Atteberry, Arshan, & Kurlaender, 2009). In the end, these and other methods of "gaming" the system result in extra, albeit artificial, boosts in aggregate test scores (Haladyna, Nolen, & Haas, 1991). These are the boosts that those being held accountable via stronger accountability policies often desperately need, and these are the boosts that those across states who are promoting such educational policies often desperately need to preserve and prolong their political careers (see, for example, Slavin, 1989). The educators being held accountable often justify the actions in which they sometimes engage to yield such artificial boosts. They do this more so when they are under immense pressure to keep their jobs, keep their tenure, earn merit pay, and the like, and more so when they view that the accountability policies being devised to hold them accountable are unjust, unwarranted, and unreasonable (Amrein-Beardsley et al., 2010; see also Smith & Noble, 1997).

It is here that this M&P Theory of Change has not only used intellectual genocide as a Machiavellian means to an end, primarily by (a) purging undesirable students, (b) cleansing the curricula, (c) sterilizing the core subject and concept areas taught, (d) exterminating courses and activities that do not matter or do not matter as much when high-stakes tests approach, (e) shrinking students' time on the tasks that ultimately matter the most, (f) decimating students' other and often more important opportunities to learn, and (g) implementing other test-boosting practices well known to achieve higher, albeit artificial, test scores. It is here, too, that this theory of change has caused intellectual genocide primarily by denying many students proper, appropriate, fair, and equal opportunities to learn the things that will ultimately matter the most, versus the things that multiple choice test items normally included on large-scale standardized achievement tests typically reflect. It is here that this theory of change has thwarted many students' "pursuit of happiness" in America's public schools.

As demonstrated in Carson's *Silent Spring*, the ecosystems into which high-stakes tests are continuously being introduced, namely America's public schools, are what are being continuously destroyed. The culprits here, however, include those who continue to use social engineering tactics of the ever-present past to exaggerate the severity of the problem (e.g., politicians advancing such educational reform measures) and those who continue to amplify public alarm about the problem (e.g., conservative media outlets and think tanks). The culprits also include those who continue to refine and manufacture the solutions to the problem, lobby their solutions to those in political and policy-making positions, and profit by continuously providing the best if not only solutions to the educational problem (e.g., educational measurement, testing, and other educational research-based companies and non-profits).

As in the aforementioned examples of the U.S. Department of Agriculture, the U.S. Department of the Treasury, and the U.S. Department of Housing and Urban Development, the control measures put in place by the U.S. Department of Education and others are ultimately causing markedly worse effects than the initial troubles that needed to be controlled or reformed in the first place. Ironically, they too are working alongside and under the advisement of many of the industrialists with vested private and financial interests (e.g., CTB McGraw-Hill, the Educational Testing Service, Harcourt Educational Measurement, Pearson, Riverside Publishing). Yet they are buying into the products nonetheless. Metaphorically speaking, America's public schools are being silenced, because educational opportunities and experiences are being annihilated in many cases by high-stakes tests, particularly in the schools most in need of genuine guidance and support.

The 30-year road trip to nowhere

But why is it that this M&P Theory of Change has held so firm? Why is it that this theory of change has become commonplace when there is still virtually no scientific research evidence verifying that it works? Why is it that after 30 years of tinkering with this theory of change, we still have the ever-present grave concerns about America's public school system that so alarmed the public 30 years ago in *A Nation at Risk* (U.S. Department of Education, 1983)? If this theory of change, socially engineered and now so deeply engrained in the minds of many, indeed worked, why would we still have such concerns about America's public school system, after the implementation of 30 years of repeated educational policies aligned with the same theory?

Tyack and Cuban wrote about a similar saga in their book *Tinkering toward Utopia* (1995), arguing that in fact most educational reform initiatives are ahistorical and as such fail, for many reasons but particularly when bureaucratic controls become the means to bring about utopian ends (see also Slavin, 1989). This definitely resonates here as virtually all of the research evidence indicates that these unrealistic and highly bureaucratized goals, as governmentally fixed, will likely elude us forever. These goals are indeed illusions leading us down a socially engineered road to a non-existent idealized destination. In fact, the M&P Theory of Change and its policy derivatives might be viewed as the greatest failed social engineering project of our time (Jehlen, 2009).

The question remains whether in fact the road to utopia is really the road trip to nowhere. Thirty years of evidence would support the latter. It seems that the nation's educational policymakers' attempts to socially engineer the public's belief systems and values have indeed worked so well that, to the nation's detriment, it cannot successfully socially engineer its way out of this nonsense, and off this prolonged and futile path. Educational reform in this case is failing ironically due in part to a lack of accountability to its own history.

Box 1.1 Top 10 assertions

1. Governmental and private groups often attempt to methodically sway public attitudes and behaviors to achieve idealized ends. By definition, a public policy is a tool used to do this.
2. The U.S. Department of Education's release of *A Nation at Risk* in 1983 prompted a fear about America's "failing" public schools. Since then, scholars have not only discredited the report, they have produced evidence to show the report was intentionally used for politically charged purposes.
3. *A Nation at Risk*, regardless, resulted in three decades of measurement-driven education policies, fueled by a Measure and Punish (M&P) Theory of Change.
4. The M&P Theory of Change is that by holding districts, schools, teachers, and students accountable for performance on the states' large-scale standardized achievement tests, administrators will supervise the schools better, teachers will teach better, and students will learn more, particularly in the nation's lowest performing schools.
5. The most recent federally-backed *Race to the Top* (RttT) program follows the M&P Theory of Change, requiring that states receiving RttT awards use students' test scores for even more consequential purposes and adopt even stronger accountability mechanisms than those required by *No Child Left Behind* (NCLB).
6. The M&P Theory of Change is fundamentally flawed and misguided at its core. Accordingly, it has never yielded its intended effects. It has never improved the nation's global standing, nor has it positively impacted the ever-persistent achievement gap.
7. The U.S. currently spends well over $1.7 billion on tests. This excludes the costs that come along with tests' associated resources (e.g., test preparation workbooks) and other related costs (e.g., administrative and instructional time for test preparation and administration).
8. Because of the increased reliance on tests, the M&P Theory of Change has caused intellectual genocide, primarily by denying many students proper, appropriate, fair, equal, and equitable opportunities to learn the things that will ultimately matter the most.
9. Metaphorically speaking, public schools in the U.S. are being silenced, given the educational opportunities and experiences that high-stakes tests are in many cases annihilating, especially in the schools most in need of guidance and support.
10. Educational reform, in this case, is failing due in part to a lack of accountability to its own history.

Notes

1 As part of the negotiations and final agreement to ban DDT, the industrial companies producing it were permitted to export the pesticide outside the U.S. until the mid-1980s (Griswold, 2012).

2 The U.S. is not the only country buying into this theory of change. In Australia, the former Prime Minister, Julia Gillard, while Minister for Education was also vying for Australia to be ranked no lower than #5 by 2025 on the PISA (Programme for International Student Assessment) international tests. What is the road to utopia in Australia? Test preparation, teaching to the PISA tests,

and increased school choice. This is also the case in Great Britain, another country infected by what is being termed the GERM (the Global Education Reform Movement) advancing such initiatives (Sahlberg, 2012; see also Ravitch, 2012b).

3 David Coleman is the main architect and proponent of the Common Core State Standards. He is an original member of the board of Michelle Rhee's StudentsFirst organization (see note 6 below); he previously ran an educational data and assessment company, and he was recently appointed as the president of the College Board curriculum and testing company.

4 Sandy Kress served as an advisor to President George W. Bush with respect to NCLB, and is also credited as one of its crafters. Kress, however, is not an academic and has not conducted research on the topic.

5 The quality of the journal in which this article was published is suspect. For more information see www.nonpartisaneducation.org/.

6 Michelle Rhee, the former chancellor of Washington DC's public schools and founder and current CEO of StudentsFirst, enacted a strict testing policy to which high stakes were attached, after which noteworthy cases of cheating were uncovered. Rhee denied such cheating occurred, but this ultimately led to her undoing as well as that of the then DC mayor.

7 Margaret Spellings, another one of NCLB's architects, followed Rod Paige as the U.S. Secretary of Education in the George W. Bush administration. She continues to speak out in favor of NCLB and criticizes the current federal administration under President Obama for turning soft on the matter.

8 In the late 1970s, as part of the minimum competency movement, the state of Florida implemented the first recorded state policy aligned with this theory of change. In Florida, state policymakers implemented a statewide test on which students were required to show minimum competence prior to being graduated from high school. Early gains in Florida test scores were used as an example of how this theory of change could improve education. However, soon thereafter, the immediate gains hit a plateau and then returned to where they were before the testing policy was implemented, leading Florida to abandon its new and promising testing policy. It was also discarded because of differential pass rates among racial groups and a discernible increase in high school dropout rates, particularly among ethnic minorities and students from low socioeconomic backgrounds (Linn, 2000; Serow, 1984). As well, because the content that was tested for "minimum" competence became the "maximum" on which students, particularly in urban schools, became competent (Bracey, 1995; U.S. Department of Education, 1983), the testing policy backfired. This is essentially what we continue to witness today.

9 The authors of the AERA's (2000) position statement on high-stakes testing wrote:

> When schools, districts, or other administrative units are compared to one another or when changes in scores are tracked over time, there must be explicit policies specifying which students are to be tested and under what circumstances students may be exempted from testing. Such policies must be uniformly enforced to assure the validity of score comparisons. In addition, reporting of test score results should accurately portray the percentage of students exempted.